Action at a Distance in Supercoiled DNA: Effects of Sequence on Slither, Branching, and Intramolecular Concentration

Dennis Sprous and Stephen C. Harvey
Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294 USA

ABSTRACT We report a computer modeling study of DNA supercoiling in model plasmids over the size range of 140-1260 bp. We used a computer model with basepair resolution. Molecular dynamics was used to produce ensembles at 300 K and to investigate intramolecular motions. The plasmid models varied by their sequence. The sequence types employed for comparison included a curve-bearing plasmid, a heterogenous sequence plasmid, and a homogenous sequence. Within the three sequence types tested at the 1260-bp plasmid size, we observed several sequence-dependent phenomena. Writhe, radius of gyration, slither motion, and branching probability were seen to be sequence dependent. Branching probability was the least in the homogenous plasmid and the greatest in the curve-bearing plasmid. The curve imposed a symmetry on the plasmid that was absent in the heterogenous sequence. Significant localizations and enhancements of intramolecular concentration were seen to a persistence length. Molecular dynamics allowed us to observe the mechanism of branch formation and reabsorption. We observed a size-dependent change in the types of motion observed in plasmids. Slither motion predominated in plasmids up to 600 bp in size, whereas global rearrangements were more important in the 1260 mer.

INTRODUCTION

Action at a distance is a phenomenon seen often in DNA studies. Interactions of distal sites require proper orientations as well as through-space proximity. One of the simplest means by which distant sites can be brought together is by looping caused by intrinsically curved DNA, by protein-induced bending, or by both. However, to complicate issues, DNA is inherently dynamic. There is mechanistic evidence for both global rearrangements and linear diffusion ("slither") motions (Krasnow and Cozzarelli, 1983; Spengler et al., 1985; Parker and Halford, 1991; Marko and Siggia, 1994). The first is a refolding of the plasmid, and the latter is the concerted slide of bases along a fixed global geometry. Direct observation of these motions is currently impossible because of the static nature of the most useful structural method, electron microscopy (Dustin et al., 1991).

Proteins associating and dissociating from DNA dynamically regulate DNA function and structure. The classical example is the loop seen in the arabinose operon (Lobell and Schleif, 1991). IHF (Higgins et al., 1989) and the μ-repressor (Vogel et al., 1991) are known to organize DNA loops dynamically and in so doing regulate specific transcription events. Replication (Echols, 1990) and recombination (Craig, 1991; Landy, 1989) also involve protein-induced DNA loops.

At the heart of all this are proteins bending DNA. DNA is not a straight rod. Intrinsic curvature arises from sequence effects, especially A tracts (Trifonov, 1985; Bolshoy et al., 1991). Proteins bend the helix. It has been shown that substitution of an intrinsically curved DNA sequence for a protein binding site will produce in vitro models functionally equivalent to the protein bound state of the system (Bracco et al., 1989, Gartenberg and Crothers, 1988; Goodman and Nash, 1989; Perez-Martin and Espinosa, 1993). Intrinsic DNA curvature is stable in solution (Hagerman, 1984). Thus, protein-induced DNA bending and intrinsic DNA curvature are to first approximations equivalent. As in protein:DNA complexes, it has been seen that intrinsically curved sequences will produce DNA loops in electron micrographs (Griffith et al., 1986). These loops exist both with (Laundon and Griffith, 1988) and without (Griffith et al., 1986) supercoiling pressure. In the presence of supercoiling pressure, two closely spaced loops were seen to increase the population of branched structures at the expense of interwound structures (Laundon and Griffith, 1988). Unfortunately, electron micrographs are a static presentation of a dynamic system.

Computer modeling offers a complement to static experimental studies. With computer models it is still possible to produce static structures corresponding to low-energy minima. A tacit assumption behind such studies is that the low-energy minimum found is near the center of the ensemble. However, it has been demonstrated that minimum energy structures are at best peripheral to the ensemble averages (Sprous et al., 1996). Thus, a more difficult, exciting, and relevant endeavor is the generation of thermodynamic ensembles at 300 K. With ensembles, relevant physical properties can be calculated for supercoiled DNA.

Any supercoiled DNA computer modeler makes a decision whether to model a larger DNA plasmid in low detail or to model a smaller DNA plasmid in greater detail (Malhotra et al., 1993; Schlick, 1995). At one extreme, the B-spline model developed by Hao and Olson (1989a,b;

Received for publication 17 July 1995 and in final form 18 January 1996.
Address reprint requests to Dr. Stephen C. Harvey, Department of Chemistry and Molecular Genetics, University of Alabama at Birmingham, UAB Station, BHSB, Birmingham, AL 35294-0005. Phone: 205-934-5028; Fax: 205-975-2547; E-mail: harvey@neptune.cmc.uab.edu.

0006-3495/96/04/1893/16 $2.00

1990) can represent a 1000-bp plasmid with fewer than 20 control points. At the other extreme, in the future on more powerful computers than today's machines, a modeler could attempt an all atom representation. For the time being, on more contemporary computers, several lower-resolution models have been devised. These models include bead representations (Chirico and Langowski, 1994; Langowski et al., 1994), finite elements (Bauer et al., 1993; Yang et al., 1993), spline functions (Hao and Olson, 1989a,b, 1990; Schlick et al., 1994) and rigid or elastic rods (Klenin et al., 1991; Vologodskii et al., 1992; Kremer et al., 1993; Vologodskii and Cozzarelli, 1993; Langowski et al., 1994). Our model (Tan and Harvey, 1989) represents each basepair as a plane defined by three points. Successive triads defining single basepairs are connected by springlike bonds. Thus the common helicoidal parameters of roll, tilt, and twist are reproduced in our model, and sequence-dependent details and anisotropic bending can be incorporated into this model. Though some degree of curvature can be incorporated into lower-resolution models, our model has the critical advantage of being able to address questions related to real sequences. To date, no one else has addressed specific questions of homogeneous versus heterogenous DNA. Unlike the other models, this model is able to follow groove orientations.

Simulation studies that treat curve-bearing plasmids are a smaller subset of supercoiled DNA simulations. Some critical information was seen in the earlier studies. A key observation of a finite-element study (Bauer et al., 1993) was that a single curve lowered the minimum linking deficit required to induce writhing. A Monte Carlo study illustrated that curves serve to dampen local internal motions, which correlates well with observed dynamic light-scattering data (Kremer et al., 1993). A later Monte Carlo study (Klenin et al., 1995) examined the effect of curved regions on intramolecular interactions and will be compared with our current study. Our own studies (Sprous et al., 1996; Tan et al., 1996) have shown that curves organize loops, hinder slither motion, and speed convergence times.

This paper and the previous companion papers (Sprous et al., 1996; Tan et al., 1996) are concerned with the production of plasmid ensembles at 300 K. We have employed a variety of structure types in these studies. The previous papers were concerned with relatively small plasmids, up to 600 bp. This paper concentrates on a series of 1260-mer plasmid sequences. To our knowledge, this is the largest plasmid simulation carried out with basepair-specific parameterizations.

We concentrate first on a single theme: a substitution of 50 intrinsically curved basepairs has dramatic effects both locally and globally for a 1260-mer plasmid. However, the effect is finite. The effect of the curve decays as distance from it increases. We will present and discuss some observed branching seen at the 1260-mer plasmid size as it relates to specific sequence types. By molecular dynamics, we were able to view the mechanism by which a branch is formed and reabsorbed. Lastly, observations about the speed of convergence as a function of plasmid size will be discussed.

MATERIALS AND METHODS

Models

We use a molecular mechanical model that we call 3dna for our study. It has been described in detail elsewhere (Tan and Harvey, 1989). Briefly, it represents each basepair as a plane defined by three particles. This level of detail is sufficient for basepair-specific parameterization of rise, roll, tilt, and twist.

Like all molecular models, 3dna has a potential energy function (McCammon and Harvey, 1987). The energy penalty function is composed of several terms:

$$E_{tot} = \sum k_{bonds}(\rho_i - \rho_0)^2 + \sum k_{angles}(\theta_i - \theta_0)^2 + \sum k_{torsion}(\phi_i - \phi_0)^2 + E_{nonbonds}$$

where the various k are force constants, variables subscripted with i are actual values of the ith particular bond, angle, or torsion, and a variable subscripted with 0 is the ideal measure for that constraint. A structure with no strained constraints will have an energy of zero. Any deviation from the ideal will produce a measurable energy. Relatively strong force constants hold the triad of particles together as a plane. Weaker force constants associated with roll, tilt, and twist are deformed.

Equilibrium values of roll and tilt in our models are related to sequence by the rules of Bolshoy et al. (1991). The equilibrium value of twist is related to sequence by the rules of Kabsch et al. (1982). Current structural methods are insufficient for the task of nailing down the exact origin of DNA curvature (Haran et al., 1994). However, the point of this paper is not critically tied to the function used to relate structure to sequence. It is tied to the ideal that there are distinct types of DNA structure.

Force constants used for the helicoidal parameters of roll, tilt, rise, and twist are based on the elastomechanical properties of DNA. The rise force constant and the twist force constant are based on experimentally measured DNA longitudinal (Putnam et al., 1982) and torsional (Barkley and Zimm, 1979; Hogan et al., 1983; Shore and Baldwin, 1983; Thomas and Bloomfield, 1983) moduli. The bending constraints, k_{roll} and k_{tilt}, are based on the known DNA persistence length (Hagerman, 1981; Rizzo and Schellman, 1981; Kam et al., 1981; Hogan et al., 1983). Previously (Sprous et al., 1996; Tan et al., 1996), we normally employed an isotropic bending mode such that the two bending constraints were equal. Theoretical and experimental data (Zhurkin et. al, 1979; Fratini et al., 1982; Nelson et al., 1987; Srinivasan et al., 1987; Sarai et al., 1988; DiGabriele and Steitz, 1993; Goodsell et al., 1993; Grzeskowiak et al., 1993) have shown that it is easier to roll (bending into or out of the major groove) rather than to tilt (bending into—and deforming—the phosphoribose backbone). All the 1260-mer simulations except the particular homogenous plasmid were done with anisotropic bending. In the case of anisotropic bending, our value of k_{roll} is 27.6 kcal mol^{-1} rad^{-2} and our value of k_{tilt} is 55.3 kcal mol^{-1} rad^{-2}. In the case of isotropic bending, k_{roll} and k_{tilt} were both set to a value of 43.7 kcal mol^{-1} rad^{-2}.

Both the bending and the torsional force constants can be verified by independent analysis. The bending force constants can be confirmed by a simple analysis, making use of molecular dynamic simulations of linear 3dna segments to generate large ensembles. The mean second, fourth, and sixth powers of the end-to-end distance are related to the persistence length by a theory developed by Hermans and Ullman (1952). This analysis has shown that our force constants produce a persistence length of 540 ± 50 Å in our models, in excellent agreement with experiment (Appendix). Further, the ratio $\Delta Wr/\Delta Lk$ is dependent on the balance of torsional and bending constraints. It is known that the slope of $\Delta Wr/\Delta Lk$ as a function of ΔLk in DNA is roughly 0.70 (Boles

et al., 1990). Earlier simulations showed that 3dna has the same behavior (Tan and Harvey, 1989).

In 3dna, a 50-Å nonbonded exclusion radius prevents strand crossovers. This corresponds to the experimental exclusion volume seen at an ionic strength of roughly 0.15 M (Stigter, 1977; Shaw and Wang, 1993).

DNA is a two-stranded polymer that has a physical linkage (Lk) of one strand about the other. Linkage is constant short of chemical compromise of the phosphoribose backbone. In a relaxed plasmid, linkage approaches the natural twist of the plasmid:

$$Lk_0 \approx Tw_0$$

Plasmids writhe when the linkage is unequal to the natural twist of the plasmid. In this case, the stress on the natural twist of DNA is accommodated by writhing, with total link being the sum of twist (Tw) and writhe (Wr) (White, 1969):

$$Lk = Tw + Wr$$

A quantity that is commonly used to set the degree of supercoiling in computer simulations is the ratio between the linking deficit, $Lk - Lk_0$, and the link number in the relaxed molecule, Lk_0. We will follow Sinden (1994), who calls this the "specific linking difference":

$$\sigma_{sp} = (Lk - Lk_0)/Lk_0 = \Delta Lk/Lk_0$$

This quantity is called the "superhelical density" by other authors (e.g., Vologodskii et al., 1992). We prefer to follow the nomenclature of Sinden (1994), who reserves that name for the corresponding experimentally measured quantity, the ratio between the number of measurable superhelical turns, τ, and Lk_0:

$$\sigma = \tau/Lk_0$$

Because of the torsional stiffness of DNA, ΔTw is small, and almost all changes in link are manifested as superhelical turns, so the specific linking difference and the superhelical density are nearly equal.

In nature, the average superhelical density is -0.05 (Zacharias et al., 1988). All 3dna models in this study were built at a specific linking difference of -0.05, which corresponds to the biological level of superhelical density.

Various structural types of various sizes were used in this series of studies (Table 1). The types are roughly described as being homogenous, heterogenous, and curve bearing. The homogenous plasmids are based on an ideal, average B-DNA helix. For these, each equilibrium value of roll and tilt is set to 0°. The twist equilibrium value is set to 34° for each successive basepair. The heterogenous sequence is a DNA sequence that is approximately straight according to Bolshoy rules (Bolshoy et al., 1991). The equilibrium conformations of the linear equivalents of both the crv1260 and the het1260 plasmids are shown in Fig. 1 *A*. Thus the characteristic of the het1260 plasmid is one of gentle undulations, whereas the crv1260 plasmid is identified by the presence of a severe curve. The curve-bearing plasmid is identical to the heterogenous, except for a 50-bp change to the sequence $(A_5G_5)_5$, which produces a 90° bend.

TABLE 1 Itemization of the various ensembles produced with the 3dna computer model

Tag	Type	Bending	Size (base pairs)
hom140	Homogeneous	Isotropic	140
crv140	Curve-bearing	Isotropic	140
het140	Heterogenous	Isotropic	140
hom315	Homogeneous	Isotropic	315
crv315	Curve-bearing	Isotropic	315
het315	Heterogenous	Isotropic	315
hom600	Homogeneous	Isotropic	600
crv600	Curve-bearing	Isotropic	600
het600	Heterogenous	Isotropic	600
hom1260	Homogeneous	Isotropic	1260
crv1260	Curve-bearing	Anisotropic	1260
het1260	Heterogenous	Anisotropic	1260

This paper concentrates on the 1260-mer ensembles; the smaller plasmids have been discussed elsewhere (Sprous et al., 1996; Tan et al., 1996).

Protocols

There are two protocols that need to be described here. The first is the minimization done to identify Lk_0. The second is the 1260-mer production molecular dynamics protocol.

Lk_0 is that linking number that has the lowest energy in a minimally writhed state. To determine this linking number we do the following. Several different closed unwrithed plasmids (flat circles) of differing linking number are generated. The structures are minimized by 200-step steepest descent followed by 50-step conjugate gradient minimization. That structure having the lowest energy is then identified as Lk_0.

For interwound plasmids of reasonable size, Wr depends on ΔLk but only weakly on plasmid size. Thus the homogenous 1260-mer starting structure for production dynamics at 300 K was produced by scaling up a smaller well-minimized plasmid. Extensive simulated annealing (Sprous et al., 1996; Tan et al., 1996) was performed to eliminate the unrelaxed aspects. Starting structures for the other plasmids were selected from the homogenous plasmid ensemble.

All ensembles went through the same production dynamics protocol. An initial equilibration phase was performed. This phase consisted of velocity rescaling every 10 ps for 10 ns. Actual production dynamics consisted of velocity rescaling every 25 ns. Coordinates were saved every 0.25 ns, because variations in Wr and the radius of gyration, r_g, are slow on this time scale. At the end of 25 ns, both velocities and structures were archived. The time step throughout this study was 25 fs. On an SGI R4400-based computer we were able to simulate roughly 0.75 ns/CPU h.

Our choice of nearly free dynamics has advantages and disadvantages. In a previous study (Sprous et al., 1996), we saw that frequent reassignment of velocities led to a loss of momentum, which in turn led to much slower sampling of conformational space. Exploiting momentum by frequent velocity rescaling or free dynamics led to much more efficient sampling of conformational space. In solution, motions would be damped by collisions with the solvent. Hence, this technique presents an accelerated image of structural transitions that do occur in DNA. Motions are observable, though accurate kinetics are not. In principle, both Langevin and Brownian dynamic (Chirico and Langowski, 1994) simulations are capable of presenting accurate kinetic information, provided that the hydrodynamic properties are accurately represented. Performing Langevin dynamics with 3DNA would require magnitudes-greater simulated time to reach convergence. Thus, it is computationally forbidden.

Analysis

r_g was determined by the positions of the sets of pseudoatoms at the center of each basepair:

$$r_g^2 = \sum[(x_i - \langle x\rangle)^2 + (y_i - \langle y\rangle)^2 + (z_i - \langle z\rangle)^2]/N$$

where x_i, y_i, and z_i are the coordinates of the center pseudoatom for basepair i; $\langle x\rangle$, $\langle y\rangle$, and $\langle z\rangle$ are the coordinates of the molecular center of mass, and the sum is over all N basepairs, $i = 1, N$.

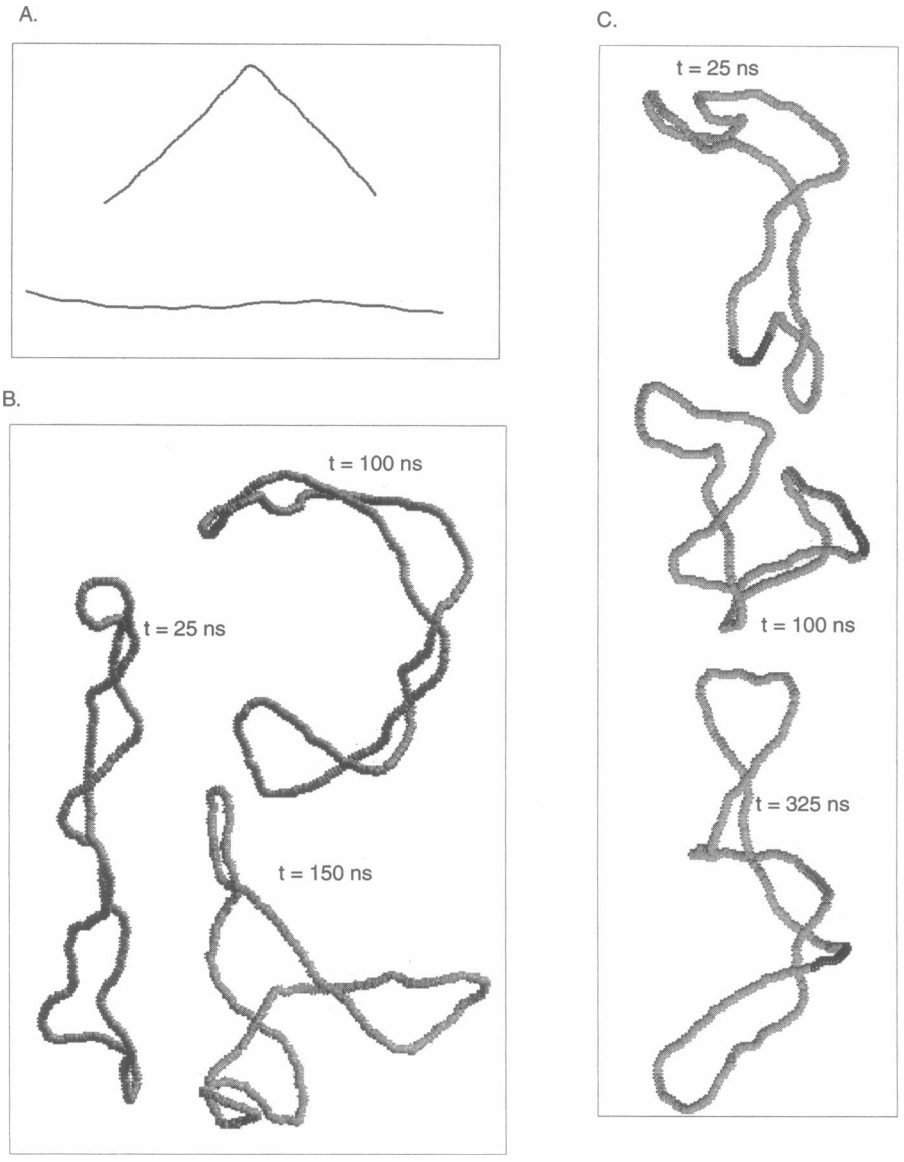

FIGURE 1 Various structures. *A*, Linear, lowest-energy equivalents to the crv1260 and the het1260 plasmids. *B*, Structures selected from the het1260 ensemble at various times. Dark shading denotes the curved region. *C*, Structures selected from the crv1260 plasmid ensemble.

We calculated writhe by using the Gauss integral (White, 1969). Linking number was determined by the projection method (Benham, 1986).

For both Wr and r_g, we calculated average, standard deviation, skew, and kurtosis. Average and standard deviation are common statistics. Skew and kurtosis are significantly more rare. They are defined by the following equations:

$$g_1 = (1/s^3(n-1))\sum(x-\langle x\rangle)^3 \quad \text{(skew)}$$

$$g_2 = (1/s^4(n-1))\sum(x-\langle x\rangle)^4 - 3 \quad \text{(kurtosis)}$$

Skew is a measure of the asymmetry of the distribution, i.e., how far the mode differs from the mean, whereas kurtosis is a measure of the distribution's peakedness. Normal distributions have skew and kurtosis of zero. A negative skew signifies a population heavier to the left of the mean, and a positive skew signifies a population heavier to the right of the mean. Leptokurtic populations are more heavily populated at the tails than at the mean, and kurtosis has a positive sign. Platykurtic populations are more heavily populated in intermediate regions at the expense of the populations at the mean and the tails and have a negative sign. A bimodal population would be an extreme platykurtic distribution.

Skew and kurtosis must be reported in context of their significance. Significance tests (t_s) were employed for both g_1 and g_2:

$$t_1 = |g_1/s_{g_1}|$$

$$t_2 = |g_2/s_{g_2}|$$

where s is the standard error of the given parameter and n is the population of the sample:

$$s_{g_1} = [6/n]^{1/2}$$

$$s_{g_2} = [24/n]^{1/2}.$$

Convergence in molecular dynamics simulations is difficult to guarantee. There is the danger that erroneous conclusions may be drawn from simulations that appear to be converged if there are significant structural rearrangements that occur more slowly than the time scale of the simulation. Different parameters may converge at very different rates. The most slowly converging parameter that we have seen is the antipodes profile.

The antipodes profile was described previously (Sprous et al., 1996). Briefly, an antipodes profile is a plot of the distances between basepairs that are diametrically opposite in the primary sequence. For any plasmid geometry other than a perfectly flat, unwrithed circle, those distances will not be identical at any particular time. However, for a homogenous plasmid, every basepair's parameterization is identical to the next. Thus, every basepair must spend an equal amount of time at any given conformation as its fellows. This requirement demands that, for all basepairs i and j at the time of convergence, the time average value $\langle d_{i,i+N/2} \rangle$ must be equal to the time average value $\langle d_{j,j+N/2} \rangle$.

The instantaneous antipodes profile resembles a sine wave function. As time increases, the time average antipodes profile will resemble a sine wave of decreasing amplitude. As seen in Fig. 2, the amplitudes of antipodes profiles for homogenous plasmids will reach zero at very long times. To estimate convergence time in the homogenous plasmids, we found that a single exponential decay equation works well:

$$A(t) = A_0 e^{(-t/\tau)}$$

where $A(t)$ is the amplitude at some time t, A_0 is the initial amplitude, and τ is the characteristic decay time.

For heterogeneous sequences, this simple equation does not hold. However, it was observed in smaller plasmids (Sprous et al., 1996; Tan et al., 1996) that all other parameters including r_g and Wr reached convergence more swiftly than the antipodes profiles. Also, it was observed that r_g and Wr reached convergence quicker for heterogeneous plasmids than for homogenous plasmids. The values of τ for the homogeneous plasmid are thus an upper limit for the time required to reach convergence for a given size plasmid.

One of the simplest explanations of action at a distance mechanisms is that curves selectively bring specific distal sites together. This effectively increases intramolecular concentration, leading to biological activity. Intramolecular concentration is expressed by the J factor (Jacobson and Stockmayer, 1950), in ligase studies (Shore et al., 1981; Shore and Baldwin, 1983), and the concept has been extended in simulations of DNA supercoiling (Vologodskii et al., 1992; Klenin et al., 1995).

Our calculation of intramolecular concentration is based on an extrapolation method (Hagerman and Ramadevi, 1990). Distances between basepairs of interest are collected across the simulation and sorted in descending order. Pools of equal population are then created. The smallest member of each pool and the largest member of that pool are used to define a

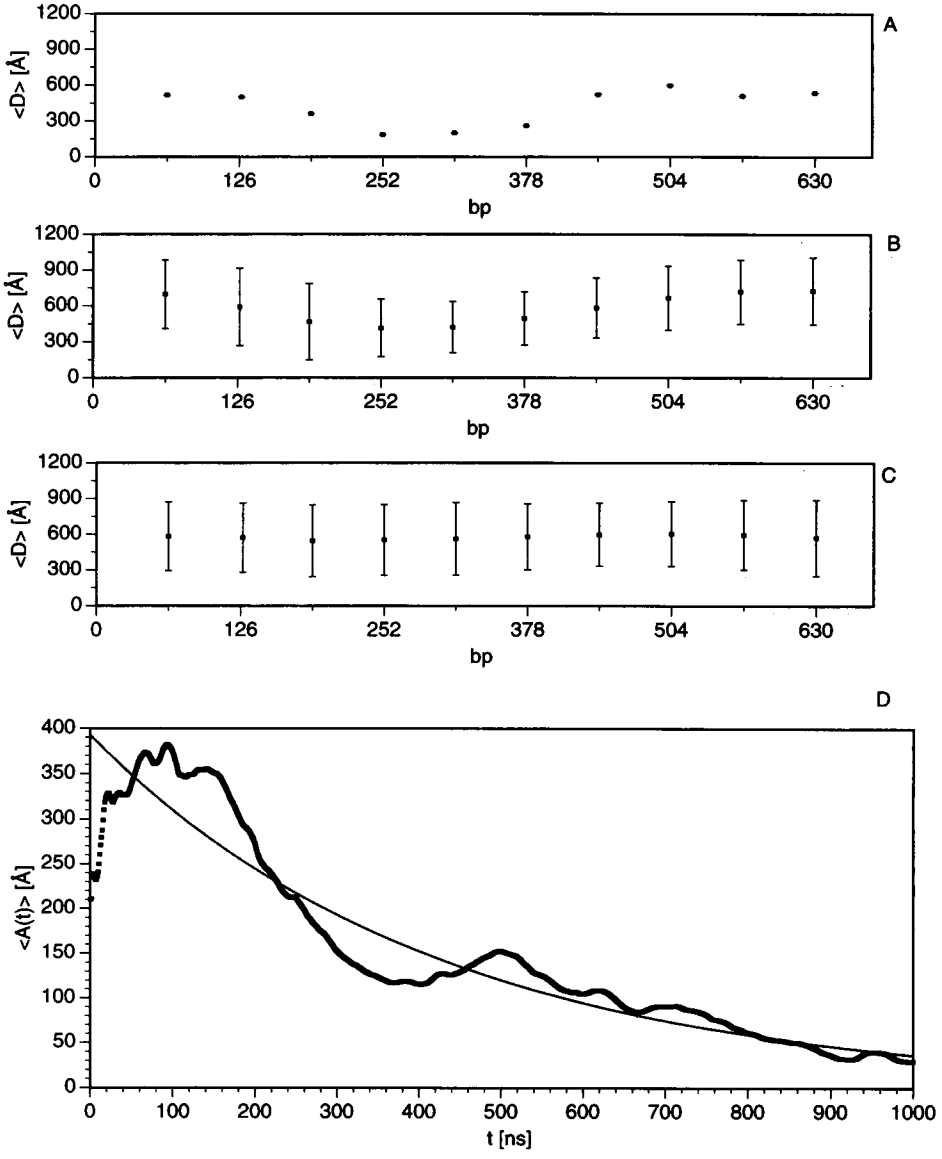

FIGURE 2 Antipodes profile plots and average antipodes profile amplitude for the hom1260 ensemble. *A*, Average antipodes profile at $t = 0$ ns. *B*, Average antipodes profile at $t = 500$ ns. *C*, Average antipodes profile at $t = 1000$ ns. *D*, Average amplitude of the antipodes profile as a function of time.

volumetric element. Concentration within the pool is calculated according to the following relation:

$$C[r_m] = (N_{pool}/N_{total})(10^{27}/N_A)/[4/3\pi(r^3_{largest} - r^3_{lowest})].$$

where N_{pool} is the population in the pool, N_{total} is the total population of the simulation, 10^{27} is a conversion factor for cubic ångstroms to liters, N_A is Avogadro's number, $r_{largest}$ is the upper limit of the pool, and r_{lowest} is the lower limit of the pool. r_m is the midpoint between $r_{largest}$ and r_{lowest}. Linear regression is performed on $(C[r_m],r_m)$ to yield $C[0]$. $C[0]$ is reported as the intramolecular concentration.

In addition to concentration, geometric orientation is also required for reactions to take place. To explore how a curve influences this parameter, we have analyzed the relation of the major groove of one basepair to a distal basepair. The angle formed by the major groove and the center of mass of basepair i to the center of mass of basepair $i + j$ was collected for j = 150, 300, 450, and 600 bp. As this angle will range from 0° to 180°, a random orientation will average 90°. Any other average value indicates a preferential orientation.

The two most distinctive features of supercoiled DNA are pseudonodes (apparent crossover points) and loops. However, there are no true nodes in supercoiled DNA, and the location of pseudonodes depends on the direction of view. Locations of loops from one view to another are more consistent. We have developed an algorithm to detect loop tips. The algorithm mimics the process used by a human to locate pseudonodes and loops. The algorithm starts by locating all the pseudonodes that are pairs of points at which the DNA appears to cross over itself. The continuous stretches of DNA between each pair of points are examined next. These are considered to form loops, and loops that totally enclose other loops are eliminated. What remain are the terminal loops, and the tips of these loops are taken as the points on the DNA farthest from the pseudonodes (Tan et al., 1996).

Loop-tip positions are presented as functions of time and present instantaneous as well as dynamic structural information. Loop tips are not just branches. Loop tips are also other irregularities in the superhelix. Stray points in loop-tip plots that appear are signals of superhelical breathing motions and other irregularities in the superhelix. Long continuous lines of loop-tip density that persist for tens of nanoseconds have been seen to be branches. Lines with nonzero slopes reveal slither motions as they occur.

Basepair numbering was done relative to the center of the curve insert for the crv1260 plasmid loop-tip plots. Thus, basepairs −75 and 75 are symmetrically arranged about the curve. For the het1260 plasmid, the sequence is identical save for the 50-bp curved substitution sequence. The hom1260 numbering was done simply 1 to 1260, as it had no special reference point. Plasmids are circles, so, for the het1260 and the crv1260 plasmids, basepairs 630 and −630 are adjacent. Likewise, for the hom1260 plasmid, basepairs 1 and 1260 are adjacent.

RESULTS AND DISCUSSION

A single curve can have dramatic results

We will concentrate on presenting a comparison of two ensembles that differ by 50 of 1260 bp. The linear lowest-energy equivalents of the plasmids are presented in Fig. 1 *A*. In the linear equivalents to the plasmids the effect of the curve on equilibrium structure is pronounced. The het1260 linear equivalent is a long rod, whereas the crv1260 linear has a 50-bp curve that produces a total bend of ~90°. For the ensemble of structures at 300 K this change effects both global and local parameters.

Dynamic structure: the DNA superhelix is rich in fluctuations

Fig. 1 presents typical structures from both the het1260 (Fig. 1 *B*) and the crv1260 (Fig. 1 *C*) ensembles. The structures are not remarkably different at first glance, but the curve 50-bp insert maintains a recognizable shape, whether located near a loop tip (t = 100 ns) or not. The curve can facilitate the bending of the overall interwound molecule (t = 325 ns).

The loop-tip plots of crv1260 and het1260 are quite distinct (Fig. 3). The loop-tip plot of hom1260 (Fig. 4) has a strong resemblance to the het1260 ensemble's loop-tip plot. First, of course, there is the nearly continuous density near basepair 0 in crv1260 that is absent in het1260 and hom1260. This is expected, as basepair 0 is the center of the curve. More significantly, there is a sequence-dependent effect on slither motion and branching probability. The curve acts to prevent long-range slither motion throughout the plasmid, and it increases the population of branched structures.

In the het1260 and the hom1260 plasmids there is slither motion, but this is disrupted in the crv1260 plasmid. In the het1260, the initial loop tips at basebairs 570 and −60 drift through the entire 1260-bp range over a time span of 450 ns. The hom1260 slither motion changes direction after ~475 ns. In the crv1260 plasmid, initial loops are at basepairs 20 and −560. The loop tip at basepair 20 is stabilized by the curved segment and does not slither much beyond basepair 50. The loop tip at basepair −560 disintegrates. It is replaced by a complex series of transient loops that slither through short distances before disintegrating.

Through most of the 450-ns range, het1260 is unbranched. Likewise, the hom1260 ensemble is largely unbranched over its full 1000-ns range. The stray points that pop up from structure to structure in no consistent fashion are irregularities in the superhelix rather than true branches. The structures at t = 25 ns and t = 100 ns in Fig. 1 *B* are both fairly typical of what is seen in the het1260 ensemble. The highly bent superhelix at 100 ns (Fig. 1 *B*) is an intermediate in the formation of a branched structure such as the trefoil seen at 150 ns. A similar structure was seen before the formation of a trefoil in the homogenous 1260-mer simulation. The formation of the superhelical bend is correlated to a drop in r_g but to no noticeable change in Wr.

The crv1260 plasmid is characterized by a single well-formed superhelix but multiple loops and local regions of superhelical irregularity. A trefoil is not uncommon. The occurrence of only two loop tips is rare. Regions where the superhelix opens and irregularities form are common. This produces a denser appearance in the crv1260 loop-tip plot than in the corresponding het1260 plot. The structures in Fig. 1 *C* are typical of the crv1260 ensemble. In the top two cases the curve region is at or near a loop tip. The brief discontinuity of loop-tip density in Fig. 3 (*bottom*) near 310–340 ns is of interest. The structure at

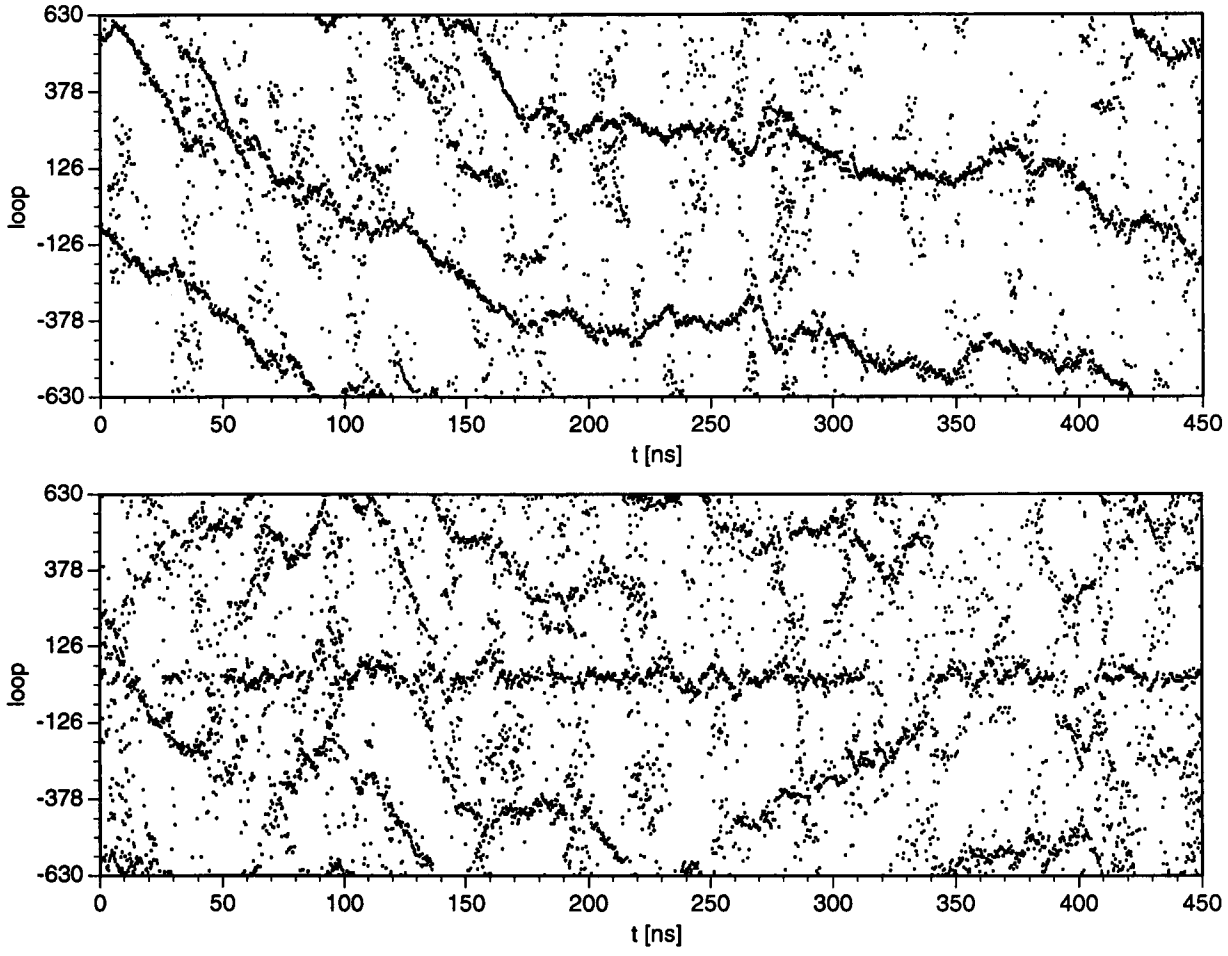

FIGURE 3 het1260 and crv1260 loop-tip location as a function of time. (*Top*) Loop-tip locations for the het1260 plasmid. (*Bottom*) Loop-tip locations for the crv1260 plasmid.

325 ns shows that the curve is not near a loop tip but has migrated to the center of the superhelix and has bent the whole superhelix (Fig. 1 *C*). A bent superhelix features prominently in the observed mechanism of branch formation and is discussed below.

Statistical properties of the ensembles vary according to sequence

Writhe and radius of gyration are global parameters that depend on the entire plasmid's conformation. In Fig. 5, population as a function of writhe and radius of gyration is presented. The two structures have distinct peaks. The crv1260 peak is taller and narrower and is distinctly shifted from the het1260 peak. These results are consistent with the idea that curves limit conformational space.

The causes of this are easier seen in the statistics given in Table 2. Average value, standard deviation, skew, kurtosis, maximum value, and minimum value are given for Wr and r_g for all the 1260-mer ensembles. The hom1260 ensemble is distinct from the other, more natural, heterogeneous sequence types. There are some key differences among the ensembles. The most striking difference is that the curve-bearing ensemble is normally distributed about almost all its parameters, whereas there are significant deviations from normality for the other two ensembles.

Within the r_g parameter the mean values for the three ensembles do not differ significantly. The differences among the three values of $\langle r_g \rangle$ are smaller than the standard deviation of any ensemble. However, there are some key differences to be seen in the ensembles. The root cause is that the curved region limits and organizes the amount of conformational space available to crv1260 plasmid. The r_g standard deviation of the crv1260 ensemble is 12 Å smaller than the standard deviation of the het1260 ensemble and more than 19 Å smaller than that of the hom1260. Both the het1260 and the hom1260 ensembles differ from normality for the r_g parameter. The hom1260 ensemble has a noticeable skew to the left and a significant tendency to kurtosis. Also, the het1260 ensemble has a significant kurtosis. The crv1260 ensemble is normally distributed and is limited to a

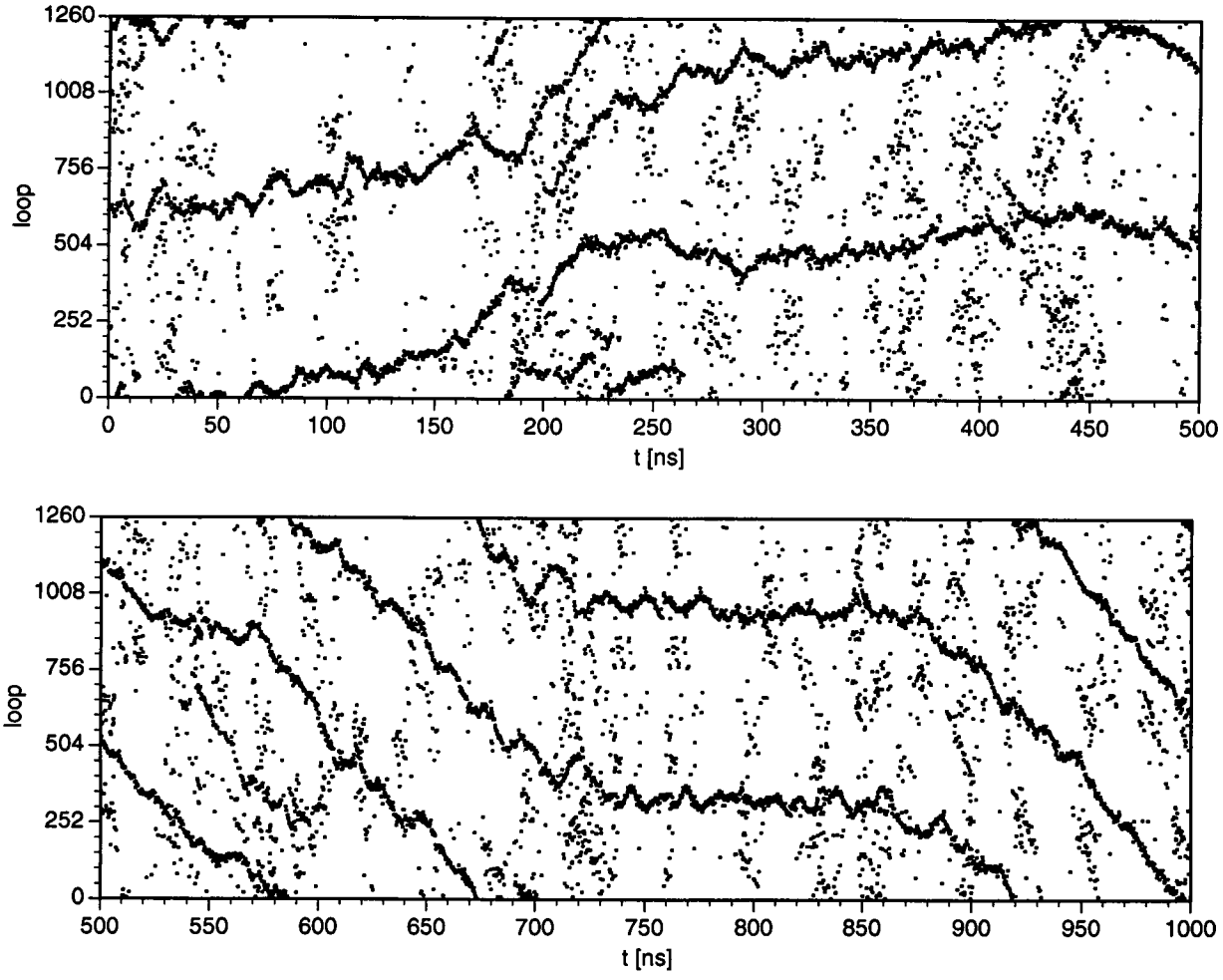

FIGURE 4 hom1260 loop-tip location as a function of time for the hom1260 plasmid. The top graph presents the first 500 ns, and bottom graph presents the last 500 ns.

much smaller range of possible structures than the other two ensembles. Here, the curve has organized the crv1260 ensemble into a somewhat limited region of conformational space.

The *Wr* parameter is simpler. The mean values of the three ensembles are significantly different, although the specific linking difference is the same in all three plasmids. The standard deviations of the three ensembles are close to one another. There is a significant leftward skew in the het1260 ensemble. Kurtosis is insignificant in all ensembles. Again, the crv1260 and the hom1260 distributions are normal.

How far can a curve be felt?

That the global parameters are altered this significantly between the crv1260 and the hom1260 ensembles is important. In terms of sequence content, the two plasmids differ by 50 of 1260 bp. Would these effects hold in still larger plasmids, or would the effect of the curve become less important for larger plasmids? Are the changes seen in global parameters evidence of far-reaching effects? Or would these effects be restricted because of the finite persistence length? (Schellman, 1974; Flory, 1989)

Recent simulations by Olson and collaborators (Zhang et al., 1994; Yang et al., 1995) lend support to the idea that a curve has far-reaching effects. In the first study, as the binding of the DNA is tightened around a histone (as the DNA is more sharply bent) large changes are seen in the global conformation of the entire plasmid. This result is consistent across plasmids of 300-, 500-, and 1000-bp length. The second study presents an image of intrinsic curvature organizing plasmids in subtle and far-reaching ways regardless of size. However, both studies were done by minimization techniques to reach the global minimum conformation—which corresponds to a low-temperature structure. Work done recently in this lab (Sprous et al., 1996; Tan et al., 1996) demonstrated that the global minimum is at best peripheral to the ensemble average within plasmids at 300 K. It is unclear whether the great changes

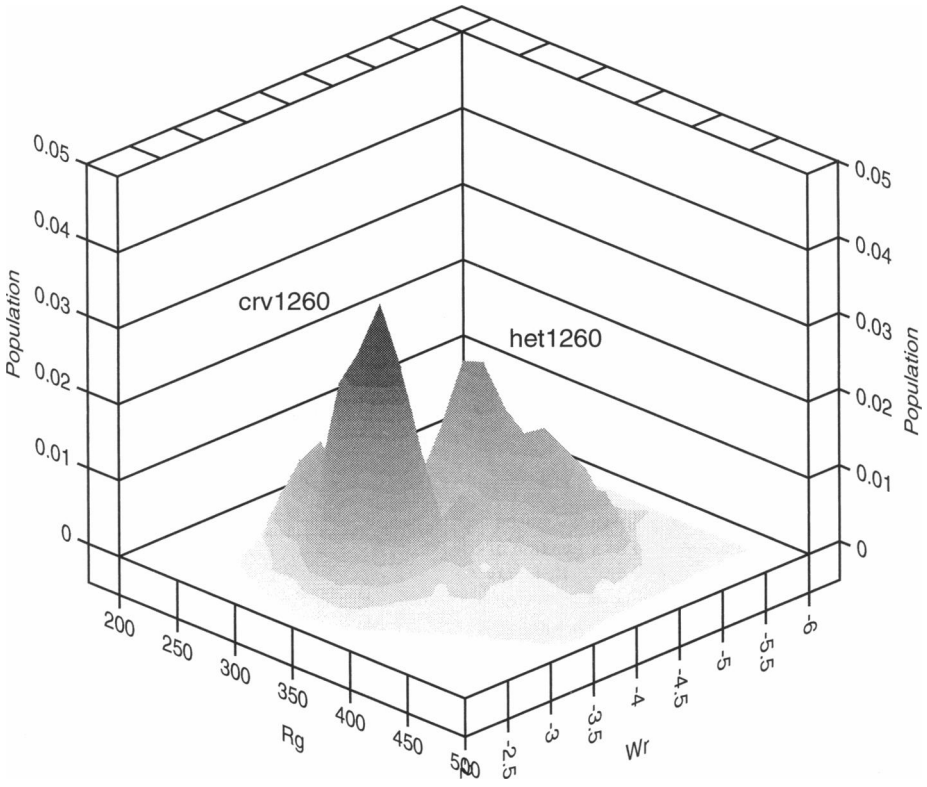

FIGURE 5 Population density $P(Wr, R_g)$ for both the crv1260 and the het1260 plasmid ensembles.

seen in the Olson papers would hold true for real plasmids at ambient conditions.

Recent work by Klenin and collaborators presents an alternative possibility (Klenin et al., 1995). They studied a 2700-bp curve-bearing plasmid by Monte Carlo methods at 300 K. Their model is less detailed than 3dna, which allows them to attempt larger plasmids. In Fig. 7 of that work, a nearly 10-fold decrease in intramolecular interactions was seen as the distance between equidistant sites increased from 140 to 1122 bp for a bend angle of 120°. As our plasmid is approximately the size of their weakest measured interaction, it is not surprising that we observe global changes. How far a curve affects a plasmid is a question at the heart of the action at a distance mechanism. We examine this issue next.

Distant sites on a DNA plasmid are capable of influencing one another. A reasonable explanation of the phenomenon demands both 1) simple increases in intramolecular concentrations and 2) proper geometric orientation. To explore these issues, we have calculated intramolecular concentrations. Also, we have looked at groove orientations of distant sites.

Fig. 6 presents the average intramolecular concentration as a function of basepair number for basepairs separated by 150, 300, 450, and 600 bp. For the crv1260 plasmid the maximum for intermolecular concentration as a function of basepair number should occur for basepairs symmetrically located with respect to the curve, so they should be near basepair $j/2$, where j is the separation distance in basepairs. For the het1260 it is not possible to

TABLE 2 Statistical properties of writhe and radius of gyration

tag	$\langle Wr \rangle$	s	Max	Min	g_1	t_1	g_2	t_2
hom1260	−5.12	0.33	−4.03	−6.15	0.08	2.1	−0.13	1.7
het1260	−4.57	0.35	−3.37	−5.74	−0.38	6.0	0.02	0.2
crv1260	−3.24	0.34	−2.10	−4.40	0.12	2.1	0.05	0.4
	$\langle r_g \rangle$							
hom1260	354	55	507	228	−0.11	2.8	−0.46	6.0
het1260	355	48	493	236	0.05	0.8	0.43	3.0
crv1260	325	35	440	240	0.005	0.1	0.08	0.7

Averages are indicated by angle brackets. *Wr* is writhe. Other properties are standard deviation *s;* maximum value Max; minimum value Min; skew, g_1; significance test of skew, t_1; kurtosis, g_2; and significance test of kurtosis, t_2. Critical values of t are the following: $t(P < 0.01) = 2.58$ and $t(P < 0.001) = 3.29$. Populations for the ensembles were 4000 structures for hom1260 and 1800 each for crv1260 and het1260.

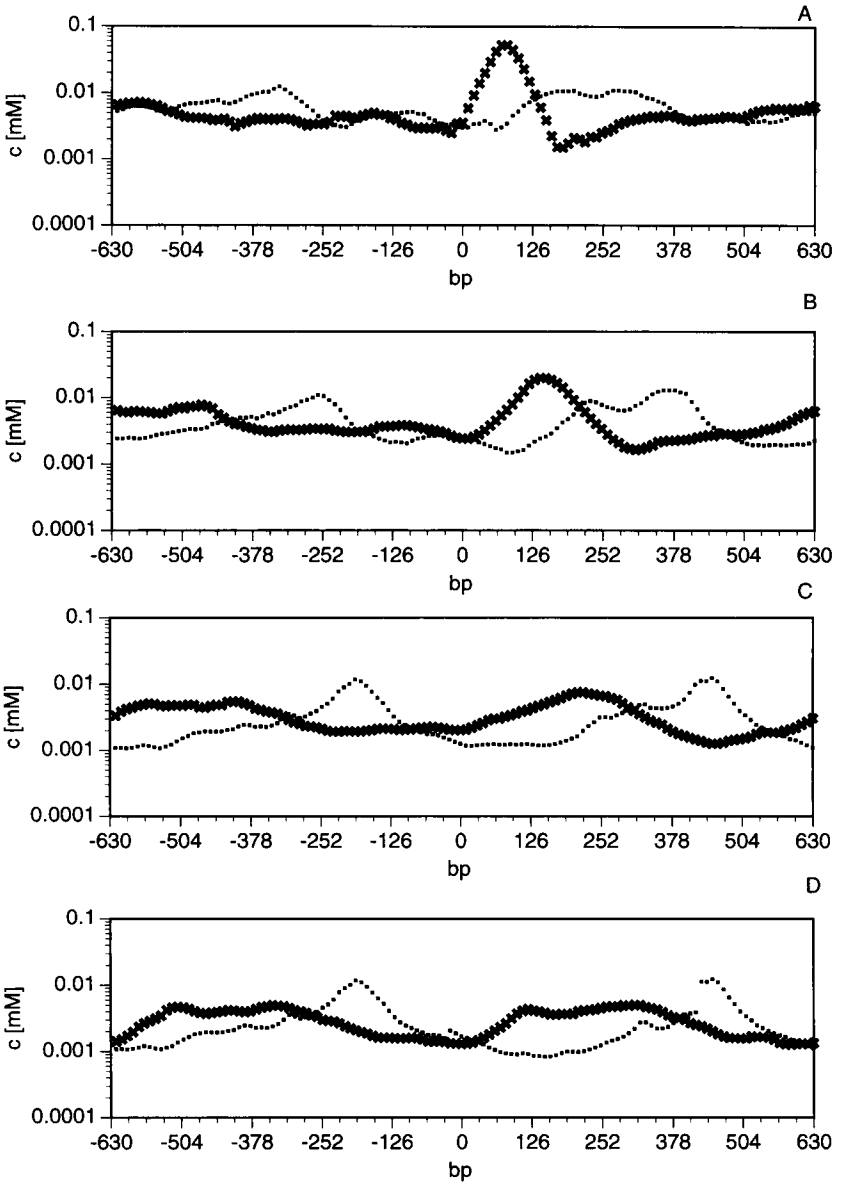

FIGURE 6 Time average intramolecular concentration of basepair i to basepair i-j as a function of basepair number for various values of j: ✖, points derived from the crv1260 ensemble; ■, points derived from the het1260 ensemble. Average intramolecular concentration for sites separated by (*A*) 150, (*B*) 300, (*C*) 450, and (*D*) 600 bp.

predict where the maximum will be. As the sequence is not homogenous, one can expect that there will be preferences.

The het1260 ensemble average intramolecular concentration is presented along with the crv1260 data (Fig. 6). Two points are seen from this analysis of the crv1260 ensemble. First, the maxima are at the predicted point in each plot for the crv1260 ensemble. Second, the height of the normalized maximum decreases as the separation between basepairs increases.

The maxima for the crv1260 sets are at basepairs 80, 140, 220, and 260 for 150-, 300-, 450-, and 600-bp separation. The plots show that for sites separated by 150 and 300 bp the effect of the curve is both an enhancement and a localization: the peaks are both comparatively tall and in the expected place. In the case of sites separated by 450 and 600 bp the effect of the curve is more a localization than an enhancement: the peaks are not comparatively so tall, but they are in the expected places. This is seen in Fig. 7, which presents the concentration enhancement. We calculate this value by dividing the maximum concentration by the average concentration of the points outside the peak. The curve-bearing plasmid experiences the greatest concentration enhancement near 150-bp separation and significant enhancements up to 300-bp separation. By 450 bp, the het1260 and crv1260 concentration enhancements are very similar. 150 bp is a single persistence length. We conclude that there is evidence that a curve has strong concentrating effects for a single persistence length on both of its flanks. Beyond that distance, some localization effects are seen, but they are not particularly significant.

An important question posed concerns the statistical error associated with the het1260 intramolecular concentration

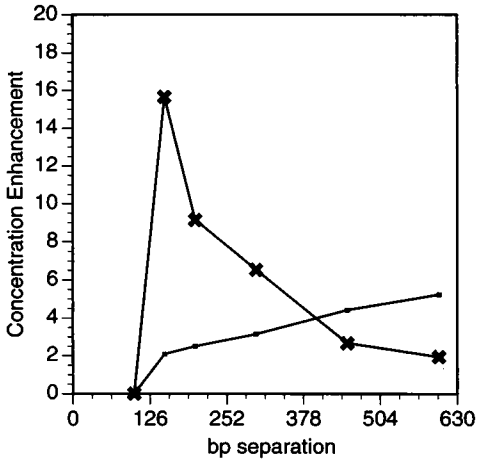

FIGURE 7 Concentration enhancement. The value of the maximum concentration divided by the value of the average concentration of equidistant sites as a function of separation, measured in basepairs: ✖, data from the crv1260 plasmid; ●, data from the het1260 plasmid.

plot. What would the profile look like at infinite time (i.e., what would the profile look like at convergence)? The exact shape cannot be guaranteed. However, there would be peaks and valleys in the final profile. This is because the heterogeneous plasmid will have preferred conformations. As the plasmid was designed specifically to lack strong curvature features (see Fig. 1), it is not possible to predict where the peaks and valleys will be. Although convergence cannot ever be guaranteed, our final profile matches the general shape expected for a heterogeous plasmid. The trends discussed above should be independent of the exact locations of peaks.

Groove orientation is a significant measure of intramolecular interaction in its own right. Systems exist that are strongly influenced by the orientation of grooves, including the arabinose operon (Lobell and Schleif, 1991). The plots in Fig. 8 present average major groove orientation angle between basepair i and basepair $i + j$ for $j = 150$ and $j = 300$ bp. The range of the angle is from 0° to 180°, so an average of 90° over several points signifies a random orientation. The results for the het1260 ensembles do not deviate more than ±10° from the average 90° for any distance of separation.

With the het1260 as a reference, the crv1260 plots are clearly well organized. Deviations from 90° as great as 30° are seen in the 150-bp separation case and greater than 20° in the 300-bp separation case, and in both the 450- and the 600-bp separation cases deviations of greater than 15° are seen. For separations of 150, 300, and 450 bp the regions of angular organization span the regions where intermolecular concentration maxima were. In the case of the 600-bp

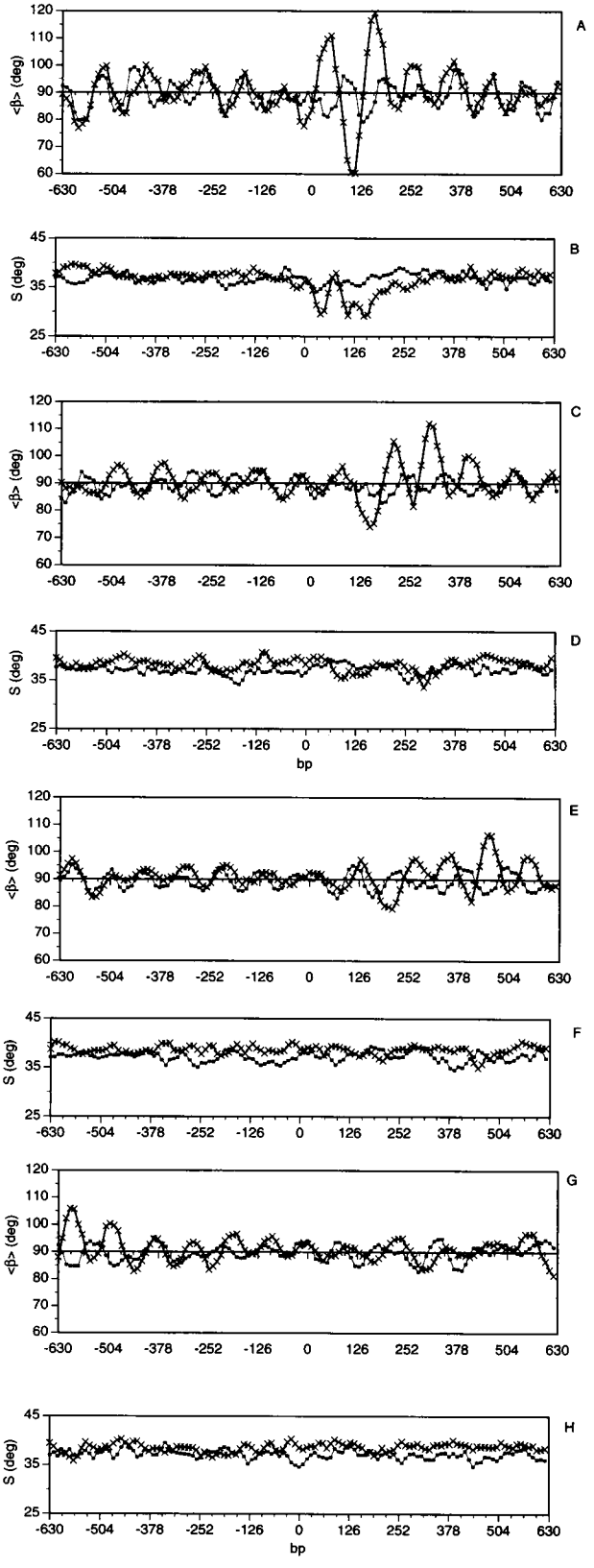

FIGURE 8 Time average major groove orientation angle of basepair i to distal basepair i-j ($\langle\beta\rangle$) and standard deviation of the same angle ($s = (\langle\beta - \langle\beta\rangle)^2\rangle^{1/2}$) as a function of basepair number for various values of j: ✖, points derived from the crv1260 ensemble; ■, points derived from the het1260 ensemble. *A*, $\langle\beta\rangle$ for $j = 150$ bp; *B*, s for $j = 150$ bp; *C*, $\langle\beta\rangle$ for $j = 300$ bp; *D*, s for $j = 300$ bp; *E*, $\langle\beta\rangle$ for $j = 450$ bp; *F*, s for $j = 450$ bp; *G*, $\langle\beta\rangle$ for $j = 600$ bp; *H*, s for $j = 600$ bp.

separation the region of enhanced angular organization spans a zone that is coupled to the region of enhanced intramolecular concentration seen in Fig. 6. As was seen in the intramolecular concentration analysis, the effect of a curve on groove orientation between basepairs decreases as distance between the basepairs increases.

We should be cautious in our conclusions concerning the limit of curvature effects because superhelical density may affect the limit at which a curve could be felt. Experimental work of Laundon and Griffith (1988) indicated that a curve could have further-reaching effects than what we present here. Though Laundon and Griffith (1988) do not state the superhelical density of their plasmids, for both it appears to be greater than $\sigma_{sp} = -0.05$, at which these simulations took place. Higher superhelical densities can be generated intracellularly, and they would be expected to increase the distance over which the effects of curvature are felt. Lower superhelical densities would introduce a greater degree of floppiness into plasmids, leading to an expected decrease of curvature's reach. The curvature reach presented here matches that of several biologically active systems (Benoist and Chambon, 1981; Dunn et al., 1984; Hahn et al., 1984; Dandanell and Hammer, 1985; Lobell and Schleif, 1991).

We draw three conclusions from the distance and angle plots. First, within a persistence length a curve can certainly be felt. Second, at 300 K the effects of a curve decay as distance from the curve increases. Lastly, beyond a persistence length some localization is seen but is not necessarily a significant enhancement. In larger plasmids, however, Klenin et al. (1995) report effects seen at distances up to ~1100 bp. Thus, our conclusions agree in spirit with the data presented by Klenin et al. (1995). Simulation work that appears to contradict our findings (Zhang et al., 1994) was done by minimization with nicked circles, i.e., in the absence of supercoiling pressure.

Branching is sequence dependent

Vologodskii and collaborators (1992) have presented a Monte Carlo study of supercoiled DNA. A key goal of that study was to present "at least to a first order, quantitative treatment of branching." Their study used a model in which a single element was equivalent to a tenth of a Kuhn segment, or roughly 30 bp. The model was homogenous. The results of that study showed that a 3.5-kb plasmid would have two, three, and four loop-tip structures in ratios of roughly 65:25:10. A 5.25-kb plasmid would have two, three, four, and five loop-tip structures in ratios of roughly 15:55:25:5. An implication of that paper is that the probability of branching for a 1260 mer would be low.

Experimental counts of branching (Laundon and Griffith, 1988; Andrian et al., 1990; Boles et al., 1990; Vologodskii et al., 1992) vary greatly and are functions of length, ionic strength, superhelical density, and nucleotide sequence, but it is generally believed that smaller plasmids such as 1260 mer are interwound and unbranched (Vologodskii et al., 1992). In our study we have held the length, ionic strength, and specific linking difference constant and have seen that nucleotide sequence can affect branching probability.

Inspection of Fig. 3 shows two long lines of continuous loop-tip density in the heterogeneous loop-tip plot. A third line appears and persists for tens of nanoseconds in the 100–200-ns time range. Hence, the rough number of stable branches is slightly greater than 2. The hom1260 simulation loop-tip plot is presented in Fig. 4. Two stable lines of loop-tip density are seen through most of the time range. From 160 to 270 and from 550 to 600 ns, a third line persists for tens of nanoseconds. This rough estimate presents branching in the hom1260 simulation to be on par with the that in the het1260 simulation. In the plot for the curve-bearing plasmid (Fig. 3, *bottom*), at almost any given time one can pick out three lines of stable loop-tip density. Thus, the number of branches in the crv1260 simulation approaches three. As was mentioned before, loop tips are not just branches but can be other irregularities in the superhelix. However, counting the number of reported loop tips does show the trend outlined above. The average number of loop tips is 3.1 ± 1.1 in the case of the hom1260, 3.2 ± 1.2 for the het1260 ensemble, and 3.5 ± 1.2 in the crv1260 ensemble (standard error is at worst 0.02 for the averages). Thus, the introduction of structural irregularities increases branching probability, with curve sequences enhancing this effect.

A possible reason for this trend is that the heterogeneous character of the het1260 leads to transient substructures that act as nucleation sites for branching. In the case of the hom1260 plasmid, every bending constraint is straight. Thus the potential for irregularities leading to branching is lower. It is reasonable that slither motion is involved in branch formation. Normally slither motion simply pushes basepairs along a fixed global geometry. In the crv1260 plasmid this is difficult. However, the momentum is still there, and the only place left to go is out to form a branch. An explicit slither motion was required in Vologodskii's Monte Carlo study (Vologodskii et al., 1992) to produce branching number changes in a reasonable amount of computer time.

By visual inspection of molecules and observed loop-tip density lines, all but the crv1260 simulations favored the shapes of figure 8's rather than trefoils. However, the phenomenon of branching in these small plasmids is somewhat unexpected, considering Vologodskii's work, which has been noted to agree closely with experimental observations (Vologodskii et al., 1992). There are differences between our model and Vologodskii's model that may account for some of these trends. First, Vologodskii's model is homogeneous, which we have noted is the condition that has the lowest propensity for branch formation. Beyond that, as our model employs a basic unit for every basepair and Vologodskii's employs a basic unit roughly every 30 bp, our model has a greater number of possible hinges on which to bend than Vologodskii's. It would be interesting to see whether branching propensity decreased in our model if

larger segments of DNA were used as the basic unit. Likewise, it would be interesting to see whether employing more basic units in Vologodskii's model would increase the observed branching propensity. This is not to say that the bending modulus in either model is inaccurate. Increasing the number of hinges affects the entropic component of the model, which may have repercussions on branching probability.

Our results agree with the experimental evidence that branching is a function of sequence (Laundon and Griffith, 1988). We conclude that branching can occur under some circumstances at the level of a 1260 mer. Heterogeneous sequences lead to a greater degree of branching, and curve sequences to still more. The ability of sequence and curvature to affect dynamic behavior is possibly more important than the strictly structural effects.

Mechanism of branch formation

Fig. 4 presents loop-tip location as a function of time for the hom1260 ensemble. Over the 1000-ns time span of the simulation, trefoil structures exist twice: first from 180 to 265 ns and later from 550 to 600 ns. By analyzing our archived structures, we were able to see the formation of the branch and later reabsorption of the branch back into the superhelix.

In Fig. 9, loop-tip position, $r_g(t)$, and $Wr(t)$ are presented over the 150- to 300-ns time range. The r_g parameter decreases over this range. *Wr*, by contrast, continues its oscillations without a clear trend. When we viewed numerous structures in the hom1260 ensemble over this time window, it became obvious that the superhelix was first bending. After bending, a bulge formed. Lastly, the bulge organized itself and became a branch. $r_g(t)$ is sensitive to this global change in conformation. By contrast, writhe does not give a distinct hint that anything is happening.

In Fig. 10 representative structures of the bend, bulge, and branch mechanisms are presented. For reabsorption, the sequence of events is reversed. The branch turns to a bulge. Then the bulge dies into a bent superhelix. $R_g(t)$ is correlated to these changes. The branch that formed near 180 ns disappears near 265 ns.

Convergence speed and the balance of slither versus global rearrangements

As explained in Materials and Methods, for homogenous plasmids the antipodes profiles can be modeled as a sine

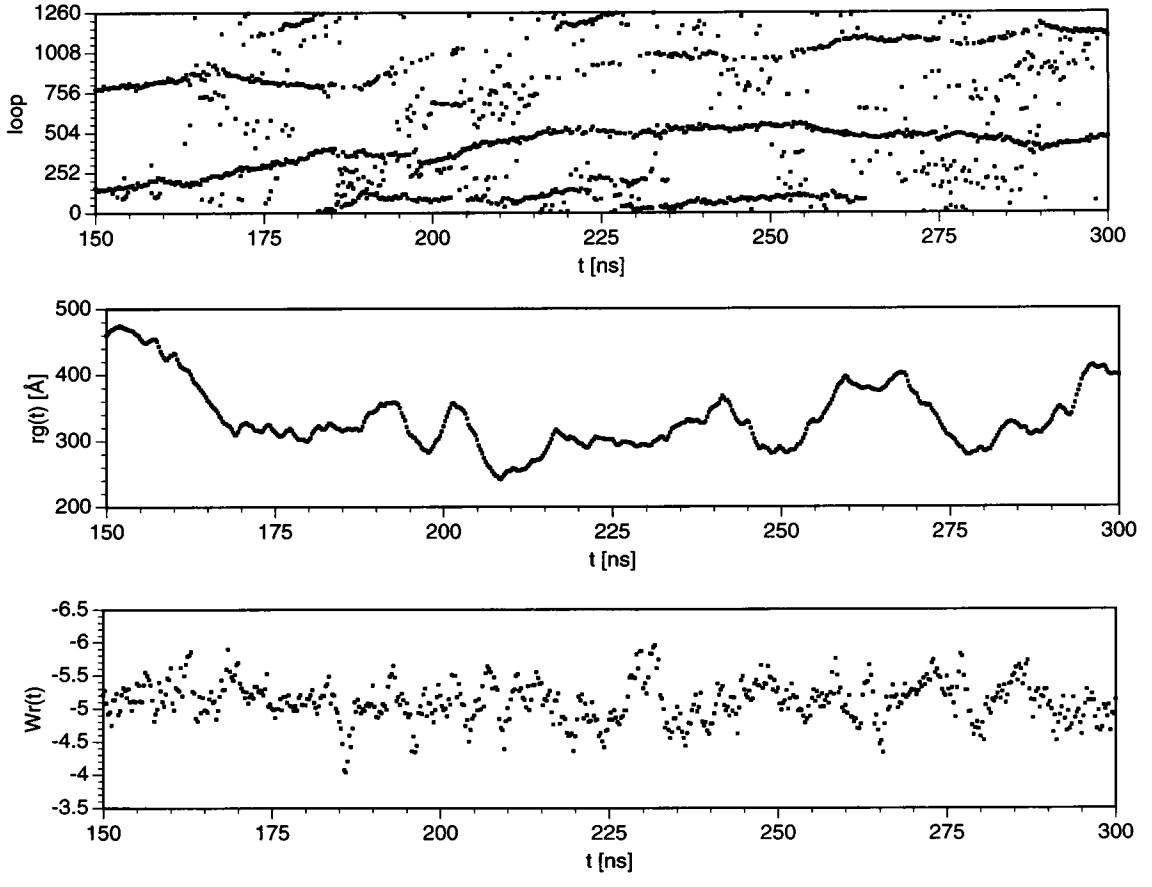

FIGURE 9 Parameters over the time range of branch formation (150–300 ns) for the hom1260 ensemble. (*Top*) Loop-tip location as a function of time. (*Center*) Radius of gyration as a function of time. (*Bottom*) Writhe as a function of time.

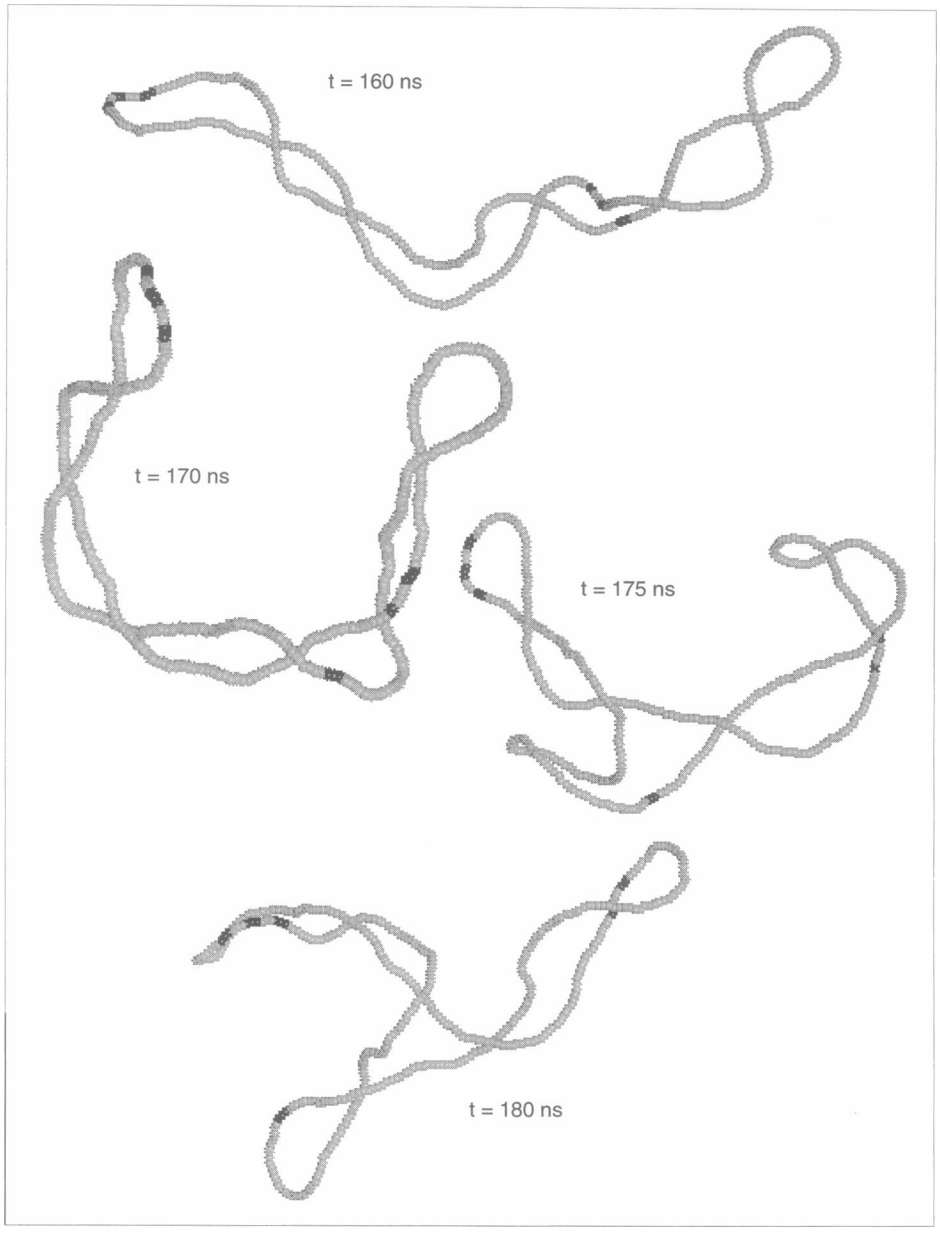

FIGURE 10 Structures from the hom1260 ensemble, illustrating the steps involved in branch formation. Black bands are included for visual reference. The single band is centered at basepair 0, the double bands at basepair 420, and the triple bands at basepair 840.

function. The decay of the amplitude of that sine wave has a characteristic time, τ. Table 3 presents the τ values seen in various DNA ensembles.

Within the smaller plasmids a trend can be seen in the growth of τ. As plasmid size roughly doubles, τ increases roughly by a factor of 7. This trend fails to continue to the size of the 1260 mer. For the homogeneous 600 mer, τ was 150 ns. The decay constant for the hom1260 case is 422 ns, not the predicted 1050 ns.

TABLE 3 Antipodes decay parameters for selected ensembles

Tag	A_0 (Å)	τ (ns)	r^2
hom140	12	3.0	0.90
hom315	120	21.0	0.78
hom600	284	150.0	0.96
hom1260	392	422.0	0.93

See text for definition of parameters.

Why did the trend fail? The answer lies in the balance between slither motion and global rearrangement motions. For the smaller plasmids, convergence demanded slither, because global rearrangements were difficult in plasmids containing only a few persistence lengths of DNA. Slither was nearly the only motion seen in the 140 mer. The 600 mers had more variety, but not much. The 1260 mers by contrast had a cornucopia of disturbances in the superhelix.

A 1260 mer is approximately eight persistence lengths long, whereas our earlier plasmids were one, two, and four persistence lengths. Naturally, as size increases, plasmids become more floppy, making global rearrangement motions

easier. Slither still occurs, but the plasmid's global geometry changes much more quickly than it did in the smaller plasmids. For the smaller plasmids, sampling conformational and intramolecular interactions was accomplished mainly by slither motions. In the 1260 mer, conformational space and intermolecular interactions were sampled by global rearrangements.

SUMMARY

Sequence affects both global and local parameters. The branching probability, writhe, and radius of gyration were altered significantly by the introduction of a curved 50-bp sequence. Intramolecular concentration profiles were enhanced strongly for up to a single persistence length on either side of the curve. The distributions of properties were made more nearly normal by the presence of a curve.

Heterogenous qualities and especially curvature hinder long-range slither motion. One apparent outcome of this constraint is that the momentum is directed outward, where it produces a greater amount of irregularity and branching in the superhelix. In previous studies of smaller plasmids (Sprous et al., 1996; Tan et al., 1996) such transitions were more difficult. Effects on dynamic behavior by curvature and sequence are possibly as important as the purely structural effects of curvature.

We have benefitted from stimulating discussions with Pat Higgins, Jacob Lebowitz, Robert Tan, and Wolfgang Zacharias. We are also indebted to the referees for thorough and thoughtful reviews.

APPENDIX. VERIFICATION OF THE PERSISTENCE LENGTH OF THE 3dna MODEL

Consider a random chain polymer of length L and persistence length P. In an infinitely large ensemble of such chains, the end-to-end distance R can be described only statistically. The second and fourth moments of the probability distribution for the end-to-end distance, $\langle R^2 \rangle$ and $\langle R^4 \rangle$, are given by the following expressions (Hermans and Ullman, 1952):

$$\langle R^2 \rangle = 2P\{L - P + Pe^{-L/P}\} \quad \text{(A1)}$$

$$\langle R^4 \rangle = (20/3)L^2P^2 - (208/9)LP^2 + (8/27) \cdot (e^{-3L/P} - 1)P^4 + 32(1 - e^{-L/P})P^4 - 8e^{-L/P}LP^3 \quad \text{(A2)}$$

These relationships can be used to verify the persistence length of a model DNA by generating ensembles for molecules of various lengths, measuring $\langle R^2 \rangle$ and $\langle R^4 \rangle$, and comparing these values with those calculated for different values of P.

In principle, these equations can be used for segments of any length L. In practice, the method breaks down for DNA segments shorter than 150 bp, because small changes in $\langle R^2 \rangle$ and $\langle R^4 \rangle$ correspond to large changes in persistence length.

To produce the ensembles, we first created a collection of six starting structures for molecular dynamics simulations. To do this, we chose the roll and tilt angles randomly from a normal distribution with a mean of zero and a standard deviation of 5°. This corresponds a persistence length of roughly 550 Å.

Long molecular dynamics simulations were then carried out on each of the six structures. Equilibration lasted 10 ns, with rescaling of velocities every 100 ps. Production molecular dynamics runs at 300 K lasted for 200 ns, with no velocity rescaling or reassignment. Structures were archived every 500 ps. The structures from all six runs were then combined for analysis.

This process was carried out for a 250-bp segment and for a 500-bp segment. The average persistence length from all simulation sets was 540 ± 50 Å.

REFERENCES

Adrian, M., B. ten Heggeler-Bordier, W. Wahli, A. Z. Stasiak, A. Stasiak, and J. Dubochet. 1990. Direct visualization of supercoiled DNA molecules in solution. *EMBO J.* 9:4551–4554.

Barkley, M. D., and B. H. Zimm. 1979. Theory of twisting and bending of chain macromolecules; analysis of the fluorescence depolarization of DNA. *J. Chem. Phys.* 70:2991–3007.

Bauer, W. R., R. A. Lund, and J. H. White. 1993. Twist and writhe of a DNA loop containing intrinsic bends. *Proc. Natl. Acad. Sci. USA.* 90:833–837.

Benham, C. J. 1986. Superhelical DNA. *Comm. Mol. Cell. Biophys.* 4:35–54.

Benoist, C., and P. Chambon. 1981. In vivo sequence requirements of the SV40 early promoter region. *Nature (London).* 290:304–310.

Boles, T. C., J. H. White, and N. R. Cozzarelli. 1990. Structure of plectonemically supercoiled DNA. *J. Mol. Biol.* 213:931–951.

Bolshoy, A., P. McNamara, R. E. Harrington, and E. N. Trifonov. 1991. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. USA.* 88:2313–2316.

Bracco, L., D. Kotlarz, A. Kolb, S. Diekmann, and H. Buc. 1989. Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli. EMBO J.* 8:4289–4296.

Chirico, G., and J. Langowski. 1994. Kinetics of DNA supercoiling studies by Brownian dynamics simulations. *Biopolymers.* 34:415–433.

Craig, I. 1991. Methylation and the fragile X. *Nature (London).* 349: 742–743.

Dandanell, G., and K. Hammer. 1985. Two operator sites separated by 599 base pairs are required for deoR repression of the deo operon of *Escherichia coli. EMBO J.* 4:3333–3338.

DiGabriele, A. D., and T. A. Steitz. 1993. A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* 231:1024–1039.

Dunn, T. M., S. Hahn, S. Ogden, and R. F. Schleif. 1984. An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. USA.* 81:5017–5020.

Dustin, I., P. Furrer, A. Stasiak, J. Dubochet, J. Langowski, and E. Egelberg. 1991. Spatial visualization of DNA in solution. *J. Struct. Biol.* 107:15–21.

Echols, H. 1990. Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination. *J. Biol. Chem.* 265: 14697–14700.

Flory, P. J. 1989. Statistical Mechanics of Chain Molecules. Hasner Publishers, New York.

Fratini, A. V., M. L. Kopka, H. R. Drew, and R. E. Dickerson. 1982. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATT(BrC)GCG. *J. Biol. Chem.* 257:14686–14707.

Gartenberg, M. R., and D. M. Crothers. 1988. DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature (London).* 333:824–829.

Goodman, S. D., and H. A. Nash. 1989. Functional replacement of a protein-induced bend in a DNA recombination site. *Nature (London).* 341:251–254.

Goodsell, D. S., M. L. Kopka, D. Cascio, and R. E. Dickerson. 1993. Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. USA.* 90:2930–2934.

Griffith, J., M. Bleyman, C. A. Rauch, P. A. Kitchin, and P. T. Englund. 1986. Visualization of the bent helix in kinetoplast DNA by electron microscopy. *Cell.* 46:717–724.

Grzeskowiak, K., D. S. Goodsell, M. Kaczor-Grzeskowiak, D. Cascio, and R. E. Dickerson. 1993. Crystallographic analysis of C-C-A-A-G-C-T-T-G-G and its implications for bending in B-DNA. *Biochemistry.* 32: 8923–8931.

Hagerman, P. J. 1981. Investigation of the flexibility of DNA using transient electric birefringence. *Biopolymers.* 20:1503–1535.

Hagerman, P. J. 1984. Evidence for the existence of stable curvature of DNA in solution. *Proc. Natl. Acad. Sci. USA.* 81:4632–4636.

Hagerman, P. J., and Ramadevi. 1990. Applications of the method of phage T4 ligase-catalyzed ring closure to the study of DNA structure: I. Computational analysis. *J. Mol. Biol.* 212:351–362.

Hahn, S., T. Dunn, and R. Schleif. 1984. Upstream repression and CRP stimulation of the *E. coli* L-arabinose operon. *J. Mol. Biol.* 180:61–72.

Hao, M. H., and W. K. Olson. 1989a. Global equilibrium configurations of supercoiled DNA. *Macromolecules.* 22:3292–3303.

Hao, M. H., and W. K. Olson. 1989b. Modeling DNA supercoils and knots with B-spline functions. *Biopolymers.* 28:873–900.

Hao, M. H., and W. K. Olson. 1990. Molecular modeling and energy refinement of supercoiled DNA. *J. Biomol. Struct. Dyn.* 7:661–692.

Haran, T. E., J. D. Kahn, and D. M. Crothers. 1994. Sequence elements responsible for DNA curvature. *J. Mol. Biol.* 244:135–143.

Hermans, J. J., and R. Ullman. 1952. The statistics of stiff chains with applications to light scattering. *Physica.* 18:951–971.

Higgins, N. P., D. A. Collier, M. W. Kilpatrick, and H. M. Krause. 1989. Supercoiling and integration host factor change the DNA conformation and alter the flow of convergent transcription in phage Mu. *J. Biol. Chem.* 264:3035–3042.

Hogan, M. E., J. LeGrange, and B. Austin. 1983. Dependence of DNA helix flexibility on base composition. *Nature (London).* 304:752–754.

Jacobson, H., and W. H. Stockmayer. 1950. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* 18:1600–1606.

Kabsch, W., C. Sander, and E. N. Trifonov. 1982. The ten helical twist angles of B-DNA. *Nucleic Acids Res.* 10:1097–1104.

Kam, Z., N. Borochov, and H. Eisenberg. 1981. Dependence of laser light scattering of DNA on NaCl concentration. *Biopolymers.* 20:2671–2690.

Klenin, K. V., M. D. Frank-Kamenetskii, and J. Langowski. 1995. Modulation of intramolecular interactions in superhelical DNA by curved sequences: a Monte Carlo simulation study. *Biophys. J.* 68:81–88.

Klenin, K. V., A. V. Vologodskii, V. V. Anshelevich, A. M. Dykhne, and M. D. Frank-Kamenetskii. 1991. Computer simulation of DNA supercoiling. *J. Mol. Biol.* 217:413–419.

Krasnow, M. A., and N. R. Cozzarelli. 1983. Site specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented Res sites. *Cell.* 32:1313–1324.

Kremer, W., K. Klenin, S. Diekmann, and J. Langowski. 1993. DNA curvature influences the internal motions of supercoiled DNA. *EMBO J.* 12:4407–4412.

Landy, A. 1989. Dynamic, structural and regulatory aspects of λ specific recombination. *Annu. Rev. Biochem.* 58:913–950.

Langowski, J., U. Kapp, K. Klenin, and A. V. Vologodskii. 1994. Solution structure and dynamics of DNA topoisomers. Dynamic light scattering studies and Monte Carlo simulations. *Biopolymers.* 34:639–646.

Laundon, C. H., and J. D. Griffith. 1988. Curved helix segments can uniquely orient the topology of supertwisted DNA. *Cell.* 52:545–549.

Lobell, R. B., and R. F. Schleif. 1991. AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. *J. Mol. Biol.* 218:45–54.

Malhotra, A., H. A. Gabb, and S. C. Harvey. 1993. Modeling large nucleic acids. *Curr. Opin. Struct. Biol.* 3:241–246.

Marko, J. F., and E. D. Siggia. 1994. Fluctuations and supercoiling of DNA. *Science.* 265:506–508.

McCammon, J. A., Harvey, S. C. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, London.

Nelson, H. C. M., J. T. Finch, B. F. Luisi, and A. Klug. 1987. The structure of an oligo(dA). Oligo(dT) tract and its biological implications. *Nature (London).* 330:221–226.

Parker, C. N., and S. E. Halford. 1991. Dynamics of long-range interactions on DNA. *Cell.* 66:781–791.

Perez-Martin, J., and M. Espinosa. 1993. Protein-induced bending as a transcriptional switch. *Science.* 260:805–807.

Putnam, B. F., E. W. Prohofsky, and L. L. van Zandt. 1982. Calculating composite bending, shearing, torsional and base tilting force constants of B- and A-form poly(dG)-poly(dC). *Biopolymers.* 21:885–894.

Rizzo, V., and J. Schellman. 1981. Flow dichroism of T7 DNA as function of salt concentration. *Biopolymers.* 20:2143–2163.

Sarai, A., J. Mazur, R. Nussinov, and R. L. Jernigan. 1988. Origin of DNA helical structure and its sequence dependence. *Biochemistry.* 27: 8498–8502.

Schellman, J. A. 1974. Flexibility of DNA. *Biopolymers.* 13:217–226.

Schlick, T. 1995. Modeling superhelical DNA: recent analytical and dynamic approaches. *Curr. Opin. Struct. Biol.* 5:245–262.

Schlick, T., W. K. Olson, T. Westcoff, and J. P. Greenberg. 1994. On higher buckling transitions in supercoiled DNA. *Biopolymers.* 34:564–597.

Shaw, S. Y., and J. C. Wang. 1993. Knotting of a DNA chain during ring closure. *Science.* 260:533–536.

Shore, D., and R. L. Baldwin. 1983. Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J. Mol. Biol.* 170:957–981.

Shore, D., J. Langowski, and R. L. Baldwin. 1981. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA.* 78:4833–4837.

Sinden, R. R. 1994. DNA Structure and Function. Academic Press, New York.

Spengler, S. J., A. Stasiak, and N. R. Cozzarelli. 1985. The stereostructure of the knots and catenanes produced by phage λ integrative recombination: implications for the mechanism and DNA structure. *Cell.* 42:325–334.

Sprous, D., R. K.-Z. Tan, and S. C. Harvey. 1996. Molecular modeling of closed circular DNA thermodynamic ensembles. *Biopolymers.* In press.

Srinivasan, A. R., Torres, R., Clark, W., and Olson, W. K. 1987. Base sequence effects in double helical DNA. I. Potential energy estimates of local base morphology. *J. Biomol. Struct. Dyn.* 5:459–496.

Stigter, D. 1977. Interactions of highly charged colloidal cylinder with applications to the double stranded DNA. *Biopolymers.* 16:1435–1448.

Tan, R. K.-Z., and S. C. Harvey. 1989. Molecular mechanics model for DNA supercoiling. *J. Mol. Biol.* 205:573–591.

Tan, R. K.-Z., D. Sprous, and S. C. Harvey. 1996. Molecular dynamics simulation of small DNA plasmids: effects of sequence and supercoiling on intramolecular motions. *Biopolymers.* In press.

Thomas, T. J., and V. A. Bloomfield. 1983. Chain flexibility and hydrodynamics of the B and Z forms of poly(dG-dC)-poly(dG-dC). *Nucleic Acids Res.* 11:1919–1930.

Trifonov, E. N. 1985. Curved DNA. *CRC Crit. Rev. Biochem.* 19:89–106.

Vogel, J. L., Z. J. Li, M. M. Howe, A. Toussaint, and N. P. Higgins. 1991. Temperature-sensitive mutations in the bacteriophage Mu c repressor locate a 63-amino-acid DNA-binding domain. *J. Bacteriol.* 173: 6568–6577.

Vologodskii, A. V., and N. R. Cozzarelli. 1993. Monte Carlo analysis of conformation DNA catenanes. *J. Mol. Biol.* 232:1130–1140.

Vologodskii, A. V., S. D. Levene, K. V. Klenin, M. Frank-Kamenetskii, and N. R. Cozzarelli. 1992. Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.* 227:1224–1243.

White, J. H. 1969. Self-Linking and the Gauss integral in higher dimensions. *Am. J. Math.* 91:693–728.

Yang, Y., I. Tobias, and W. Olson. 1993. Finite element analysis of DNA supercoiling. *J. Comp. Chem.* 98:1673–1686.

Yang, Y., T. P. Westcott, S. C. Pederson, I. Tobias, and W. K. Olson. 1995. Effects of localized bending on DNA supercoiling. *Trends Biol. Sci.* In press.

Zacharias, W., A. Jaworski, J. E. Larson, and R. D. Wells. 1988. The B- to Z-DNA equilibrium in vivo is perturbed by biological processes. *Proc. Natl. Acad. Sci. USA.* 85:7069–7073.

Zhang, P., I. Tobias, and W. K. Olson. 1994. Protein induced changes in DNA folding. *J. Mol. Biol.* 242:271–290.

Zhurkin, V. B., Y. P. Lysov, and V. I. Ivanov. 1979. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.* 6:1081–1096.