Circular Structures in Retroviral and Cellular Genomes*

(Received for publication, May 4, 1995, and in revised form, August 7, 1995)

Fred G. Albert†, Edward C. Bronson‡, Daniel J. Fitzgerald, and John N. Anderson§
From the Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907

A computer program for predicting DNA bending from nucleotide sequence was used to identify circular structures in retroviral and cellular genomes. An 830-base pair circular structure was located in a control region near the center of the genome of the human immunodeficiency virus type I (HIV-I). This unusual structure displayed relatively smooth planar bending throughout its length. The structure is conserved in diverse isolates of HIV-I, HIV-II, and simian immunodeficiency viruses, which implies that it is under selective constraints. A search of all sequences in the GenBank data base was carried out in order to identify similar circular structures in cellular DNA. The results revealed that the structures are associated with a wide range of sequences that undergo conformation, including most known examples of DNA inversion and subtelomeric translocation systems. Circular structures were also associated with replication and transposition systems where DNA looping has been implicated in the generation of large protein-DNA complexes. Experimental evidence for the structures was provided by studies which demonstrated that two sequences detected as circular by computer preferentially formed covalently closed circles during ligation reactions in vitro when compared to nonbent fragments, bent fragments with noncircular shapes, and total genomic DNA. In addition, a single T → C substitution in one of these sequences rendered it less planar as seen by computer analysis and significantly reduced its rate of ligase-catalyzed cyclization. These results permit us to speculate that intrinsically circular structures facilitate DNA looping during formation of the large protein-DNA complexes that are involved in site- and region-specific recombination and in other genomic processes.

Sequence-directed bending of DNA is one factor that causes local variations in the structure of genomes (reviewed by Crothers et al. (1990), Trifonov (1991), Hagerman (1992), and Harrington (1992)). The major determinant of bent DNA arises from oligo(A) tracts, which deflect the axis of the helix from a straight line. In bent sites, the tracts are spaced near the helical repeat of DNA, which causes their deflections to sum producing curvature of the DNA axis. Studies of bent DNA have most often focused on simple curved segments that are less than 100 bp in length. However, A tract bending has the potential to give rise to complex higher order structures, which may be important in regulating genome packaging and function. For example, the spatial organization of bending elements and straight DNA segments in replication origins and promoters is thought to be an important feature in the initiation of replication (Eckdahl and Anderson, 1987; 1990) and transcription (McAllister and Achberger, 1989; Pérez-Martín, 1994). Similarly, a multi-domain structure of nucleosome positioning DNA has been proposed to be a determinant for the deposition of nucleosomes at specific sites in chromatin DNA (Drew and Travers, 1985; Fitzgerald et al., 1994). It is generally assumed that intrinsic curvature facilitates the wrapping of DNA around regulatory and structural protein complexes, and it seems likely that the intrinsic shape of the curved DNA plays some role in the assembly mechanism. Thus, a description of higher order DNA structures beyond the level of simple curved loci may be needed for a complete understanding of the biological relevance of bent DNA.

Small planar DNA circles produced by ligation of synthetic oligonucleotide precursors have served as model systems for the study of one complex higher order structure in DNA (Ulanovsky et al., 1986; Ko et al., 1990; Crothers et al., 1992). To our knowledge, the only comparable structures that have been found in natural DNA are in kinetoplast minicircles of the trypanosome Crithidia fasciculata. These sequences have been shown to exist as circular structures by electron microscopy (Griffith et al., 1986). In this report, computer programs for calculating bending from nucleotide sequence were used to identify circular structures of various sizes in cellular and retroviral genomes. Ligation studies were then carried out in order to provide experimental evidence for the structures. The results show that large circular structures are associated with a wide range of DNA recombination systems and we suggest that this unusual structure plays a general role in DNA looping.

EXPERIMENTAL PROCEDURES

Computer Analysis—The standard program for calculating DNA bending from nucleotide sequence has been previously described (Eckdahl and Anderson, 1987). The program incorporates the dinucleotide twist angles reported by Kabsch et al. (1982) and uses the parameters of the wedge model for bent DNA (Ulanovsky and Trifonov, 1987). The program assumes that bending occurs only at AA/TT and uses wedge values estimated by Ulanovsky and Trifonov, (1987). A derivative program, which incorporates wedge values for all 16 dinucleotides (Bolshoy et al., 1991), has also been described previously (Wang et al., 1994) and was used for the analysis in Fig. 1E. The programs calculate three-dimensional coordinates of the helical axis along a sequence and then compute the ENDS ratio from these coordinates. The ENDS ratio is used to quantify predicted bending and is defined as the ratio of the contour length of a DNA segment to the shortest distance between the ends of the segment.

ENDS ratios were computed at the indicated window widths for PCR, polymerase chain reaction; Exo III, exonuclease III; HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; ori, origin of replication; IHH, integration host factor; HU, histone-like protein; Fis, factor for inversion stimulation; RT, reverse transcriptase.
sequences from release 73 of the GenBank data base (Benson et al., 1994). Details of the strategies used for analysis of both the data base and random sequences have been described previously (VanWyhe et al., 1991) and are outlined under "Results." The subdivisions of the GenBank library that were used in the studies shown in Fig. 3 were phage, bacterial, invertebrates, plants, organelles, nonmammalian vertebrates, rodents, primates, other mammals, and viruses. The automatic search of the structural RNA, synthetic, and unannotated categories were omitted. ENDS ratio analyses were performed on all sequences with lengths greater than 1000 bp. A total of 10,823 sequences (37% of the library) were analyzed. Upon request, the sequence locus names, nucleotide positions, and lengths of the highest 100 ENDS ratio peaks in each subdivision of the library will be provided. The accession numbers of the selected retroviral sequences used in Fig. 2 are given in Table I of Bronson and Anderson (1994). The accession numbers of sequences A–W from Fig. 4 and the base pair positions of ENDS ratio peaks in the sequences are: A, X62121 (6650); B, J01347 (1540); C, M18274 (3870); D, M32345 (1160 and 2700); E, J01801 (520); F, M11774 (291); G, M11195 (1500); H, M65025 (710); I, J00478 (850); J, Z11876 (530); K, M33720 (2620); L, M36467 (580); M, M57546 (3890); N, M36467 (580); O, M75456 (3890); P, M30176 (1450); Q, V00686 (1570); R, X14385 (5230); S, X04658 (4020); T, X02451 (2070); U, J01456 (1820); V, K01979 (550); W, M18394 (1100); M, M36467 (580); N, M57546 (3890); O, M30176 (1450); P, M63034 (1374); Q, V00686 (1570); R, X14385 (5230); S, X04658 (4020); T, X02451 (2070); U, J01456 (1820); V, K01979 (550); W, M18394 (1100). Variation is expressed as the standard deviation (± S.D.) or the standard error (± S.E.) of the mean.

General DNA Techniques and Electrophoresis—Standard molecular biology techniques were used for DNA manipulations (Sambrook et al., 1989). Double-stranded DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977) using a sequencing kit from U. S. Biochemical Corp. Oligonucleotides were synthesized by the Purdue University Oligonucleotide Synthesis Facility. Unless otherwise indicated, DNA fragments were analyzed by DNA sequencing in order to identify wild-type and mutant fragments. Following precipitation, DNA was dissolved in electrophoresis sample buffer and resolved on 1.3% agarose gels containing ethidium bromide. The apparent size of fragments on polyacrylamide gels was estimated by comparing mobilities of fragments to a 123-bp ladder (Life Technologies, Inc.) and to Had II digests of pUC18. Detection of Molecules and Ligase Reactions—An EcoRI fragment corresponding to the circular structure detected by computer in intron D of the RNA polymerase I gene from Caenorhabditis elegans (accession no. M29235) was prepared by PCR. Plasmid DR36 (Bird and Riddle, 1989) containing this segment was kindly provided by Dr. David Bird (University of California, Riverside). An XhoI/BamHI fragment from the plasmid was used as a template to amplify the 192-bp segment (bp 3809–4000), and the primers were designed such that the PCR product could be inserted into the EcoRI site of pUC18. EcoRI-digested recombinant plasmids isolated from transformed Escherichia coli (DH5α) were analyzed by electrophoresis on polyacrylamide gels and then by DNA sequencing in order to identify wild-type and mutant sequences. Of the 32 characterized plasmids, one point mutation was recovered, we concluded that it was introduced during PCR. The addition of wild-type sequence at position 3856. Ligation reactions were performed in a total volume of 20 μl with 0.1 units of T4 polynucleotide kinase as described by Drak and Crothers (1991). Unless indicated otherwise, labeled DNAs (1.0 μg/ml) mixed with nonradioactive EcoRI-digested recombinant plasmids (350 μg/ml) were equilibrated at 5°C in 30 m Tris (pH 7.8), 10 mM MgCl₂, 10 m dithiothreitol, 100 μM dNTPs. These high DNA concentrations and a 10-fold overabundance of noncircular fragments (Zahn and Blattner, 1987), Unless otherwise indicated, ligation was initiated by the addition of T4 DNA ligase (Promega) to a final concentration of 5 units/ml. Samples were removed at the indicated times and added to stop buffer (66 mM EDTA, 5% glycerol, 0.05% bromphenol blue) in order to quench the reaction. One ligated sample set was treated with exonuclease III (Exo III) (1 unit/ml) for 1 h at 37°C prior to the addition of stop buffer. Ligated products were resolved on 2% agarose gels in TAE containing ethidium bromide (Shore et al., 1981). Gels were subsequently dried and autoradiographed.

The circular structure that was detected by computer in the phase λ genome (bp 870–1750) was amplified and cloned as described above using BamHI-digested λ DNA as the template in the PCR. The ligation products of this 880-bp segment were compared to the ligation products of control (noncircular) EcoRI fragments from the troponin I (Koppe et al., 1989) and tropomyosin (Schmidt and Werner, 1991) genes, which are 700 and 1192 bp in length, respectively. EcoRI-digested recombinant plasmids (20 μg/ml) were equilibrated at 0–1°C in 30 m Tris (pH 7.8), 10 mM MgCl₂, 10 m dithiothreitol, 1 m ATP, and 1.8 m sucrose. Ligation was initiated by the addition of T4 DNA ligase. The final concentration of ligase was 10 units/ml unless otherwise indicated. In some experiments, 10 μg/ml of the ligation reaction were incubated at 3–120 min after the addition of ligase, mixed with stop buffer, and electrophoresed on 1.3% agarose gels containing ethidium bromide. DNA in the gels was transferred to nitrocellulose, and the blots were probed with nick-translated insert DNAs. In other experiments including the one shown in Fig. 11, ligation reactions contained all three digested plasmids (20 μg/ml each) and 0.5-mL samples were removed at the indicated times and added to 1 mL of ethanol. Following precipitation, DNA was dissolved in electrophoresis sample buffer and resolved on 1.3% agarose gels containing ethidium bromide. Photographic negatives of the gels and blots were scanned by computer, and DNA forms were quantified using the NIH image 1.52 program. Open circular, closed circular, and linear forms were identified by comparing the gel mobilities of unidentified DNA samples treated and not treated with Exonu I, by the restriction nuclease test described in Fig. 2 of Shore et al. (1981), and by polyacrylamide gel electrophoresis (Ulanovsky et al., 1986) (data not shown). Photographic negatives of ethidium bromide-stained gels are presented in Figs. 5, 8, and 11.

RESULTS

Computer Analysis of Retroviruses—A map of the human immunodeficiency virus type I (HIV-1, strain HXB2) is shown at the bottom of Fig. 1B. Given are the positions of the three long open reading frames (gag, pol, and env) and the locations of regulatory elements in the vicinity of the integrase gene at the end of pol. These elements include a transcriptional enhancer with two subdomains containing AP-1 and NF-κB binding sites (Verdin et al., 1990; Van Lint et al., 1991) and the central polypurine tract which serves as an initiation site for synthesis of the plus strand HIV DNA (Hungrnes et al., 1993; Charnreau et al., 1992; Hungrnes et al., 1993). A single-stranded gap that flanks this tract is seen in unincorporated copies of the linear double-stranded viral DNA (Charnreau and Clavel, 1991). A major DNase I-hypersensitive site is found between the enhancer subdomains when the provirus is packaged into cellular chromatin (Verdin, 1991).

Computer-generated helical projections of the HIV-1 proviral DNA are shown in Fig. 1A. The standard program that produced the projections is based on a model of intrinsic curvature, which assumes that DNA bending arises solely from AA/TT stacks. Deflections from linearity are quantified by the ENDS ratio, which is defined as c/d, where c is the contour length of a DNA segment and d is the shortest distance between its ends (Eckdahl and Anderson, 1987). In panel B, ENDS ratios were computed at window widths of 20–200 bp (top), 20–400 bp (middle), and 20–1000 bp (bottom) so that the corresponding short, intermediate, and long bent segments could be characterized in a single analysis. At the shorter window widths, the major site of bending occurred in the first 100 bp of reverse transcriptase. Helical projections of this segment are shown in Fig. 1C. The structure is characterized by an ENDS ratio value of 1.30 at a window width of 120 bp. This value is greater than 4 standard deviations above the mean of random and GenBank data base sequences (Fitzgerald et al., 1994). The site consists of two 40-bp regions of bending centered at nucleotides 2689 and 2757. The bending elements in the two regions are out of phase with each other by about 5 bp, which produces an S-shaped molecule. This structure is similar to the S-shaped configuration that is characteristic of nucleosome positioning.
DNA (Fitzgerald et al., 1994). Minor DNase I-hypersensitive sites, indicative of phased nucleosomes, have been mapped immediately adjacent to this structure on the 3' side (Verdin, 1991). At the intermediate window widths, regions of bending are noted at the gag-pol overlap, and at the 3' ends of pol and env. These deflections, as well as minor deflections at the 5' ends of gag and env, can readily be seen by the helical projections shown in panel A. A single unusually high ENDS ratio peak with a value of 73 is seen at the 830-bp window. The computer-generated structure corresponds to a circle of 830 bp (Fig. 1, A and D). The circular structure encompasses most of the integrase gene at the end of pol, and its central position at nucleotide 4800 is 22 bp away from the internal polypurine tract. The circular structure was also seen when the sequence was analyzed by a computer program that incorporates all 16 dinucleotide wedge angles.

In order to assess the conservation of the bent structures in retroviruses, the 60 full-length retroviral sequences from re-
B–D 3–20 (served in the 17 sequences from oncoviruses or spumaviruses, visna (data not shown). No comparable structures were ob-
in some nonprimate lentiviruses including the prototype virus
bending was also noted in the homologous regions of integrase
smooth planar bending throughout its length. Near planar
positions of the sites varied in the different viral groups (data
was seen in all isolates of HIV-1, HIV-2, and SIV, although
magnitudes of the ENDS ratio peaks in some SIV isolates
results in Fig. 3 show ENDS ratio values of the highest peaks
plotted against the window widths that gave the maximal peak
values. Data base sequences were scanned at window widths
from 20 to 1100 bp, and an ENDS ratio value >15 was used as
the criterion for a peak. This value is greater than 20 standard
deviations above the mean of the data base sequences and
greater than 10 standard deviations above the mean of the
highest ENDS ratio peak in each data base sequence (Fig. 3).
No ENDS ratio peaks >10 were seen when 1500 kbp of ran-
domized sequences were subjected to the analysis, which also

Fig. 2. Predicted DNA bending in HIV-1, HIV-2, and SIV. A
generalized map of the proviral genome showing positions of Gag, Pd, Env, P, RT, I, ltr (long terminal repeats) is given at the top of the figure.

lease 73 of the GenBank data base were examined as in Fig. 1.
In this collection, each of the known retroviral lineages is
represented by multiple sequences (Bronson and Anderson, 1994). Fig. 2A shows the positions and types of bending ele-
ments along the genomes of the primate lentiviruses HIV-1, HIV-2, and SIV. The ENDS ratio peaks that characterize each
bending element are more than 2 S.D. above the mean of
the indicated number (N) of full-length sequences representing the
seven viral groups were analyzed for bending as in Fig. 1B. The mean
ENDS ratio peak values for each group are represented. The types and
magnitudes of the peaks are represented by distinguishing symbols at
the indicated positions. The values of the ENDS ratio peaks are as
follows: 120-bp window: 1.14–1.18 (×), 1.19–1.25 (+), and >1.25 (▲);
400-bp window: 1.35–1.60 (●), >1.60 (□); windows from 400 to 1100 bp:
3–20 (open box) and >20 (bold open box). The mean ± S.D. ENDS ratio
peaks for randomized sequences at window widths of 120 and 400 bp
were 1.05 ± 0.03 and 1.14 ± 0.10, respectively. B–D, two-dimensional
projections of the helical axes of the integrase regions of HIV-1 (RF
isolate), HIV-2 (ISY isolate), and SIV (sooty mangelbe). The length,
central positions, and ENDS ratios of the depicted sequences are,
respectively: (B) 820, 4320, 96; (C) 1000, 4620, 100; (D) 1060, 5070, 36.
There is 60–65% nucleotide sequence identity between B and C and
between B and D.

which are the other two subfamilies of retroviruses. The circular
structures in the primate lentiviruses had an average
ENDS ratio of 50 with circumferences that ranged from 780 to
1060 bp. The center of the structure in every genome was
within 60 bp of the internal polypurine tract. The conservation
of the predicted structure was of interest because there is as
much as 40% divergence in the nucleotide sequence of the
circular regions shown in Fig. 2 (B–D; see legend), and the
divergence is even greater among the integrase sequences from
the viruses depicted in Fig. 2A. This is significant because insertions or deletions of as few as a single bp in the central
200-bp regions, or a small number of random substitutions
(less than 5%), can abolish the computer-generated circular
projections of the sequences shown in Fig. 2 (B–D) (data not
shown). In addition, the structures are heterogeneous in size
and shape as seen in Fig. 2 (B–D). Consequently, the exchange
of homologous segments of the integrase gene from HIV-1, HIV-2, and SIV failed to generate circular structures of the
types shown in Figs. 1 and 2 (data not shown). These results
imply that selective pressures have maintained the integrity of
the circular structure despite extensive sequence divergence.

FIG. 3. Characteristics of high ENDS ratio peaks in GenBank
sequences. ENDS ratios of the 10,823 sequences in the GenBank data
base that were greater than 1000 bp in length were computed at window widths from 20 to 1100 bp as in Fig. 1B. The mean (± S.D.)
ENDS ratio values of all sequences in the data base at each window step
(×) and the mean of the highest ENDS ratio peaks per data base
sequence (+) are shown for the indicated window widths. At a window
width of 1500 bp, the corresponding ENDS ratio were 1.7 and 2.9. Each
of the 135 sequences that displayed an ENDS ratio peak >15 is indicated
by a symbol in the top portion of the graph, and the window width
at which the peak maximum was observed is given on the x axis.
Symbols represent peak centers found in untranslated DNA (●), protein
coding DNA (▲), coding DNA within 200 bp of the end of a gene (□), and
the integrase regions of HIV-1, HIV-2 or SIV (□). The two structures
indicated by the arrows are described under "Results."

Computer Analysis of Cellular Sequences—A search of all
sequences in the GenBank data base containing >1000 nucle-
otide was carried out as in Fig. 1B in order to estimate the
frequency of high ENDS ratio peaks in cellular DNA. The
results in Fig. 3 show ENDS ratio values of the highest peaks
plotted against the window widths that gave the maximal peak
values. Data base sequences were scanned at window widths
from 20 to 1100 bp, and an ENDS ratio value >15 was used as
the criterion for a peak. This value is greater than 20 standard
deviations above the mean of the data base sequences and
greater than 10 standard deviations above the mean of the
highest ENDS ratio peak in each data base sequence (Fig. 3).
No ENDS ratio peaks >10 were seen when 1500 kbp of ran-
domized sequences were subjected to the analysis, which also

Computer Analysis of Cellular Sequences—A search of all
sequences in the GenBank data base containing >1000 nucle-
otide was carried out as in Fig. 1B in order to estimate the
frequency of high ENDS ratio peaks in cellular DNA. The
results in Fig. 3 show ENDS ratio values of the highest peaks
plotted against the window widths that gave the maximal peak
values. Data base sequences were scanned at window widths
from 20 to 1100 bp, and an ENDS ratio value >15 was used as
the criterion for a peak. This value is greater than 20 standard
deviations above the mean of the data base sequences and
greater than 10 standard deviations above the mean of the
highest ENDS ratio peak in each data base sequence (Fig. 3).
No ENDS ratio peaks >10 were seen when 1500 kbp of ran-
domized sequences were subjected to the analysis, which also

Circular Structures in Genomes 23573
illustrates the unusual nature of the peak structures in natural DNA. A total of 135 peaks were observed in the 10,823 data base entries, but this frequency is an overestimate of circular structures of the type shown in Fig. 2. Since visual inspection of the helical projections revealed that only about half of the peak sequences displayed planar bending throughout their lengths (data not shown). All 18 peak sequences with lengths between 360 and 580 bp were found in untranslated regions of the genome. We presume these elements are excluded from coding DNA because the closely spaced oligo(A/T) tracts needed to generate such peaks would impose severe coding constraints. Peak sequences with lengths from 620 to 1100 bp also displayed a preference for untranslated DNA, although significant fractions resided in sites that overlap coding and noncoding regions or in more central areas of coding DNA.

Three groups of sequences are overrepresented in Fig. 3 when compared to their frequencies in the GenBank data base. First, 20 of the 135 sequences are found in the integrase region of primate immunodeficiency viruses and the highest data base peak (ENDS ratio = 392) was from this region of HIV-1 (MN isolate). Second, 11 of the peak sequences were found in untranslated DNA from C. elegans and these sequences were particularly prevalent among the shorter peak sites uncovered by the search. This overrepresentation is consistent with the observation that the genome of this nematode is highly bent (VanWye et al., 1991). Third, a significant fraction (17/50) of the remaining peak sequences longer than 800 bp were associated with a class of molecules which includes known mobile elements and related sequences. Helical projections of sequences that constitute this group are shown in Fig. 4.

Alterations in gene expression caused by inversion of DNA segments has been extensively studied in bacteria and in yeast (reviewed by Cox (1988), Glasgow et al. (1988), Stark et al. (1992), and Van de Putte and Goosen (1992)). In these systems, a recombinase promotes strand cleavage and rejoining within inverted repeat sequences that flank the invertible element. Efficient inversion in bacteria also requires a recombinational enhancer. The enhancer and the gene for the recombinase are typically located within, or adjacent to, the invertible element. Fig. 4 (A–D) shows helical projections of four invertible segments that were detected as ENDS ratio peaks in the data base search. In each system, the circular structure terminates within 200 bp of at least one inverted repeat sequence and the structures in the two bacterial sequences terminate within 200 bp of a recombinational enhancer. The 2.1-kbp invertible DNA segment that controls pilin phase variation in Moraxella bovis is particularly striking since it consists of two circular structures containing the two pilin genes (Q and I), which are separated by a straight stretch of DNA (Fig. 4D). In order to determine if DNA circularity is a characteristic of invertible systems, we examined the inversion regions responsible for the well characterized phase variation of flagellin proteins in Salmo nella typhimurium and fimbrial protein genes in E. coli. The invertible DNA segments in both systems appear as C-shaped structures as shown in Fig. 4 (E and F). The cin invertase gene in phage P1 and P7 and the pin invertase gene from E. coli also appeared as C-shaped structures when viewed by the computer program for bending (data not shown).

DNA looping is thought to facilitate the assembly of complex nucleoprotein structures during synapse formation in inversion systems (reviewed by Schléf (1988), Wang and Giaever (1988), Eckols (1990), and Matthews (1992)). An analogous looping mechanism has been implicated in the transposition of the phage Mu transposon (Surette et al., 1989; Heichman and Johnson, 1990; Mizuuchi, 1992; Surette and Chaconas, 1992). As shown in Fig. 4G, the A gene for the Mu transposase and the B accessory transposition gene are contained within a circular structure, which is adjacent to the transpositional enhancer. Initiation of replication in plasmid R6K utilizes a DNA looping mechanism, which is mediated by the P1 initiator protein (Mukherjee et al., 1988a, 1988b). The protein binds to multiple enhancer sites in the γ-ori, and the resulting DNA-protein complex then loops to and activates the β-ori located ~1.2 kbp downstream. A circular structure is seen at the site of loop formation in the intervening DNA between the γ- and β-oris and its 5′ end terminates in the γ-enhancer element (Fig. 4H).

Class switching of immunoglobulin heavy chain constant region sequences (Cm) occurs through a recombinational event in switch regions located 5′ to each Cm (reviewed by Coffman (1993)). The commonmost form of switching involves the looping out and excision of chromosomal DNA between the most 5′ switch region (S5′) and one of the switch regions located further downstream along the chromosome. Switch recombinational sites are clustered within a region at the 5′ edge of the S5′ segment and this region is immediately downstream of a circular structure in the mouse genome (Fig. 4I).

Recombination events near the ends of linear chromosomes are responsible for antigen variation in the African trypanosome (Van der Plou et al., 1992; Pays et al., 1994), in the spirochete Borrelia hermsii (Barbour, 1993), and in the African swine fever virus (Almendral et al., 1990; de la Vega et al., 1990; González et al., 1990). These organisms presumably evade the immune system of their mammalian hosts by periodically switching the expression of members of multigene families that code for surface antigens. In the trypanosome and spirochete, a transcriptionally silent copy of a surface protein gene is typically duplicated and transferred unidirectionally from a distal site in the genome to a transcriptionally active telomeric expression locus. Similarly, duplication and translocation of sequences located near the left end of the 170-kbp swine viral genome is thought to have played a role in generating the antigen variation seen among different isolates of the virus (González et al., 1990). As shown in Fig. 4 (J and K), circular structures were noted in the expression locus several kilobase pairs upstream from the surface antigen genes in the spirochete and the trypanosome. Comparable O-shaped structures (Fig. 4L), as well as C-shaped structures (data not shown), were also noted within 1 kbp upstream of the antigen genes in both systems. Both locations have been implicated in recombinational events (Pays et al., 1989, 1994; Barbour et al., 1991; Barbour, 1993). Multiple unusual structures also characterized the left end of the swine virus genome. O-shaped (Fig. 4, M and N) and C-shaped (data not shown) structures were seen in this region at locations corresponding to sequences that are the most variable among viral isolates (see Fig. 7 in De La Vega et al. (1990)).

Type II introns encoding reverse transcriptase (RT introns) are common in the organelles of lower eukaryotes. These sequences are thought to be mobile elements, although the mechanism of their transposition is unclear (Lambowitz and Belfort, 1993). The data base search in Fig. 3 detected four RT introns from organelles, which are shown in Fig. 4 (O–R). The RT sequence from the cation etched virus also displayed a circular structure as seen in Fig. 4S. This virus is thought to be more closely related to transposons than to plant viruses (Doolittle et al., 1989). Helical projections of the 34 RT introns and related sequences analyzed by Mohr et al. (1993) were assessed in order to determine whether DNA circularity is a conserved feature of this group. The results revealed that about 80% of the sequences displayed either O-shaped or C-shaped structures (data not shown). Organelles from plants and lower eukaryotes also contain mobile elements in the form of small
plasmids whose origins are poorly understood (Ray et al., 1987; Kempken et al., 1992). The data base search identified four of these sequences as having high ENDS ratio peaks (>15) in the search from Fig. 3 except the C-shaped structures in G and U, which are longer than 1100 bp. Top panels, sequences from inversion and looping systems: A, min system from plasmid 15 B of E. coli (Sandmeier et al., 1991); B, FLP system for inversion of the 2 µm plasmid of Saccharomyces cerevisiae (Cox, 1989); C, FLP system for inversion of a plasmid from Zygosaccharomyces bailii (Utatsu et al., 1987); D, pilin inversion system from M. bovis (Fulks et al., 1990); E, hin system for flagellar variation in S. typhimurium (Szekely and Simon, 1983); F, fim system for fimbrial variation in E. coli (Abraham et al., 1985); G, genes A and B coding for the phage Mu transposase and accessory gene (Harvey et al., 1985); H, plasmid R6K replication-looping system (Kelly and Bastia, 1991); I, the Cγ switch region (Sµ) from mouse (Sakano et al., 1980). The maps adjacent to each set of projections show the positions of the circular elements (—), recombinase cleavage sites ( ), enhancers (●), and genes (M). Heavily lined boxes indicate genes for site-specific recombinases. The recombinational enhancers bind either Fis (A, D, and E) or Mu transposase A (G), and the replication enhancer in H binds the P1 initiator protein. Note that a circular structure terminates within 200 bp of each enhancer. Multiple recombination sites in I are located in the region between the two arrows (Matsuoka et al., 1990). Scale bars are given in the right lower corner of each panel. Middle panels, sites with telomeric recombination systems located: J, in a VMP expression site of Borrelia hermsi (Barbour et al., 1991); K and L, in a VSG expression site of Trypanosoma brucei (Pays et al., 1990; Boothroyd and Cross, 1982); M and N, in the left-terminal located variable region of the swine fever virus (Almendral et al., 1990; González et al., 1990). Bottom panels, mitochondrial plasmids and other reverse transcriptase coding entities from: O–R, Type II (RT) introns (Wahleithner et al., 1990; Wissinger et al., 1991; Lazowska et al., 1980; Siemeister et al., 1990); S, ORF5 of the carnation etched virus (Hull et al., 1986); T and U, S-1 and S-2 plasmids from CMS-S Zea mays mitochondria (Paillard et al., 1985; Levings and Sederoff, 1983); V and W, kinetoplast plasmid minicircles of Leishmania tarentolae and Crithidia fasciculata (Kidane et al., 1984; Ray et al., 1986). The coordinates of the ENDS ratio peaks in sequences A–W are given in "Experimental Procedures."
further illustrated by a consideration of all peak sequences that were detected among the 3643 sequence entries that constitute the viral, organelle, and bacterial subdivisions of GenBank. There were 11 viral sequences and 5 sequences from organelles with ENDS ratio peaks >40. All of these structures are shown in Fig. 2 (top panel) and Fig. 4 (O, P, T, V, and W). Five of the 11 peak structures from the bacterial data base with ENDS ratio peaks >40 are also shown in Fig. 4. Of the remaining six structures from bacteria that are not shown in the figure, two were associated with ori sequences, one with a toxin gene, and three with genes that encoded pilin-like proteins. Three of the four genes were from pathogens, as were most of the bacterial recombinational systems shown in Fig. 4. To our knowledge, these genes and ori sequences have not been tested for recombinational or looping activities. Many of the bacterial sequences with ENDS ratio peaks between 15 and 40 were also from pathogens, and the circular structures were most often found in or near genes that encode toxins, membrane proteins, or proteins that form filamentous appendages. Pathogenic bacteria frequently display phenotypic variations of their surface components and toxins and genetic recombination is a prevalent mechanism that mediates these phenotypic changes (reviewed by Robertson and Meyer (1992)). Thus, it is conceivable that these systems also exhibit phenotypic variation mediated by recombination and the circular DNA structures could play a role in the process.

Experimental Analysis of Selected Structures—Two high ENDS ratio peaks with lengths of less than 200 bp were identified in the search of the GenBank data base (see Fig. 3). One sequence was from kinetoplast minicircles of C. fasciculata, and this K-DNA has been visualized as a circular structure in electron micrographs (Griffith et al., 1986). This sequence appeared as a nearly perfect circle when analyzed by the computer program (Fig. 4W). The other sequence was found within intron D of the RNA polymerase II gene of C. elegans. Experiments were performed using this intron segment in order to provide additional evidence for the existence of small circles in natural DNA. Fig. 5 gives some of the characteristics of the EcoRI fragments used in the work. Two of the fragments (pJ A71-2 and pSV2CAT) have electrophoretic properties and ENDS ratio values that are within the ranges of those seen for the bulk of natural DNA (Anderson 1986; Eckdahl and Anderson, 1987; Van Wye et al., 1991; Fitzgerald et al., 1994). The three satellite sequences display significant bending, and the sequence from dove has nearly as many phased oligo(A) tracts as is found in K-DNA and in the C. elegans intron segment (Fitzgerald et al., 1994). However, the dove satellite monomer is composed of two bent loci, which are out of phase with each other by approximately one-half helical turn. This gives rise to the computer-generated S-shaped structure shown in the figure (Fitzgerald et al., 1994). The 198-bp intron segment has an $R_L$ value (apparent chain length/real chain length) which is similar to that of the K-DNA sequence (Kitchin et al., 1986). The magnitude of the electrophoretic anomaly displayed by these segments is greater than those reported for all other natural sequences of this length. The temperature dependence of the gel anomaly was also similar to K-DNA. The $R_L$ values for both sequences declined by about 50% upon increasing the temperature from 5°C to 35°C (Diekmann and Wang, 1985; Fig. 5). Mutations in the intron sequence were recovered following PCR amplification and cloning. The computer-generated structure of one sequence carrying a point mutation that converts a $T_e$ tract to TCT$_e$ is shown in the figure. The helical projections of this sequence are less planar than those displayed by the wild-type sequence, and this difference is reflected in a reduction in the ENDS ratio and a small but reproducible reduction in electrophoretic anomaly.

A comparison of the cyclization of the DNA sequences shown in Fig. 5 should provide a stringent test for the predictive power of the computer program used in this study. In Fig. 6, DNA fragments were end-labeled and then incubated with T4 DNA ligase at 4°C for the indicated times in the presence of an excess of nonradioactive digested plasmid DNA. Following incubation, samples were electrophoresed on agarose gels in the presence of ethidium bromide in order to resolve circular and linear species (Shore et al., 1981). Inspection of the resulting autoradiograms revealed that the control and satellite DNA molecules preferentially formed linear multimers during the ligation, and nearly all of the radioactivity was sensitive to ExoIII digestion upon termination of the experiment. In contrast, both the wild-type and mutant circular structures from intron D preferentially formed closed monomer circles that were resistant to ExoIII digestion. In addition, the rate of cyclization of the wild-type sequence was 3.2 ± 0.4-fold faster than that of the mutant sequence, as revealed by four replicate gel experiments of the type shown in Fig. 6, and 2.7-fold faster, as seen from the ExoIII experiments shown in Fig. 7.

Additional evidence for the preferential cyclization of the intron segment was obtained by carrying out ligations in the presence of EcoRI-digested C. elegans genomic DNA (Fig. 8). In panel A, the $^{32}P$-labeled intron segment was mixed with a 1000-fold mass excess of C. elegans DNA prior to ligation. Gel analysis of the ligation products revealed cyclization of the intron segment persisted even in the presence of excess genomic DNA. A similar analysis in panel B showed that the
Fig. 6. **Analysis of the ligation products.** The panels show autoradiograms of the ligation products of the restriction fragments shown in Fig. 5. Ligation conditions are described under "Experimental Procedures." Samples were removed at the indicated times after addition of ligase and electrophoresed on a 2% agarose gel. The samples at time 0 were removed prior to the addition of the ligase, and those designated as E were treated with exonuclease III following 15 min of ligation. LT, linear trimer; LD, linear dimer; DC, dimer circle; LM, linear monomer; MC, monomer circle.

Fig. 7. **Cyclization kinetics of wild type and mutant intron segments.** Left panel, labeled wild-type (□) and mutant (■) intron segments from Fig. 5 were incubated with 5 units/ml ligase for the indicated times. Exo-resistant radioactivity was then measured following acid precipitation. Results are expressed as the means ± S.E. of the percentages of the values for DNA not treated with Exo. Right panel, data replotted as in Crothers et al. (1992) where (D) is the concentration of Exo-sensitive radioactivity (unreacted monomer) at time t. Exo-sensitive radioactivity is equated with unreacted monomer because autoradiograms of these samples following electrophoresis revealed >98% of the radioactivity was either unreacted or circular monomer (not shown). Cyclization rate constants $k_r$ were $5.9 \times 10^{-4}$ s$^{-1}$ for the wild-type and mutant sequences, respectively.

cyclization of the intron segment was not noticeably affected by 100-fold excess levels of genomic DNA fragments that ranged in size from 150 to 500 bp. Panel C shows the reverse approach where trace amounts of $^{32}$P-labeled genomic DNA were ligated in the presence of an excess of an EcoRI-digested plasmid containing the intron sequence. The preferential cyclization of the intron sequence when compared to the genomic DNA is also apparent from this analysis.

The torsional alignment of the two ends of a DNA molecule can influence the rate of cyclization (Shore and Baldwin, 1983a, 1983b). However, it is unlikely that this effect is responsible for marked differences in cyclization efficiencies reported in Figs. 6–8. For example, the wild-type and mutant intron segments displayed clear differences in cyclization rates as shown in Figs. 6 and 7, but both sequences have the same number of nucleotides and essentially the same number of helical turns as calculated from dinucleotide twist values (18.87 versus 18.85 turns). In addition, the failure of the control and satellite fragments to readily cyclize in the studies shown in Fig. 6 was expected since noncircular molecules in this size range resist cyclization, especially under the conditions of high DNA concentrations which favor bimolecular associations (Shore and Baldwin, 1981). To further rule out this possibility, ligation reactions were carried out in the presence of varying concentrations of ethidium bromide in order to alter the twist of control and circular molecules (Shore and Baldwin, 1983b). As shown in Fig. 9A, ethidium bromide had no noticeable effect on the preferential cyclization of the intron sequence when high DNA concentrations were used in the reaction mixtures. Likewise, cyclization of the intron sequence but not the control exon fragment was seen at all levels of the drug when the DNA concentrations were reduced by about 350-fold (Fig. 9B). We also note that ethidium bromide did not reduce the R$_{L}$ value of the intron segment when the fragment was electrophoresed in the presence of 0.3 and 0.7 $\mu$g of this drug/ml (data not shown). Cons and Fox (1990) have similarly noted that ethidium bromide failed to alter the anomalously slow gel mobility of K-DNA. Preferential cyclization of the intron sequence was also observed at ligation temperatures ranging from 5 °C to 37 °C in the absence of ethidium bromide (Fig. 9C). Temperature variations within this range, like ethidium, alter the helical twist of DNA molecules of this size (Shore and Baldwin, 1983b). Taken together, these results and the observations mentioned above strongly suggest that differences in the orientation of the ends of DNA molecules in Fig. 5 are not responsible for the preferential ligation of the intron sequence.

The studies in Figs. 10 and 11 focused on a large circular structure from phage $\lambda$. The 880-bp circular structure is found in the A gene which codes for the major subunit of the $\lambda$ DNA terminase. The enzyme cleaves the terminal cos site and also plays an active role in the packaging of DNA into proheads (Cue and Feiss, 1993). The $\lambda$ sequence was characterized by an
ENDS ratio peak of 18 which was the highest value observed for sequences in the phage subdivision of GenBank. Non-overlapping fragments from this region displayed a pronounced electrophoretic anomaly which was attributed to phased oligo(A) tracts that were found throughout this segment of the genome (Anderson, 1986). The sequence appears to belong to a set of relatively high ENDS ratio peaks which lie within the G+C-rich left arm of the genome. As noted in Fig. 10, major structural variations are seen near the beginning and end of the left arm which contain the head and tail structural genes, while minor variations are observed in the vicinity of the head-tail junction.

Non-bent molecules that are shorter than the DNA persistence length of about 150–200 bp resist cyclization in a ligation reaction (Shore et al., 1981). Consequently, the rate of cyclization of small DNA molecules with intrinsic curvature can be several orders of magnitude greater than the cyclization rate of a nonbent molecule of the same length (Koo et al., 1990; Crothers et al., 1992). The effects of sequence-dependent curvature on the cyclization of larger molecules should be more difficult to demonstrate since nonbent large fragments will readily cyclize during a ligation reaction (Shore et al., 1981). In order to investigate the cyclization of the predicted circular structure from \(\lambda\) shown in Fig. 10, the rate of cyclization of an EcoRI fragment that corresponds to the structure was compared to the rates of cyclization of two noncircular molecules of similar lengths. In Fig. 11, plasmids containing each of the sequences were digested with EcoRI and mixed together prior to ligation. The reaction was carried out using low DNA concentrations in order to favor cyclization of all DNA molecules. In an attempt to minimize random motions of the molecules, ligation reactions were carried out at 0–1 °C in the presence of 1.8 M sucrose. The results revealed that the cyclization of the \(\lambda\) segment (designated B) occurred faster than the cyclization of the control molecules A and C. This can be seen by the preferential loss of linear monomer in B and the selective increase in the corresponding circular monomer during the course of the reaction. Results similar to those shown in the figure were seen in all five experiments where this procedure was used, and in all seven experiments where the three plasmids were ligated in separate reactions and the products detected either by ethidium bromide staining or by blot hybridization analysis (data not shown). The preferential cyclization of the \(\lambda\) sequence also occurred using varying amounts of ligase (1–20 units/ml) but was not observed in the absence of sucrose in the ligation mixture (data not shown).

**DISCUSSION**

The standard computer program used in this work for predicting DNA structure from nucleotide sequence assumes that AA/TT steps are the major determinants of bending as has been proposed from numerous experimental findings (Crothers et al., 1990; Diekmann et al., 1992; Hagerman, 1992; Haran et al., 1994; Wang et al., 1994; Sprouss et al., 1995). A precondition for predicting structure by this program is that bending possibly caused by non AA/TT stacks should not significantly alter the interpretations of the analysis. This is apparently the case since the circular structure in the HIV-1 integrase gene (Fig. 1E), as well as other selected circular structures (data not shown), were also observed when the sequences were analyzed by a program that incorporated predicted wedge values for all dinucleotides. In addition, we have recently shown that programs which assume that AA/TT is the only source of bending were better predictors of electrophoretic data than a program based on all wedge values (Wang et al., 1994). Recent electrophoretic studies by Haran et al. (1994) have also stressed the relative unimportance of non AA/TT stacks in bending and critical discussions of non-A tract bending models for curvature are provided by Haran et al. (1994), Wang et al. (1994) and Sprouss et al. (1995). The standard program has previously been shown to reproduce the shapes of short DNA molecules seen in electron micrographs and to accurately predict the planarity of curvature as determined from electrophoretic studies (Eckdahl and Anderson, 1987; Fitzgerald et al., 1994; Wang et al., 1994) and from ligation analysis of small synthetic circular DNAs (data not shown). The results of this study provide additional support for the predictive power of the program and consequently for the model of A tract bending since the circular intron segment from \(C.\) elegans preferentially cyclized in ligation reactions when compared to nonbent fragments, bent fragments with noncircular shapes, and total genomic DNA (Figs. 5–8). In addition, the wild-type intron sequence cyclized significantly faster than a mutant sequence with a single base substitution and the difference was reflected in differences in electrophoretic anomaly, ENDS ratio values, and helical projections (Figs. 5 and 6). Ring closure probabilities could not be accurately determined for any of the molecules shown in Fig. 5.

---

3These are defined as the ratio of the equilibrium constants for cyclization to that for biomolecular association (Jacobsson and Stockmayer, 1950).
because the intron segments displayed very low rates of bimolecular association, which could not be quantified reliably, while the noncircular fragments did not cyclize or cyclized very slowly under all reaction conditions tested (Figs. 6, 7, and 9; data not shown). However, relative cyclization efficiencies could be estimated by making assumptions that have made by others who have encountered similar problems with short synthetic DNA molecules (Crothers et al., 1992). We estimate that the relative cyclization efficiency of the wild-type intron segment was about 3-fold greater than the mutant intron segment and more than 100-fold greater than all of the noncircular molecules shown in Fig. 5. These estimates were not affected by variation in DNA concentration, ligase concentration, ethidium bromide, or temperature (Figs. 6–9; data not shown). The sequence encoding the phage λ terminase also cyclized significantly faster than noncircular molecules of similar lengths and this sequence displays predicted planar bending throughout its 880-bp length (Figs. 10 and 11). Additional studies along these lines are needed in order to further characterize the shapes and stabilities of such large structures in solution. The identification of an ensemble of sequences with circular helical projections as in Figs. 1–4 provides a rich source of natural molecules that could be used for such an analysis.

DNA inversion systems in prokaryotes have served as models for the study of site-specific recombination (Glasgow et al., 1988; Heichman and Johnson, 1990; Stark et al., 1992; Van de Putte and Goosen, 1992; Merker et al., 1993). Typically, a recombinase binds as a dimer to each of two inverted repeat sequences and dimers of Fis interact with two sites within a transpositional enhancer located nearby. Protein-induced DNA bending and Fis-recombinase interactions then facilitate the looping of intervening DNA to form a large multi-protein com-

![Fig. 10. Predicted structures along the phage lambda genome. ENDS ratios computed at window widths from 20 to 1000 bp along the entire genome are plotted in C as a function of sequence position in kilobase pairs. The positions of the genes encoding head proteins, tail proteins and the major transition in genome base composition (↑) are indicated. Two-dimensional projections of the helix axis of the first 22 kbp are shown in B. The break in the projection at bp 7963 lies at the head-tail junction. The position of the circular structure (C) at bp 870-1750 is indicated. Projections of this 880-bp circular structure are shown in A.](image)

![Fig. 11. Analysis of ligation products of long DNA molecules. EcoRI digests of plasmids containing the 1192-bp α-prothymosin gene (A), the 880-bp segment from phage λ shown at the top of Fig. 10 (B), and the 700-bp troponin I gene (C) were incubated with T4 DNA ligase for the indicated times. Samples were electrophoresed on a 1.3% agarose gel in the presence of ethidium bromide, and a photographic negative of the gel is shown on the left. An enlargement of the bottom portion of the negative is shown in the center of the figure. L, linear monomer; OC, open circular monomer; CC, closed circular monomer. The right panel shows the ratio of circular (OC + CC) to linear monomers as a function of ligation time for A (○), B (●), and C (×).](image)
plex called the invertasome. Accessory DNA-bending proteins such as the histone-like proteins HU and IHF also aid in particle formation by facilitating the folding of the DNA (SURETT et al., 1989; HARRINGTON, 1992). Thus, the ability of DNA to loop in the invertasome should not only depend on the position and orientation of specific protein binding sites but also on the physical properties of the intervening DNA including its flexibility and intrinsic curvature. Support for this idea has been derived from the observation that intrinsically bent DNA can substitute for protein-induced bending in many systems including inversion reactions (BRAVO et al., 1989; GOODMAN AND NASH, 1989). Consequently, the short segments of bent DNA found in recombinational enhancers have been viewed as factors that facilitate protein-induced bending and thus the formation of the invertasome (JONSON et al., 1987; GLASGOW et al., 1988; HÜBLER et al., 1989). The preferential association of high ENDS ratio peaks with a variety of inversion and long-range looping systems (Fig. 4, A–H) makes it likely that the large circular structures play an analogous but perhaps more global role in particle formation. This model could also be applied to the sequences in Fig. 4 (I–N), which are thought to display regional recombinational specificity. In these systems, the large circular structures may direct recombinational complexes to the regions where strand cleavage and rejoining take place.

The circular structures described in this report may play some role in recombination in addition to a general packaging function. A consideration of such a role is particularly relevant since the requirement for a highly ordered DNA topology in the invertasome has led to the view that the DNA molecules are active participants in recombination rather than merely passive substrates for protein action. An important characteristic of inversion systems is the requirement of a negatively supercoiled DNA substrate that facilitates protein-protein interaction and the formation of the invertasome. Restriction nuclease studies and electron microscopy have revealed that bent DNA is positioned within end loops of supercoiled molecules (SILVER et al., 1986; LAUNDON AND GRIFFITH, 1988). Thus, bent DNA can order the global structure of superhelical DNA (SILVER et al., 1986; LAUNDON AND GRIFFITH, 1988; KREMER et al., 1993). It seems likely that this organizing effect would depend on the configuration of the bent segment since O- and C-shaped structures should preferentially reside at the ends of intertwined molecules while bent structures such as S-shaped regions should be preferentially excluded from these sites. These considerations imply that the circular DNA shapes have the potential to align DNA sites both within and adjacent to protein complexes on domains of supercoiled DNA and this could be important in controlling DNA topology during recombination and other long range looping events.

The circular structures in the primate immunodeficiency viruses were detected as the highest ENDS ratio sequence set in the entire GenBank data base (Figs. 1–3). This structure is not found in all retroviruses, but is apparently more conserved than sequence (see Figs. 1 and 2 and “Results”), which implies that selective pressures have maintained the circular form. To our knowledge, this region is not a hot spot for recombination, although for the role of a structure in this process cannot be excluded. Perhaps more likely, the structure functions in a DNA looping mechanism which serves to bring together the protein binding sites located in the two distinct subdomains of the intragenic transcriptional enhancer (see Fig. 1). According to this view, the structure may have a packaging function analogous to the circular elements in Fig. 4 (A, D, E, G, and H). The presence of multiple functional AP-1 recognition sequences in the upstream enhancer subdomain of HIV is of interest since these sequences are frequently components of complex regulatory elements, which contain binding sites for multiple transcription factors (SOMMERBERG et al., 1989; KERPAPOULOS AND CURTIS, 1991). Several potentially binding sites for transcription factors other than AP-1 are indeed found in this region of the HIV genome (VAN LINT et al., 1991). Thus, the DNase I-hypersensitive site that resides at this location in HIV-1 chromatin (VERDIN, 1991) may be a reflection of a complex regulatory particle assembled onto the circular DNA segment. The DNA structure is expected to be regulated since unintegrated linear viral DNA has a single-stranded gap at this site, which would prevent circle formation prior to integration. In addition, the enhancer and the nuclease-hypersensitive site are regulated by cellular factors following the integration of the HIV genome into host cell chromatin (VERDIN et al., 1990; VANLINT et al., 1991; VERDIN, 1991). Although the function of this highly unusual conformation in regulating the HIV life cycle is not yet known, the structure may provide a model system for the study of a regulatable looping mechanism that is dependent on the intrinsic circularity of the DNA.

Acknowledgments—We thank Dr. David Bird, University of California, Riverside, California, for plasmid DR36, Dr. B. J. Taparowsky for Plasmid PSV2 cat, and Sonia Nasr for the plasmid containing the satellite monomer of T. molitor. We also thank Drs. D. Asai, B. Wanner, I. TESSMAN, S. NASR and P. T. GILHAM for reviewing the manuscript, J. O. SHERIDAN for assistance with the manuscript, and Dr. P. T. GILHAM for the idea of using sucrose in ligation reactions shown in Fig. 11.

REFERENCES
ABRAHAM, J. M., FREITAG, C. S., CLEMENTS, J. R. & EISENSTEIN, B. I. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 5724–5727.
ALMENDERAL, J. M., ALMAZÁN, F., BLASCO, R. & VINUELA, E. (1990) J. Virol. 64, 2064–2072.
ANDERSON, J. N. (1986) Nucleic Acids Res. 14, 8531–8533.
BARBOUR, A. G. (1993) Trends Microbiol. 1, 236–239.
BARBOUR, A. G., CARTER, C. J., BURMAN, N., FREITA, C. S., GARAON, C. F. & BERGSTROM, S. (1991) Infect. Immun. 59, 390–397.
BENSON, D. A., BAGUSKI, M., LIPMAN, D. J. & OSTELL, J. (1994) Nucleic Acids Res. 22, 3441–3444.
BIRD, D. M. & RIDDELL, D. L. (1989) Mol. Cell. Biol. 9, 4119–4130.
BOLHOY, A., MCNAMARA, P., HARRINGTON, R. E. & TRIFONOV, E. N. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 2312–2316.
BOOTHRUDY, J. C. & CROSS, G. A. (1982) Gene (Amst.) 20, 281–289.
BRAUN, L., KOTLARZ, D., KÖBL, A., DIEKMANN, S. & BUCH, H. (1989) EMBO J. 8, 4289–4296.
BROOKS, J. C. & S. (1985) Gene, 36, 506–532.
CHARNEAU, P. & CLAVEL, F. (1991) J. Virol. 65, 2415–2421.
CHARNEAU, P., ALIZON, M. & CLAVEL, F. (1992) J. Virol. 66, 2814–2820.
COFFMAN, R. L., LEBMAN, D. A. & ROSENTHAL, P. (1993) Adv. Immunol. 54, 229–269.
CONS, B. M. G. & FOX, K. R. (1990) Biochem. Biophys. Res. Commun. 171, 1064–1070.
COW, M. M. (1989) in Mobile DNA (BERG, D. E. & HOWE, M. M. eds), pp. 661–670.
AMERICAN SOCIETY FOR MICROBIOLOGY, WASHINGTON, D. C.
CROTHERS, D. M., HARRAN, T. E. & NADEAU, J. G. (1990) J. Biol. Chem. 265, 7093–7096.
CROTHERS, D. M., DRAK, J., KAHN, J. D. & LEEVEN, S. D. (1992) Methods Enzymol. 212, 3–29.
DUE, D. & FEISS, M. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 9290–9294.
DE CHASTONAY, Y., MÜLLER, F. & TOBLER, H. (1990) Gene 93, 199–204.
DE LA VEGA, I., VINUELA, E. & BLASCO, R. (1990) Virol. 175, 234–246.
DIEKMANN, S. & WANG, J. C. (1985) J. Biol. Chem. 260, 1–11.
DODDLE, R. F., FENG, D. F., JOHNSON, M. S. & MCCURLE, M. A. (1989) Rev. Biol. 64, 1–30.
DRAK, J. & CROTHERS, D. M. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 3074–3078.
DREW, H. & TRAVERS, A. A. (1985) J. Biol. Chem. 260, 773–790.
ECHOLS, H. (1990) J. Biol. Chem. 265, 14697–14700.
EDDADHI, T. T. & ANDERSON, J. N. (1987) Nucleic Acids Res. 15, 8531–8545.
EDDADHI, T. T. & ANDERSON, J. N. (1989) Nucleic Acids Res. 18, 1609–1612.
FITZGERALD, D. J., DRYDEN, G. L., BRONSON, E. C., WILLIAMS, J. S. & ANDERSON, J. N. (1994) J. Biol. Chem. 269, 21303–21314.
FULKERSON, A. M., MARRS, C. S., STEVENS, S. P. & GREEN, R. M. (1990) J. Bacteriol. 172, 1030–1035.
GLASGOW, A. C., HUGHES, K. T. & SIMON, M. I. (1989) in Mobile DNA (BERG, D. E. & HOWE, M. M. eds), pp. 637–660.
AMERICAN SOCIETY FOR MICROBIOLOGY, WASHINGTON, D. C.
GHAGHERMAN, P. J. (1992) Biochim. Biophys. Acta 1131, 125–132.
