JOURNAL OF VIROLOGY, May 1993, p. 2707–2715
0022-538X/93/052707-09$02.00/0

Vol. 67, No. 5

Easily Unwound DNA Sequences and Hairpin Structures in the Epstein-Barr Virus Origin of Plasmid Replication

D. LYNN WILLIAMS AND DAVID KOWALSKI*

Molecular and Cellular Biology Department, Roswell Park Cancer Institute, Buffalo, New York 14263

Received 20 November 1992/Accepted 12 February 1993

The Epstein-Barr virus (EBV) origin of plasmid replication (*oriP*) includes two known *cis*-acting components, the dyad symmetry region and the family of repeats. We used P1 nuclease, a single-strand-specific endonuclease, to probe EBV *oriP* for DNA sequences that are intrinsically easy to unwind on a negatively supercoiled plasmid. Selective nuclease hypersensitivity was detected in the family of repeats on an *oriP*-containing plasmid and in the dyad symmetry region on a plasmid that lacks the family of repeats, indicating that the DNA in both *cis*-acting components is intrinsically easy to unwind. The hierarchy of nuclease hypersensitivity indicates that the family of repeats is more easily unwound than the dyad symmetry region, consistent with the hierarchy of helical stability predicted by computer analysis of the DNA sequence. A specific subset of the family of repeats is nuclease hypersensitive, and the DNA structure deduced from nucleotide-level analysis of the P1 nuclease nicks is a cruciform near a single-stranded bubble. The dyad symmetry region unwinds to form a broad single-stranded bubble containing hairpins in the 65-bp dyad sequence. We propose that the intrinsic ease of unwinding the dyad symmetry region, the actual origin of DNA replication, is an important component in the mechanism of initiation.

Infection of human B lymphocytes with Epstein-Barr virus (EBV), a human herpesvirus, results in cell immortalization (14). The EBV genome is a 172-kb circular DNA that exists as a multicopy plasmid. Several viral proteins are expressed during the latent period of infection and one of these, EBV nuclear antigen 1 (EBNA1), is required for replication of the viral genome (22, 40). A *cis*-acting sequence has been identified that allows replication and maintenance of plasmids in cells that express EBNA1 in *trans* (39). This sequence is the origin of plasmid replication (*oriP*) and is contained within a 1.8-kb fragment of the EBV genome.

oriP contains two known *cis*-acting elements, a family of 30-bp tandem repeats and a dyad symmetry region, which function in replication of the EBV genome (30). Changes in the spacing or orientation of these two elements has little effect on the activity of *oriP*. The family of repeats is required for efficient replication but can be functionally replaced by multiple tandem copies of the dyad symmetry region (38). Physical mapping has revealed that the dyad symmetry region is the actual origin of replication, since it colocalizes with the region where replication initiates within *oriP* (10). The family of repeats acts as a replication fork barrier and is the primary termination site for *oriP*-mediated replication (10). In addition to its roles in DNA replication, the family of repeats can enhance transcription from heterologous promoters in a strictly EBNA1-dependent fashion (29).

Both the dyad symmetry region and the family of repeats contain multiple EBNA1-binding sites (28). The dyad symmetry region contains four copies of an EBNA1-binding sequence, two within a 65-bp dyad sequence and two flanking the dyad. The family of repeats consists of 20 imperfect copies of a 30-bp sequence in tandem (20× 30-bp repeats), and each copy contains an EBNA1-binding sequence. EBNA1 binding promotes association between the dyad symmetry region and the family of repeats in vitro (24), creating a DNA loop (8, 34). It has been suggested that the EBNA1-mediated association between these distal *cis*-acting elements promotes the initiation of replication (8, 24, 34).

Initiation of replication requires unwinding of the DNA double helix. Duplex unwinding allows the replication machinery to access the individual DNA strands. A general model for the initiation of bidirectional replication is that an initiator protein recognizes specific sites at the origin and induces localized unwinding of a particular sequence. Such a mechanism for localized origin unwinding has been demonstrated for simian virus 40 (SV40) (1), *Escherichia coli* (3), and phage λ (31) and has been proposed for *Saccharomyces cerevisiae* (35, 37). Chromosomal replication origins from *E. coli* and *S. cerevisiae* are intrinsically easy to unwind. Unwinding of these origins can occur in the absence of an initiator protein in negatively supercoiled DNA, as revealed by hypersensitivity to single-strand-specific nucleases (20, 25, 35). At the *E. coli* origin, the nuclease-hypersensitive sequence identified in naked supercoiled DNA (20) corresponds to the same sequence induced to unwind by the initiator protein (3). Mutational analysis has shown that the intrinsic ease of unwinding at the nuclease-hypersensitive element is required for origin function in *E. coli* and *S. cerevisiae* (20, 25, 35, 37). This novel, *cis*-acting component is termed a DNA-unwinding element (DUE) (20, 37). DUEs can be identified by hypersensitivity to single-strand-specific nuclease in supercoiled DNA (20, 35) and by a recently developed computer program that calculates helical stability from DNA sequence information (25). In the present study, P1 nuclease and helical stability analysis were used to probe for easily unwound sequences in the EBV origin of plasmid replication *oriP*.

MATERIALS AND METHODS

Enzymes. P1 nuclease, prepared by the method of Fujimoto et al. (9), was from Yamasa Shoyu (Choshi, Japan).

* Corresponding author. Electronic mail address: kowalski@vax2.med.buffalo.edu

Enzymes from commercial suppliers were as follows: restriction enzymes and T4 DNA ligase (New England Biolabs); T4 polynucleotide kinase (United States Biochemical Corp.); calf intestinal phosphatase (CIP) (Boehringer Mannheim Biochemicals).

DNA. Plasmids pWEoriP and pWE-DS were grown in *E. coli* HB101 in Luria-Bertani broth and amplified with 150 μg of chloramphenicol per ml. DNA was obtained from cells lysed by being boiled in the presence of lysozyme (15). Using this method, the superhelical density is the same for different plasmids obtained from *E. coli* HB101 and is equal to −0.065 ± 0.002 (25) as determined by two-dimensional gel electrophoresis of plasmid topoisomers (21). Negatively supercoiled DNA was isolated and purified by two rounds of equilibrium gradient centrifugation in a cesium chloride solution containing ethidium bromide (27). The location of P1 nuclease-hypersensitive sites (see below) was consistent among independent DNA preparations of the same plasmid.

Construction of pWEoriP and pWE-DS plasmids. pWEoriP contains the family of 30-bp repeats and the dyad symmetry region from the EBV origin of plasmid replication (*oriP*) on a fragment of approximately 2,200 bp inserted into pBT3A, a modified pBR322 vector (20). pWEoriP was constructed as follows. *oriP* was excised from p395 (38; a gift from John Yates, Roswell Park Cancer Institute) by using *Eco*RI and *Bam*HI and subcloned into the *Eco*RI and *Bam*HI sites of pUC12. Next, *oriP* was removed from pUC12 by using endonucleases *Eco*RI and *Pst*I and subcloned into the *Eco*RI and *Pst*I sites of pBT3A. pWE-DS contains only the dyad symmetry region of *oriP* inserted into pBT3A. The dyad symmetry region was excised from p393.2 (38; a gift from John Yates) by endonucleases *Eco*RI and *Pst*I. This fragment was then subcloned into the *Eco*RI and *Pst*I sites of pBT3A.

Single-strand-specific nuclease reactions. P1 nuclease reactions contained 10 mM Tris-HCl (pH 7.0)–1 mM disodium EDTA–1.6 μg of negatively supercoiled DNA in a volume of 18 μl. After preincubation for 10 min at 37°C, P1 nuclease (2.4 ng in 2 μl; 600 U/mg) was added. The reaction mixture was kept at 37°C for 30 min. Under these conditions, supercoiled plasmids are converted to singly nicked circular DNA. The nuclease reaction was quenched as previously described (19), and the enzyme was removed by phenol extraction.

Localization of P1 nuclease-specific nicks. Nicked circular DNA produced by the action of P1 nuclease on negatively supercoiled plasmids was linearized at a unique restriction endonuclease site, dephosphorylated with CIP, and 5′ end labeled with [γ-^{32}P]ATP and polynucleotide kinase (32). The DNA was separated by electrophoresis on a 1.2% agarose gel after irreversible denaturation with glyoxal, visualized by autoradiography, and sized as previously described (35).

Fine mapping and nucleotide sequence of sites nicked by P1 nuclease. For fine mapping of nicks, pWEoriP was treated with P1 nuclease, linearized with *Eco*RI, dephosphorylated with CIP, and then 5′ end labeled with [γ-^{32}P]ATP and polynucleotide kinase (32). End-labeled DNA was denatured, and the products were separated by electrophoresis through an 8% polyacrylamide gel. To map P1 nuclease nicks at single nucleotide resolution, DNA termini generated by specific restriction enzymes were dephosphorylated with CIP and labeled at 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase (32) or at 3′ ends with [α-^{32}P]dATP (*Eco*RI site) or [α-^{32}P]dCTP (*Nci*I site) and DNA polymerase (Klenow fragment).

Computer analysis. The DNA sequences of pWEoriP and pWE-DS were assembled on a Compaq computer by using the IBI-Pustell Sequence Analysis Programs (International Biotechnologies, Inc.). The free energy required for DNA strand separation under the conditions used for the P1 nuclease assay was calculated with Thermodyn, a recently developed computer program (25). A 100-bp window size was used in calculating the free energy across the DNA sequence in 4-bp steps. Slidewrite Plus (Advanced Graphics Software, Inc.) was used to convert the output of the Thermodyn program into graphic form. A VAX computer program called StemLoop (Genetics Computer Group, Inc.) was used to search for potential hairpins in the *oriP* sequence. The parameters used were minimum stem length, 4; minimum bonds per stem, 8 (AT, 2; GC, 3; GT, 1; and other mismatches, 0); maximum loop size, 20; and minimum loop size, 3.

RESULTS

Single-strand-specific nuclease-hypersensitive sites in EBV *oriP*. P1 nuclease is a single-strand-specific endonuclease that acts on negatively supercoiled DNA at neutral pH (19). Each supercoiled molecule is nicked only once because the

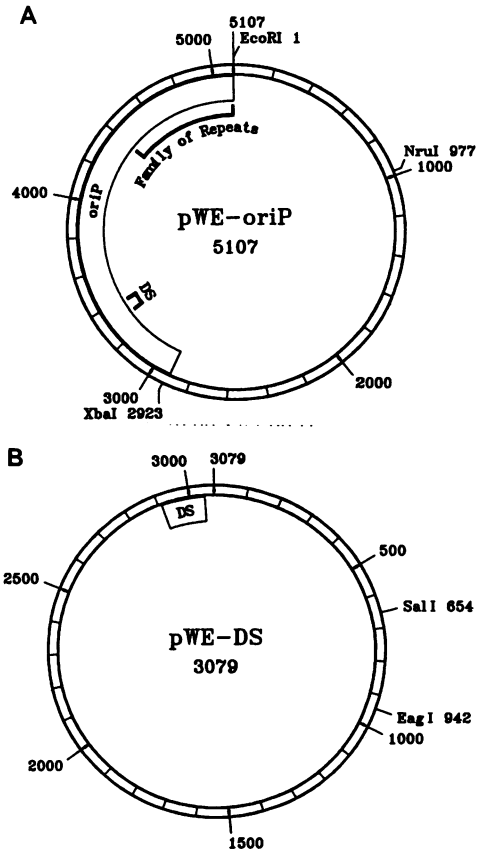


FIG. 1. Plasmid maps of pWEoriP (A) and pWE-DS (B). pWEoriP and pWE-DS are pBR322-derived plasmids containing a complete *oriP* fragment and a fragment containing only the dyad symmetry (DS) region, respectively. The *oriP* insert, the DS region, the family of repeats, and relevant restriction endonuclease sites are indicated.

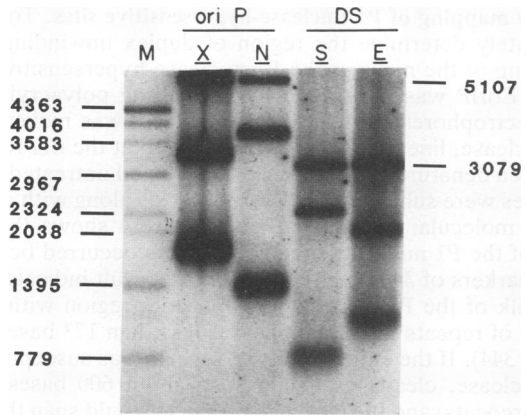


FIG. 2. Single-strand-specific nuclease-hypersensitive sites in pWEoriP and pWE-DS. P1 nuclease-treated pWEoriP and pWE-DS were cut and 5′ ^{32}P-end labeled at a single restriction endonuclease site before irreversible denaturation with glyoxal and agarose gel electrophoresis. Selected marker (lane M) sizes (in nucleotides) are indicated to the left. Sizes of the unit-length plasmids are indicated on the right. The bands that result from two different restriction digests of each plasmid are shown. Restriction endonucleases (lanes): X, *Xba*I; N, *Nru*I; S, *Sal*I; E, *Eag*I. The sizes (in bases) of the DNA single-stranded fragments produced after P1 nicking are as follows for each specific restriction digest: X, 3,300 and 1,800; N, 3,730 and 1,380; S, 2,330 and 830; E, 2,030 and 1,080. The P1 nuclease nicks map in the family of repeats in pWEoriP (position 4720 ± 50 bp) and in the dyad symmetry region in pWE-DS (position 2950 ± 50 bp).

first nick relaxes the DNA and relaxed DNA is a poor substrate for a single-strand-specific nuclease. The location and number of single-strand-specific nuclease cleavage sites depend on temperature, cation type and concentration, and level of negative supercoiling (16, 19, 32, 33). At the level of supercoiling and in the conditions used in this study, the site that is hypersensitive to a single-strand-specific nuclease identifies the most easily unwound DNA sequence in the plasmid (21).

Negatively supercoiled plasmids pWEoriP and pWE-DS were probed with P1 nuclease to determine the DNA sites that are intrinsically easy to unwind. pWEoriP is a pBR322 derivative containing a 2,200-bp insert which includes the EBV *oriP*, i.e., both the family of repeats and the dyad symmetry region (Fig. 1A). pWE-DS contains the dyad symmetry region of *oriP* (Fig. 1B). Supercoiled plasmids were quantitatively converted to nicked circular DNA by P1 nuclease. Nicked molecules were then linearized at a unique restriction site and 5′ end labeled with ^{32}P. After irreversible denaturation in glyoxal, the DNA products were separated by size, using agarose gel electrophoresis. Two subunit-length bands of equal intensity will be generated if P1 nuclease nicks a specific site at equal frequency in either DNA strand. The sizes of the two subunit-length bands indicate the distance, in bases, from the restriction site to the nicks introduced by P1 nuclease. A unit-length band is derived from the intact strand of the singly nicked circular DNA. A single pair of subunit-length bands was observed in pWEoriP after cutting with *Xba*I or *Nru*I (Fig. 2, lanes X and N). Analysis of the fragment sizes (legend to Fig. 2) assigned a P1 nuclease site to the region containing the family of 30-bp repeats within pWEoriP. P1 nuclease treatment of pWE-DS and subsequent digestion with *Sal*I or *Eag*I also showed a single band pair (Fig. 2, lanes S and E). The nuclease-hypersensitive site in pWE-DS mapped to within the dyad symmetry region. Detection of single-strand-specific nuclease nicks required negatively supercoiled DNA under our conditions (19, 33) and linearized DNA containing *oriP* was not nicked (data not shown). These results indicate that both the family of repeats and dyad symmetry regions contain DNA sequences that are intrinsically easy to unwind in supercoiled DNA.

P1 nuclease-hypersensitive sites map in DNA regions with low helical stability. The ease of unwinding can be quantified by determining the DNA helical stability, the free energy difference between the double- and single-stranded states. Helical stability can be reliably calculated from the DNA sequence by using the known thermodynamic properties of the component nearest-neighbor dinucleotides (4). We determined the DNA helical stability under the P1 nuclease digestion conditions with a recently developed computer program (25). The program calculates the helical stability of a fixed length of DNA (window) and can scan all possible nucleotide positions by sliding the window across the entire

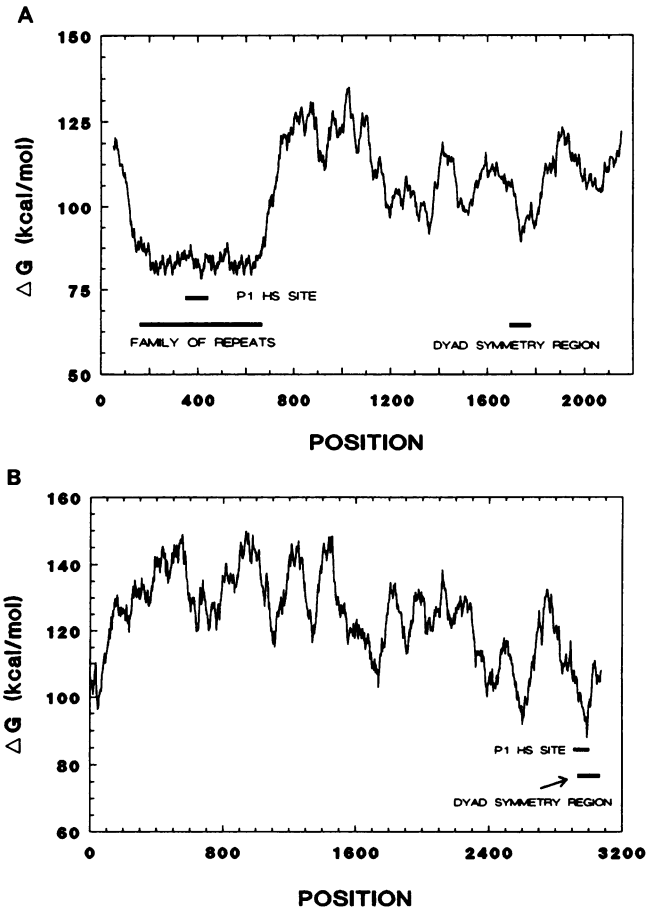


FIG. 3. Analysis of DNA helical stability. (A) EBV *oriP*. (B) pWE-DS. The helical stability (ΔG) under P1 nuclease digestion conditions was determined by computer analysis of the DNA sequence (25). The ΔG values reported correspond to the central position of a 100-bp window. The regions hypersensitive to P1 nuclease (P1 HS site, average cleavage position ± 50-bp range) are at *oriP* position 390 ± 50 bp and at pWE-DS position 2950 ± 50 bp. The positions of the family of repeats and the dyad symmetry region are also shown.

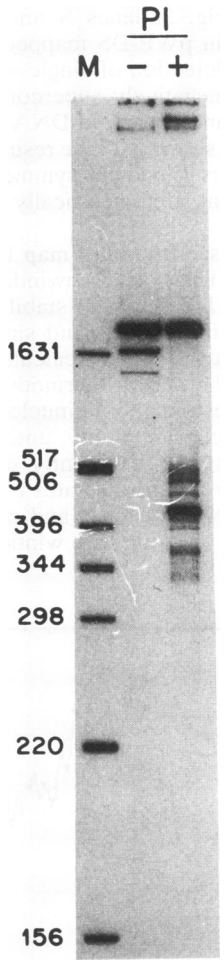


FIG. 4. Fine mapping of P1 nuclease-hypersensitive site in the family of repeats. pWEoriP was nicked with P1 nuclease and then linearized and 5′ end labeled at the *Eco*RI site. The DNA products were then denatured and electrophoresed on an 8% polyacrylamide gel. Lane M contains single-stranded DNA size markers. P1 nuclease-specific products are seen in lane P1 + between size markers 344 and 517. Larger products are not the result of P1 nuclease nicking because they are also present in DNA not treated with P1 nuclease (lane P1 −).

DNA sequence. A helical stability plot (ΔG versus nucleotide position) for *oriP* (Fig. 3A) showed that the region with the lowest free energy is between positions 200 and 650. This region corresponds to the family of repeats. The region of *oriP* that is hypersensitive to P1 nuclease (position 390 ± 50 bp) is also indicated (Fig. 3A). The analysis showed that the P1 nuclease-hypersensitive site occurs in a region of *oriP* that has a low helical stability. The analysis also showed that the dyad symmetry region, which is not cleaved by P1 nuclease in the *oriP*-containing plasmid, has a higher helical stability than the family of repeats (Fig. 3A).

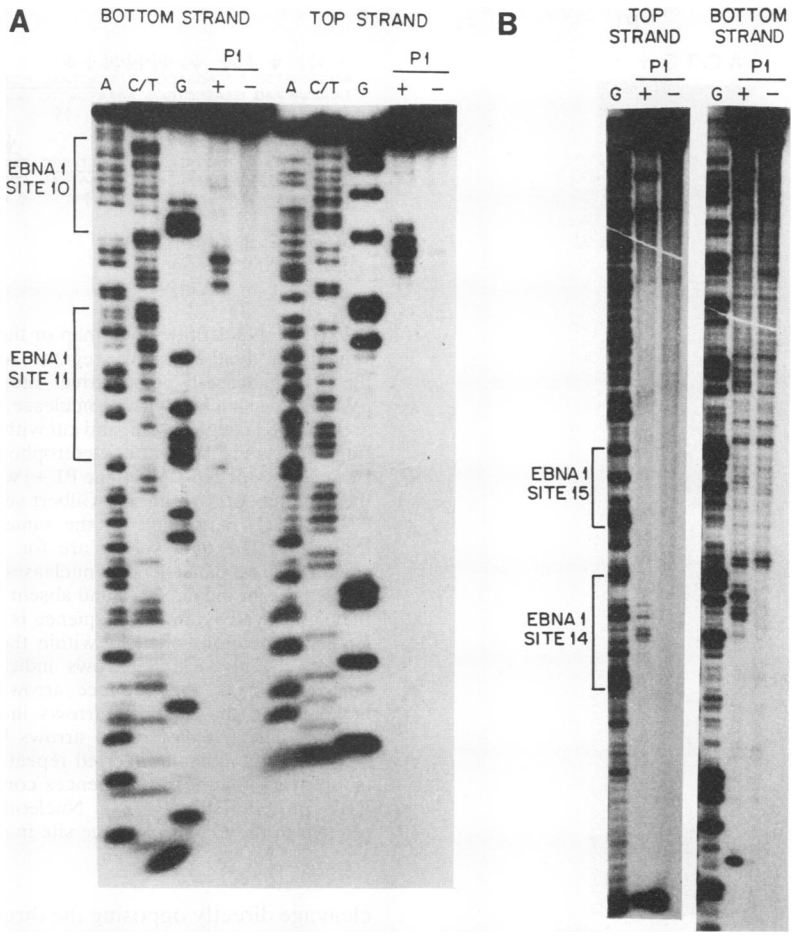
Figure 3B shows a helical stability plot for pWE-DS, the plasmid that contains the dyad symmetry region but not the family of repeats. The free energy minimum resides at position 2989 which, for a 100-bp window, includes the DNA sequence in the region from 2940 and 3039. This region colocalizes with the dyad symmetry region and the region that is hypersensitive to P1 nuclease (position 2950 ± 50 bp) (Fig. 3B).

Fine mapping of P1 nuclease-hypersensitive sites. To more accurately determine the region of duplex unwinding, fine mapping of the nicks at the P1 nuclease-hypersensitive site in pWEoriP was performed by denaturing polyacrylamide gel electrophoresis. Supercoiled pWEoriP was nicked with P1 nuclease, linearized and 5′ end labeled at the *Eco*RI site, and then denatured. P1 nuclease-treated and untreated DNA samples were subjected to electrophoresis along with appropriate molecular weight markers. Figure 4 shows that the bulk of the P1 nuclease-specific products occurred between size markers of 344 and 517 bases. This result indicates that the bulk of the P1 nuclease nicks span a region within the family of repeats that encompasses less than 173 bases (517 minus 344). If the entire family of repeats was susceptible to P1 nuclease, cleavages would span about 600 bases (20× 30-bp repeats) and the cleavage products would span the size markers from <156 to >517 bases. Thus, it is clear that only a specific portion of the family of repeats is sensitive to cutting by P1 nuclease.

***oriP* sequences cleaved by P1 nuclease.** The precise locations of the P1 nuclease cleavages within pWEoriP and pWE-DS sequences were determined by analysis at single-nucleotide resolution. Singly end-labeled restriction fragments containing the nuclease nicks were prepared and denatured. The single-stranded products were separated by electrophoresis on DNA sequencing gels alongside the products of Maxam and Gilbert (23) sequencing reactions performed on the same restriction fragments without nuclease nicks.

Figure 5A and B shows autoradiograms of sequence-level analysis of the P1 nuclease-specific nicks within pWEoriP. Brackets indicate the positions of the EBNA1-binding sites in the 20× 30-bp repeats within *oriP*. Both DNA strands were examined by labeling a single restriction site at the 5′ end and, in an independent experiment, at the 3′ end of the complementary strand. Both top and bottom strands show a similar pattern of cleavage by P1 nuclease (P1+ lanes). Major cleavages occur between EBNA1-binding sites in repeats 10 and 11 (Fig. 5A) and within the EBNA1-binding site of repeat 14 (Fig. 5B). In addition to these major cleavages, minor cleavages are detected, especially in repeats 13, 14, 15, and 16 upon longer exposure of the autoradiogram (data not shown). No P1 nuclease-specific cleavages were detected in the *Nci*I-*Bst*XI restriction fragment containing repeats 1 to 9 (data not shown). Figure 5C displays the locations of the P1 nuclease nicks within the hypersensitive region of the *oriP* DNA sequence. The nuclease-hypersensitive sequence is localized to portions of six of the 20× 30-bp repeats. The pattern of P1 nuclease nicks does not simply reflect either the repeated 30-bp sequence or the components of that sequence. For example, the sequence between the boxes containing EBNA1 sites 10 and 11 is strongly cleaved (boldface arrows), whereas similar sequences between other EBNA1 sites are cleaved weakly or not at all. Also, the boxed EBNA1 site 14 is a major cleavage site (boldface arrows), whereas other EBNA1 sites with similar or identical sequences are cleaved weakly or not at all.

Figure 6A shows an autoradiogram of sequence-level analysis of P1 nuclease nicks in the dyad symmetry region within pWE-DS. Major cleavages occur between the inverted repeat sequences and in the flanking sequences. Similar results were seen for the bottom strand (data not shown). Figure 6B displays the locations of the specific nuclease nicks within the sequence of the dyad symmetry region. The nuclease-hypersensitive sequence spans 115 bp



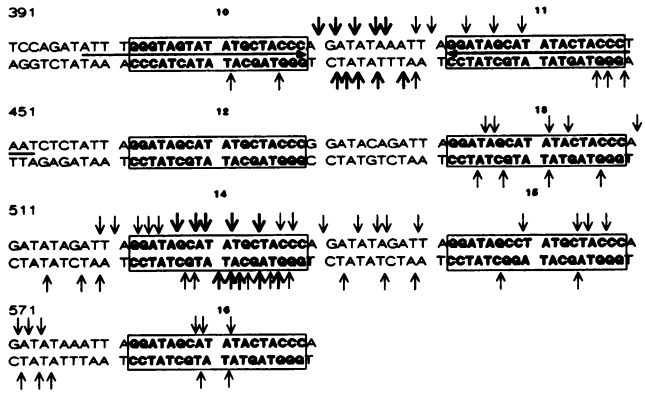
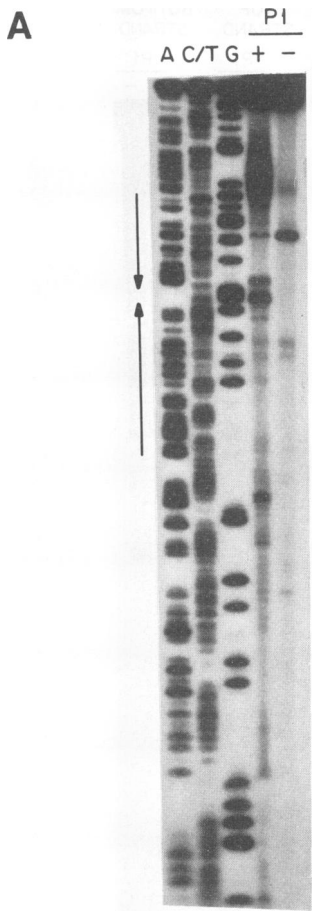


FIG. 5. Nucleotide-level map of the P1 nuclease-hypersensitive site within the family of repeats of *oriP*. (A and B) DNA sequence analysis of the P1 nuclease-hypersensitive sites in pWEoriP. P1 nuclease-nicked DNA (lane P1 +) was electrophoresed alongside the products of Maxam and Gilbert sequencing reactions (lanes A, C/T, and G) performed on the same DNA without P1 nuclease treatment. Specific restriction fragments of the DNA were singly end labeled at an *Nci*I cleavage site on the 3′ end of one strand or the 5′ end of the complementary strand. The bracketed regions on the left indicate the positions of EBNA1-binding sites within the *oriP* sequence (28). P1 nuclease-specific nicks correspond to bands present in the P1 + lane and absent in the P1 − lane. Bands present in both the + and − lanes are due to nonspecific nicking. Sequence-level analysis of P1 nuclease nicks in the 79-bp *Bst*XI-to-*Nci*I fragment is shown in panel A. Sequence-level analysis of P1 nuclease nicks in the 376-bp *Nci*I-to-*Bsp*MI fragment is shown in panel B. Supercoiled pWEoriP was nicked with P1 nuclease and then cut with restriction enzyme *Nci*I. Specific DNA fragments containing the nicks (396 and 326 bp) were isolated after separation by polyacrylamide gel electrophoresis. An aliquot of the isolated DNA fragments was ^{32}P-labeled at the 5′ end of the *Nci*I cleavage sites, and a separate aliquot was ^{32}P-labeled at the 3′ end of the *Nci*I cleavage sites. The labeled 396-bp fragment was cut with *Bsp*MI and a singly end-labeled product of 376 bp was isolated after polyacrylamide gel electrophoresis. The labeled 326-bp fragment was cut with *Bst*XI, and singly end-labeled products of 247 and 79 bp were isolated after polyacrylamide gel electrophoresis. No P1 nuclease-specific nicks were detected in the 247-bp fragment (not shown). (C) Locations of nicks within the nuclease-hypersensitive region. The vertical arrows indicate the locations of the P1 nicks. The boldface arrows represent sites of frequent nicking, and the lightface arrows represent sites of infrequent nicking. Boxes represent the EBNA1-binding sites (28). Boldface numbers above each box indicate the repeat number within the 20× 30-bp repeat region. Horizontal arrows indicate an inverted repeat sequence. Nucleotide positions are numbered relative to the *Eco*RI cleavage site in *oriP*. *oriP* resides in a 2,204-bp *Eco*RI-to-*Sst*I fragment of the EBV genome (positions 19598 to 21801).

and shows three discrete regions of major cleavage (Fig. 6B, boldface arrows). The three regions occur in similar positions on both strands and map between and flanking the inverted repeat sequence (horizontal arrows drawn between the two strands of DNA sequence).

Easily unwound sequences in EBV *oriP* and hairpin formation. EBV *oriP* contains numerous inverted repeat sequences or palindromes (17, 28). Each of the 20 EBNA1-binding sites in the family of repeats centers on a palindrome



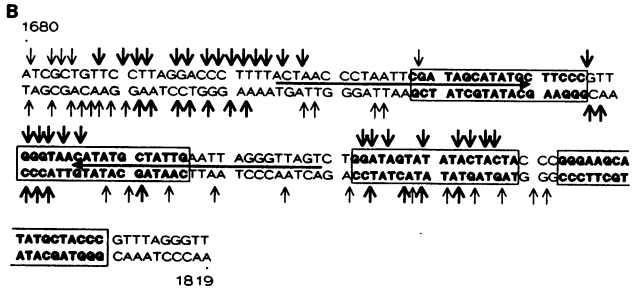


FIG. 6. Nucleotide-level map of the P1 nuclease-hypersensitive sites in the dyad symmetry region. (A) DNA sequence analysis of the P1 nuclease-hypersensitive sites in pWE-DS. Supercoiled pWE-DS was nicked with P1 nuclease, linearized and 5′ end labeled at the *Eco*RI cleavage site, and cut with *Pst*I. The 172-bp *Pst*I-*Eco*RI fragment was isolated after electrophoresis in a polyacrylamide gel. P1 nuclease-nicked DNA (lane P1 +) was electrophoresed alongside the products of Maxam and Gilbert sequencing reactions (lanes A, C/T, and G) performed on the same DNA without P1 nuclease treatment. The data shown are for the top strand of the DNA sequence (see panel B). P1 nuclease-specific nicks correspond to bands present in lane P1 + and absent in lane P1 −. The position of the 65-bp dyad symmetry sequence is indicated with arrows on the left. (B) Locations of nicks within the P1 nuclease-hypersensitive sequence. The vertical arrows indicate the locations of the P1 nuclease nicks. The boldface arrows indicate sites of frequent nicking, and the lightface arrows indicate sites of less frequent nicking. The two horizontal arrows between the complementary DNA strands indicate inverted repeat sequences that make up the 65-bp dyad. The boxed sequences contain the binding sites for the EBV protein EBNA1 (28). Nucleotide positions are numbered relative to the *Eco*RI cleavage site in *oriP*.

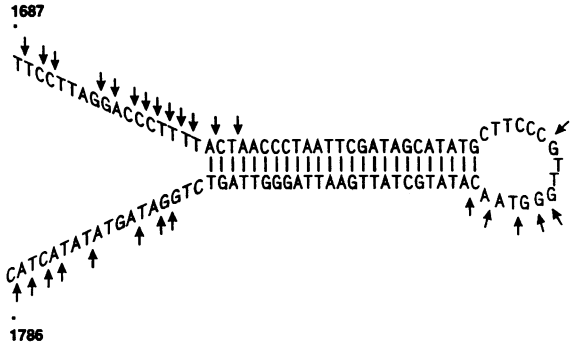
which, when the DNA is strand separated (melted), can potentially form intrastrand base pairs leading to small hairpin structures. Such small hairpins (12 bp; stem, 4; loop, 4) have been proposed to form in the family of repeats, on the basis of studies with single-strand-specific nucleases (26); however, direct evidence for this hypothesis is lacking since the cleavages were not mapped at the nucleotide level and their precise locations with respect to inverted repeat sequences are not known. As an alternative to small hairpins, large hairpin structures could potentially form in the family of repeats and in the dyad symmetry region if intrastrand base pairing occurs between, as opposed to within, EBNA1-binding sites (17, 28); however, there is no experimental evidence for such large hairpin structures.

We analyzed the *oriP* sequence with the computer program StemLoop (6) and searched for stems ≥4 bp and loops ≤20 bases. The hairpin with the longest stem (25 bp) occurred in the dyad symmetry region. This hairpin includes the 65 bases that make up the dyad sequence (Fig. 6B, horizontal arrows between the sequence). Figure 7A shows the three major regions of P1 nuclease cleavage in the dyad symmetry region superimposed on the hairpin structure. The experimental data are consistent with P1 nuclease cleavage at the loop of the hairpin structure and with extensive cleavage flanking both sides of the base-paired stem. The flanking cleavages indicate that the DNA is single stranded on both sides of the hairpin stem. A similar structure is also deduced to form in the complementary (bottom) strand on the basis of detection of three major regions of P1 nuclease cleavage directly opposing the three regions in the top strand (Fig. 6B, boldface arrows). The overall structure deduced to form within duplex DNA is a single-stranded bubble containing a hairpin in each strand.

The hairpin with the second longest stem (22 bp, including two mismatches) identified by the StemLoop program occurs at a unique site within the family of repeats. The hairpin includes an inverted repeat sequence at EBNA1 sites 10 and 11 (Fig. 5C, horizontal arrows between the sequence) which is associated with the P1 nuclease-hypersensitive site. No identical inverted repeat sequences are present elsewhere in the family of repeats as a result of imperfections in the repeat sequences. Figure 7B shows the major P1 nuclease cleavages in repeats 10 and 11 superimposed on the potential hairpin structure. The data are consistent with P1 cleavage at the loop of a hairpin (Fig. 7B). A similar hairpin structure is deduced to form in the complementary (bottom) strand on the basis of the disposition of the major nuclease cleavages (Fig. 5C, boldface vertical arrows) with respect to the inverted repeat sequence (horizontal arrows). In contrast to the hairpins in the dyad symmetry region, the hairpins in the family of repeats are not associated with major cleavages immediately flanking both sides of the inverted repeat sequence (Fig. 5C) that comprises the base-paired stem. The structure appears to be double stranded in the regions flanking the hairpin stems. The overall DNA structure deduced from the nuclease cleavage patterns in the sequences of both strands is cruciform.

Major P1 nuclease cleavages also occur in the family of repeats at EBNA1 site 14 (Fig. 5B). The StemLoop program revealed no hairpin structure that could account for the major cleavages in repeat 14. The major cleavages in repeat

A. DYAD SYMMETRY REGION



B. FAMILY OF REPEATS: EBNA1 sites 10 & 11

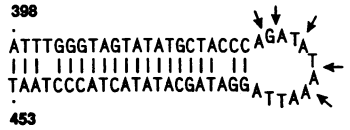


FIG. 7. Single-strand-specific nuclease cleavages at potential hairpins in *oriP*. Major cleavages introduced by P1 nuclease (arrows) are superimposed on hairpin structures predicted by the StemLoop program (6). Nucleotide positions are numbered relative to the EcoRI cleavage site in *oriP*. (A) Dyad symmetry region. The hairpin stem is 25 bp in length (including one G-T pair), and the loop contains 15 bases. (B) Family of repeats. The hairpin stem is 22 bp long (including two mismatches), and the loop contains 12 bases.

14 are flanked by numerous minor cleavages that occur in both strands over a broad region 5′ and 3′ to the major cleavages (Fig. 5C). The total region of major and minor cleavages is extensive and spans 100 bp, beginning in repeat 13 and ending in repeat 16. Single-strand-specific nuclease cleavage over an extensive region is consistent with a broad region of localized DNA unwinding (single-stranded bubble), like that detected with our nuclease-hypersensitivity assay on other DNA molecules (20, 21, 25, 35). However, the distribution of cleavages is unusual in that a portion of the broad region (part of repeat 14) is cleaved more frequently than the rest of the region (Fig. 5C).

DISCUSSION

We have shown that DNA in the dyad symmetry region derived from EBV *oriP* is intrinsically easy to unwind. The easily unwound region is identified by its hypersensitivity to the single-strand-specific enzyme P1 nuclease in a negatively supercoiled plasmid (Fig. 2, DS). In our assay conditions, the nuclease-hypersensitive site identifies the DNA region with the lowest free energy cost for unwinding in the entire plasmid (20, 21, 36). Consistent with this concept, the nuclease-hypersensitive site in the dyad symmetry region corresponds to the DNA region in pWE-DS with the lowest helical stability as determined by computer analysis (Fig. 3B).

The dyad symmetry region is the physical origin of DNA replication within the genetically defined *oriP* (10). The intrinsic ease of DNA unwinding is a property that the EBV dyad symmetry region shares with other replication origins. The intrinsic ease of unwinding in the dyad symmetry region may be required to facilitate origin unwinding to allow the replication machinery access to the parental DNA strands during initiation. Such a sequence is termed a DUE. A DUE is required for replication from the *E. coli* chromosomal replication origin (*oriC*) (20) and from the H4 autonomously replicating sequence (*ARS*) (35) and the C2G1 *ARS* in *S. cerevisiae* (37). At *E. coli oriC*, localized unwinding induced by initiator protein binding occurs in the DUE in vitro (3) and in vivo (12). Initiator protein binding induces localized unwinding at replication origins from SV40 (1) and phage λ (31) and extensive structural distortions at a herpes simplex virus origin (18). The SV40 origin requires a DUE (21a), and the other origins may also require a DUE to facilitate protein-induced unwinding. The components of the EBV dyad symmetry region are similar to those in *E. coli oriC*, the SV40 origin, and yeast origins, since the dyad symmetry region contains an easily unwound sequence as well as binding sites for an essential replication protein (EBNA1). However, DNA strand separation has not been detected when EBNA1 binds at the EBV dyad symmetry region (7, 13). Although EBNA1 is required for replication of EBV DNA (22, 40), its specific role in replication is not known. Since no other EBV-encoded protein is required, host proteins that bind EBNA1 and/or the dyad symmetry region may be required to initiate origin unwinding.

On an *oriP* plasmid containing both the family of repeats and the dyad symmetry region, the family of repeats is the only site hypersensitive to P1 nuclease. Nuclease hypersensitivity in our assay is hierarchical since hypersensitivity reflects stable DNA unwinding (21). Stable unwinding at the most easily unwound site reduces the level of negative supercoiling in a plasmid and the probability of stable unwinding at less easily unwound sites in the same molecule. Our finding of nuclease hypersensitivity in the family of repeats but not in the dyad symmetry region of *oriP* indicates that the family of repeats is more easily unwound than the dyad symmetry region. Consistent with this interpretation, computer analysis reveals that the family of repeats has a lower helical stability than the dyad symmetry region. The ease of unwinding in the dyad symmetry region can contribute to replication origin function even though the region is not the most easily unwound sequence in a plasmid containing the family of repeats. Studies on the DUE in yeast origins have shown that it is the absolute ability of the origin to unwind that is important to plasmid replication, not the relative unwinding ability with respect to other regions in the plasmid (36). In the presence of EBNA1 protein and host proteins that may participate in the initiation of replication, the dyad symmetry region and the family of repeats are likely involved in protein complexes and are not expected to compete for unwinding in the way that they do in naked supercoiled DNA. In the case of *E. coli oriC*, the initiation protein complex acts locally to unwind the DNA (3) and only a nearby easily unwound sequence that is properly positioned in relation to the protein binding sites functions as a DUE (20).

The presence of numerous inverted repeat sequences in EBV *oriP* has prompted speculation on the potential for altered DNA secondary structures, especially large hairpins (17, 28); however, there has been no experimental evidence for such structures. Low-resolution mapping of single-strand-specific nuclease cleavages in *oriP* led others to propose that small hairpins or, alternatively, transiently unwound regions form in the family of repeats (26). Our

nucleotide-level analysis of P1 nuclease cleavage patterns indicates that the stably unwound DNA detected in our assay condition is associated with large hairpins (Fig. 7). The overall DNA structure deduced to form in the dyad symmetry region is a broad single-stranded bubble with a large hairpin in each strand. In our assay conditions, negatively supercoiled DNA stably unwinds at the sequence with the lowest helical stability (20, 21, 36) which, in pWE-DS, is the dyad symmetry region. Once the DNA is strand separated, intrastrand base pairing and hairpin formation at the dyad sequence is expected to be energetically favored over single-stranded structures. The large hairpins detected in the family of repeats do not form within a single-stranded bubble. Instead, these hairpins appear to adopt a typical cruciform structure which is double stranded in the regions flanking the hairpin stems.

Cruciform extrusion can be kinetically forbidden in negatively supercoiled DNA (5, 11); however, cruciform extrusion can be kinetically favorable if the inverted repeat is near a broad DNA sequence with low helical stability (2). Consistent with the latter possibility, the cruciform deduced to form in the family at repeats 10 and 11 occurs near a broad region (100 bp) of P1 nuclease cleavages in repeats 13 through 16. The region is similar in breadth to regions of low helical stability detected by our nuclease-hypersensitivity assay on other DNA molecules (20, 21, 25, 35); however, nuclease-hypersensitive sites previously characterized in our laboratory do not occur in or near long inverted repeat sequences that could extrude large cruciforms and do not contain a small region of enhanced cleavage, as seen in repeat 14. The enhanced cleavage frequency in repeat 14 compared with that of repeats 13, 15, and 16 may be related to the extrusion of the large cruciform in repeats 10 and 11, possibly reflecting the equilibrium distribution between two populations of DNA molecules: (i) a majority with a small single-stranded bubble in repeat 14 plus a cruciform in repeats 10 and 11 and (ii) a minority with a broad single-stranded bubble in repeats 13 through 16 and no cruciform. Our findings indicate that DNA in the family of repeats offers considerable conformational flexibility. It remains to be determined how this conformational flexibility influences protein interactions and the diverse biological functions of the family of repeats.

ACKNOWLEDGMENTS

We thank Darren Natale, John Yates, Charles Miller, and Ruea-Yea Huang for critical reading of the manuscript.

This research was supported in part by a grant from the Buffalo Foundation and by a grant (GM44119) from the National Institutes of Health.

REFERENCES

1. **Borowiec, J. A., and J. Hurwitz.** 1988. Localized melting and structural changes in the SV40 origin of replication induced by T-antigen. EMBO J. **7**:3149–3158.
2. **Bowater, R., F. Aboul-ela, and D. M. J. Lilley.** 1991. Large-scale stable opening of supercoiled DNA in response to temperature and supercoiling in (A+T)-rich regions that promote low-salt cruciform extrusion. Biochemistry **30**:11495–11506.
3. **Bramhill, D., and A. Kornberg.** 1988. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. Cell **52**:743–755.
4. **Breslauer, K. J., R. Frank, H. Blocker, and L. A. Marky.** 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA **83**:3746–3750.
5. **Courey, A. J., and J. C. Wang.** 1983. Cruciform formation in negatively supercoiled DNA may be kinetically forbidden under physiological conditions. Cell **33**:817–829.
6. **Devereux, J., P. Haeberl, and O. Smithes.** 1984. A comprehensive set of sequence analysis for the VAX. Nucleic Acids Res. **12**:387–395.
7. **Frappier, L., and M. O'Donnell.** 1991. Epstein-Barr nuclear antigen 1 mediates a DNA loop within the latent replication origin of Epstein-Barr virus. Proc. Natl. Acad. Sci. USA **88**:10875–10879.
8. **Frappier, L., and M. O'Donnell.** 1992. EBNA1 distorts *oriP*, the Epstein-Barr virus latent replication origin. J. Virol. **66**:1786–1790.
9. **Fujimoto, M., A. Kuninaka, and H. Yoshino.** 1974. Purification of a nuclease from *Penicillium citrinum*. Agric. Biol. Chem. **38**:777–783.
10. **Gahn, T. A., and C. L. Schildkraut.** 1989. The Epstein-Barr virus origin of plasmid replication, *oriP*, contains both the replication and termination sites of DNA replication. Cell **58**: 527–535.
11. **Gellert, M., M. H. O'Dea, and K. Mizuuchi.** 1983. Slow cruciform transitions in palindromic DNA. Proc. Natl. Acad. Sci. USA **80**:5545–5549.
12. **Gille, H., and W. Messer.** 1991. Localized DNA melting and structural pertubations in the origin of replication, *oriC*, of *Escherichia coli in vitro* and *in vivo*. EMBO J. **10**:1579–1584.
13. **Hearing, J., Y. Mulhaupt, and S. Harper.** 1992. Interaction of Epstein-Barr virus nuclear antigen 1 with the viral latent origin of replication. J. Virol. **66**:694–705.
14. **Henle, W., V. Diehl, G. Kohn, H. zur Hausen, and G. Henle.** 1967. Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells. Science **157**:1064–1065.
15. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
16. **Iacono-Connors, L., and D. Kowalski.** 1986. Altered DNA conformations in the gene regulatory region of torsionally-stressed SV40 DNA. Nucleic Acids Res. **14**:8949–8962.
17. **Karlin, S.** 1986. Significant potential secondary structures in the Epstein-Barr virus genome. Proc. Natl. Acad. Sci. USA **83**: 6915–6919.
18. **Koff, A., J. F. Schwedes, and P. Tegtmeyer.** 1991. Herpes simplex virus origin-binding protein (UL9) loops and distorts the viral replication origin. J. Virol. **65**:3284–3292.
19. **Kowalski, D.** 1984. Changes in site specificity of single-strand specific endonucleases on supercoiled PM2 DNA with temperature and ionic environment. Nucleic Acids Res. **12**:7071–7086.
20. **Kowalski, D., and M. J. Eddy.** 1989. The DNA unwinding element: a novel, *cis*-acting component that facilitates opening of the *Escherichia coli* replication origin. EMBO J. **8**:4335–4344.
21. **Kowalski, D., D. A. Natale, and M. J. Eddy.** 1988. Stable DNA unwinding, not "breathing," accounts for single-strand-specific nuclease hypersensitivity of specific A+T-rich sequences. Proc. Natl. Acad. Sci. USA **85**:9464–9468.

21a. **Lin, S., and D. Kowalski.** Unpublished data.

22. **Lupton, S., and A. J. Levine.** 1985. Mapping genetic elements of Epstein-Barr virus that facilitate extrachromosomal persistence of Epstein-Barr virus-derived plasmids in human cells. Mol. Cell. Biol. **5**:2533–2542.
23. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
24. **Middleton, T., and B. Sugden.** 1992. EBNA1 can link the enhancer element to the initiator element of the Epstein-Barr virus plasmid origin of DNA replication. J. Virol. **66**:489–95.
25. **Natale, D. A., A. E. Schubert, and D. Kowalski.** 1992. DNA helical stability accounts for mutational defects in a yeast replication origin. Proc. Natl. Acad. Sci. USA **89**:2654–2658.
26. **Orlowski, R., and G. Miller.** 1991. Single-stranded structures are present within plasmids containing the Epstein-Barr virus latent origin of replication. J. Virol. **65**:677–686.
27. **Radloff, R., W. Bauer, and J. Vinograd.** 1967. A dye-buoyant-

density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA of HeLa cells. Proc. Natl. Acad. Sci. USA **57**:1514–1521.

28. **Rawlins, D. R., G. Milman, S. D. Hayward, and G. S. Hayward.** 1985. Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA1) to clustered sites in the plasmid maintenance region. Cell **42**:859–868.
29. **Reisman, D., and B. Sugden.** 1986. *trans*-activation of an Epstein-Barr viral transcriptional enhancer by the Epstein-Barr viral nuclear antigen 1. Mol. Cell. Biol. **6**:3838–3846.
30. **Reisman, D., J. Yates, and B. Sugden.** 1985. A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two *cis*-acting components. Mol. Cell. Biol. **5**:1822–1832.
31. **Schnos, M., K. Zahn, R. B. Inman, and F. R. Blatner.** 1988. Initiation protein induced helix destabilization at the lambda origin: a prepriming step in DNA replication. Cell **52**:385–395.
32. **Sheflin, L. G., and D. Kowalski.** 1984. Mung bean nuclease cleavage of a dA+dT-rich sequence or an inverted repeat sequence in supercoiled PM2 DNA depends on ionic environment. Nucleic Acids Res. **12**:7087–7104.
33. **Sheflin, L. G., and D. Kowalski.** 1985. Altered DNA conformations detected by mung bean nuclease occur in promoter and terminator regions of supercoiled pBR322 DNA. Nucleic Acids Res. **13**:6137–6154.
34. **Su, W., T. Middleton, B. Sugden, and H. Echols.** 1991. DNA looping between the origin of Epstein-Barr virus and its enhancer site: stabilization of an origin complex with Epstein-Barr nuclear antigen 1. Proc. Natl. Acad. Sci. USA **88**:10870–10874.
35. **Umek, R. M., and D. Kowalski.** 1988. The ease of DNA unwinding as a determinant of initiation at yeast replication origins. Cell **52**:559–567.
36. **Umek, R. M., and D. Kowalski.** 1990. The DNA unwinding element in a yeast replication origin functions independently of easily unwound sequences present elsewhere on a plasmid. Nucleic Acids Res. **18**:6601–6605.
37. **Umek, R. M., and D. Kowalski.** 1990. Thermal energy suppresses mutational defects in DNA unwinding at a yeast replication origin. Proc. Natl. Acad. Sci. USA **87**:2486–2490.
38. **Wysokenski, D. A., and J. L. Yates.** 1989. Multiple EBNA1-binding sites are required to form an EBNA1-dependent enhancer and to activate a minimal replicative origin within *oriP* of Epstein-Barr virus. J. Virol. **63**:2657–2666.
39. **Yates, J. L., N. Warren, D. Reisman, and B. Sugden.** 1984. A *cis*-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc. Natl. Acad. Sci. USA **81**:3806–3810.
40. **Yates, J. L., N. Warren, and B. Sugden.** 1985. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. Nature (London). **313**:812–815.