JOURNAL OF VIROLOGY, May 1987, p. 1495–1506
0022-538X/87/051495-12$02.00/0

Vol. 61, No. 5

Sequences of the Epstein-Barr Virus (EBV) Large Internal Repeat Form the Center of a 16-Kilobase-Pair Palindrome of EBV (P3HR-1) Heterogeneous DNA

HAL B. JENSON,[1,2]* PAUL J. FARRELL,[3] AND GEORGE MILLER[1,2,4]

Departments of Epidemiology and Public Health,[1] Pediatrics,[2] and Molecular Biophysics and Biochemistry,[4] Yale University School of Medicine, New Haven, Connecticut 06510, and Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, London W2 1PG England[3]*

Received 10 November 1986/Accepted 30 January 1987

We have previously characterized several genomic rearrangements of Epstein-Barr virus (EBV) DNA contained in one of the defective EBV genomes harbored by the P3HR-1 (HR-1) line (H. B. Jenson, M. S. Rabson, and G. Miller, J. Virol. 58:475–486, 1986). One recombinant clone of heterogeneous DNA (het DNA) from this defective genome is an *Eco*RI fragment of 16 kilobase pairs (kbp) which is a palindrome. DNA digestion fragments specific for the center of this palindrome were present in cells which contained het DNA but not in cells which lacked het DNA. Thus, the palindrome was not an artifact of DNA cloning. The organization of the center of this palindrome was studied by DNA sequencing. The comparable region of the parental HR-1 genome was also studied by DNA sequencing. The central 3,495 base pairs (bp) of the palindrome were composed of sequences derived exclusively from internal repeat 1 of EBV, represented by the *Bam*HI W fragment. At each end of the central 3,495 bp was a symmetrical recombination with sequences of *Bam*HI-Z, located more than 50 kbp away on the standard EBV genome. The central 3,495 bp were composed of an unduplicated 341 bp flanked by two perfect palindromic repeats of 1,577 bp. The 341-bp unique region was a portion of a 387-bp region of standard HR-1 *Bam*HI-W which was identical to the central 387 bp of the palindrome. This central 387-bp region contained numerous stretches of dyad symmetry capable of forming a large stem-and-loop structure. The palindromic rearrangement had created two novel open reading frames in het DNA derived from standard HR-1 *Bam*HI-W sequences. These two het DNA open reading frames had different amino termini but identical carboxy termini derived from the large open reading frame in standard HR-1 *Bam*HI-W (HR-1 BWRF1). The *Bam*HI-W sequences found in het DNA did not include either the TATA box of standard HR-1 *Bam*HI-W or the exons which are present in the potentially polycistronic latent mRNAs encoding EBV nuclear antigens. These marked alterations in genomic structure may relate to the unique biologic properties of virus stocks containing het DNA by creation of new polypeptides or by formation or deletion of regulatory or functional signals.

Study of Epstein-Barr virus (EBV) from the P3HR-1 (HR-1) cell subclone of the Burkitt lymphoma line Jijoye (21) provides a unique approach to understanding the biology of EBV. HR-1 is a spontaneous deletion mutant of EBV which fails to immortalize lymphocytes (7, 29, 32, 37). In addition, HR-1 virus is now known to contain defective EBV virions, also immortalization deficient, which account for several important properties of the HR-1 strain. These include the ability to *trans* activate latent EBV in X50-7 cells, resulting in complete viral replication and production of immortalization-competent virions (28) and enhanced ability to activate early replicative antigen expression in cell lines such as Raji (15, 16, 29, 32). Unlike standard EBV episomes, the defective virions are maintained in the HR-1 culture by cell-to-cell spread rather than by vertical partitioning (27). Thus, the defective virions are replicons which contain their own, as yet uncharacterized, origins of DNA replication.

The genomes of the defective viruses, called heterogeneous DNA (het DNA), contain numerous deletions and rearrangements compared with the standard EBV genome (9, 23, 28). The structure of 36 kilobase pairs (kbp) of het DNA isolated from a cellular subclone of the HR-1 line has been mapped by restriction endonuclease analysis and comparison with standard EBV DNA (23). One het DNA fragment, an *Eco*RI fragment of 16 kbp, is arranged as a palindrome of *Bam*HI subfragments flanking a central 1.2-kbp *Bam*HI subfragment.

The purpose of this study was to address several unanswered questions about this unique 16-kbp palindrome. Was the palindrome a perfect palindrome? How many recombination events occurred in the formation of the palindrome? Were new open reading frames created? Therefore, the central 3,495 base pairs (bp) of this palindrome were studied by DNA sequencing to determine their exact structural organization and to define the genomic alterations which occurred in the generation of het DNA. The nucleotide sequences of the corresponding region of the standard HR-1 genome were also determined for comparison of the open reading frames and regulatory signals of het DNA with parental HR-1 EBV.

(This work was presented in preliminary form at the Tenth International Herpesvirus Workshop, Ann Arbor, Mich., August 1985.)

MATERIALS AND METHODS

Cells and cellular DNA. As a source of parental HR-1 virus, we used the P3H3 cell line obtained in 1981 from W. Henle. A prototype cellular subclone lacking het DNA is clone HH514-16 (clone 16). Clone HH543-5 (clone 5) contains het DNA in addition to the standard HR-1 genome and

* Corresponding author.

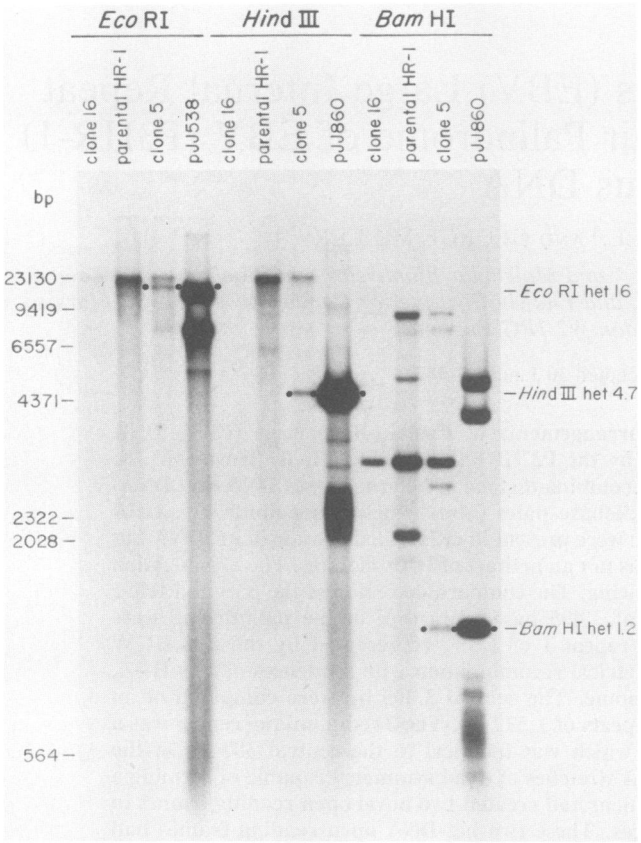


FIG. 1. Identification of the restriction endonuclease fragments from the palindrome of het DNA. Total intracellular DNA from parental P3HR-1, HR-1 clone 16 (het$^-$), and HR-1 clone 5 (het$^+$) was digested separately with *Eco*RI, *Hin*dIII, or *Bam*HI. The chimeric plasmid construct containing the cloned het DNA restriction endonuclease fragment was included for comparison. Plasmid pJJ538 consists of *Eco*RI het 16 cloned in pSV2-neo. Plasmid pJJ860 consists of *Hin*dIII het 4.7 cloned in pSV2-neo. Plasmid pJJ860 was also used for the *Bam*HI digestion since *Bam*HI het 1.2 is wholly contained within *Hin*dIII het 4.7. The Southern blot was probed with a specific probe from het DNA with homology to the central 373 bp of *Eco*RI het 16. The sizes of the cloned digestion fragments of *Eco*RI het 16 with homology to the center of the palindrome were *Eco*RI (16 kbp), *Hin*dIII (4.7 kbp), and *Bam*HI (1.2 kbp). These center digestion fragments of the palindrome of het DNA and of the cloned DNA fragments are indicated (●) for each restriction endonuclease. The cloned het DNA digestion fragments from the center of this palindrome corresponded identically in size to intracellular DNA digestion fragments found in HR-1 clone 5 cells. None of these fragments was detected in clone 16 cells, which lack het DNA. These fragments were not detected in parental HR-1 cells because the defective het DNA genome constitutes a rare subpopulation of the total number of EBV genomes. Additional digestion fragments of het DNA with homology to this region of *Bam*HI-W were present in parental HR-1 intracellular DNA which were not present in clone 5 DNA.

exhibits the biologic properties of parental HR-1 virus (20, 28, 33).

Plasmids. *Eco*RI het 16 (23) was cloned into pACYC-184 (pJJ66-18) from intracellular DNA of HR-1 clone 5. *Hin*dIII het 4.7 was subcloned from this construct into pSV2-neo (pJJ860) by using standard methods (23). *Eco*RI-A from HR-1 viral DNA (Life Sciences, Inc., St. Petersburg, Fla.) was cloned into pACYC-184 (pHH363). A single *Bam*HI-W from this construct was subcloned into pSV2-neo (pA1). This was the source of standard HR-1 *Bam*HI-W for DNA sequencing. Recombinant plasmids were maintained in *Escherichia coli* DH1 (*recA*$^-$) (18).

Southern blotting. Total intracellular DNA from clone 16, clone 5, and parental HR-1 cells was extracted by the procedure of Davis and Kingsbury (11). Intracellular DNA was digested with *Eco*RI, *Hin*dIII, or *Bam*HI. The fragments were separated by electrophoresis in 0.75% agarose. After transfer to nitrocellulose (38), the blot was probed with an [α-^{32}P]dCTP-labeled single-stranded DNA probe prepared by the prime-cut method (3). The probe consisted of a 373-bp *Hin*fI het DNA fragment cloned in M13mp8 (pJJ860.240). This insert was specific for the very center of *Eco*RI het 16 and extended from base pairs 1986 to 2358 of *Hin*dIII het 4.7 (see Fig. 3).

DNA sequencing. *Hin*dIII het 4.7 was isolated from the pSV2-neo plasmid pJJ860 (23) by digestion with *Hin*dIII and purification with Elutip-d (Schleicher & Schuell, Inc., Keene, N.H.) (36). *Bam*HI-W was isolated from the pSV2-neo plasmid containing HR-1 *Bam*HI-W (pA1) by digestion with *Bam*HI and purification with Elutip-d. The linear DNA was self-ligated to produce circular and concatenated forms (12). A random library of DNA fragments was constructed by sonication (12) and cloning into the M13mp8 vector (26). Some specific clones for DNA sequencing were obtained by clone turnaround in M13mp9 (2). Clones were sequenced by the dideoxy-chain termination method (4, 35). All sequences were determined on both strands. Remaining ambiguities were resolved by substituting dITP for dGTP in the sequence reactions (30). For *Hin*dIII het 4.7, each nucleotide in the sequence was determined an average of 18 times. For HR-1 *Bam*HI-W, each nucleotide was determined an average of 12 times.

DNA sequence data was collected with a microcomputer (H. B. Jenson, Comput. Appl. Biosci., in press) and compiled by using the programs of Staden (40). Sequence analysis and comparisons were made by using the programs of Staden (41) and Devereux (13).

RESULTS

Identification of the palindrome in HR-1 clone 5 virus. The existence of viral DNA fragments in HR-1 clone 5 cells corresponding to fragments in recombinant plasmid DNA was verified to exclude the possibility that the palindrome arose during DNA cloning. A single-stranded DNA probe with homology to 373 bp of the central region of *Eco*RI het 16 was used to probe total intracellular DNA from parental HR-1, clone 16, and clone 5 cells digested with *Eco*RI, *Hin*dIII, or *Bam*HI. The chimeric plasmid construct containing the cloned het DNA was placed in an adjacent lane for size comparison (Fig. 1). In clone 5 cells, a 16-kbp *Eco*RI fragment was detected. The *Hin*dIII digestion revealed a 4.7-kbp fragment, and the *Bam*HI digestion revealed a 1.2-kbp fragment each in HR-1 clone 5 cells. Each of these DNA fragments found in clone 5 cells was identical in size to the corresponding cloned het DNA fragment. None of these fragments was detected in clone 16 cells, which lack het DNA (20). These fragments were not detected in parental HR-1 cells because the defective het DNA genome constitutes a rare subpopulation of the total number of EBV genomes.

Parental HR-1 cells also contained het DNA fragments not evident in clone 5 cells. This is most clearly shown in the *Bam*HI digestion in Fig. 1. Parental HR-1 contained *Bam*HI

fragments of approximately 2 and 5 kbp which were not seen in the defective viruses in clone 5 cells or in clone 16 cells which lack defective viral DNA. Thus, parental HR-1 virus contains other defective DNA fragments with homology to the center of *Eco*RI het 16 in addition to those present in clone 5 cells.

DNA sequence of the center of the palindrome. *Hin*dIII het 4.7, the central *Hin*dIII subfragment of *Eco*RI het 16, contained sequences from standard HR-1 *Bam*HI-W and *Bam*HI-Z joined by a symmetrical rearrangement on each side of the palindrome (Fig. 2). The point of recombination of *Bam*HI-Z to *Bam*HI-W was at thymidine residue 1 in the consensus eucaryotic promoter TATAAA found in *Bam*HI-W. This promoter was subsequently excluded from *Hin*dIII het 4.7 het DNA. The sequence of the central 3,495 bp of *Hin*dIII het 4.7 was derived entirely from sequences comparable to standard HR-1 *Bam*HI-W (Fig. 3). The DNA sequence and analysis of the *Bam*HI-Z sequences of defective and standard HR-1 virus will be presented in a subsequent report (H. B. Jenson and G. Miller, manuscript in preparation).

The DNA sequence of the central 3,495 bp of *Eco*RI het 16 contained a unique central region of 341 bp flanked by two perfect palindromic repeats of 1,577 bp each (Fig. 2). There were no discrepancies in the sequencing of *Hin*dIII het 4.7 that suggested that any base-pair differences might exist between the two sides of the palindrome outside of the central unique region. M13 clones which had one end in the central unique 341-bp region and extended into the palindromic region of *Hin*dIII het 4.7 specifically excluded differences between the two sides within 200 bp on either side of the 341-bp region, but because the two palindromic sides could not be cloned separately before the DNA sequence was known, *Hin*dIII het 4.7 was sequenced an excessive number of times to increase the likelihood of sequencing both sides. If individual base-pair differences did exist, they would be apparent in the sequence analysis. Each nucleotide of *Hin*dIII het 4.7 was sequenced an average of 18 times, with some regions sequenced 50 times. Thus, for each nucleotide of *Hin*dIII het 4.7, the probability that all of the contributing M13 clones came from the same side was, on the average, 1 in $2^{(18-1)}$ (approximately 1 in 1.3×10^5, or 0.00076%), with certain regions even lower at 1 in $2^{(50-1)}$ (approximately 1 in 5.6×10^{14}, or 0.00000000000018%).

DNA sequence of standard HR-1 *Bam*HI-W. To understand the rearrangements responsible for the defective het DNA genome, the region corresponding to the center of the palindrome from the standard HR-1 genome was sequenced for comparison. Each nucleotide was sequenced an average of 12 times. The DNA sequence of standard HR-1 *Bam*HI-W is shown in Fig. 4. All 3,495 bp in the center of the palindrome in het DNA were derived from sequences of standard HR-1 *Bam*HI-W.

Structure of the central 387 bp of the palindrome. The central 387 bp of *Hin*dIII het 4.7 was composed of the unique central 341 bp (base pairs 2184 to 2524 in Fig. 3) flanked by two 23-bp palindromic repeats. An identical 387-bp region was found in standard HR-1 *Bam*HI-W (base pairs 2272 to 2658 in Fig. 4). Thus, by comparison with standard HR-1 DNA, the palindrome was created by a single rearrangement located at base pairs 2160 to 2161 of *Hin*dIII het 4.7 (Fig. 3). There was no evidence for any additional rearrangement of het DNA sequences within this central 387-bp region.

This 387-bp region was distinctive for a remarkable degree of dyad symmetry, which has also been found in B95-8 EBV (8). Within this region, 80% could be paired to an inverted

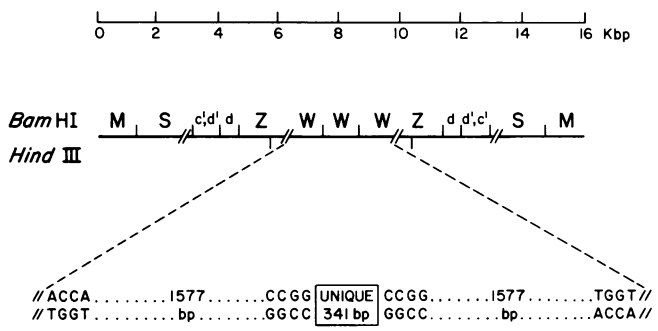


FIG. 2. *Bam*HI and *Hin*dIII digestion map and palindromic structure of *Eco*RI het 16. Homologies to standard EBV *Bam*HI fragments are shown above each het DNA *Bam*HI subfragment. *Eco*RI het 16 was a palindrome around a central unique region of 341 bp. The two symmetric intramolecular recombination events (present on each side of the palindrome) which joined regions of standard EBV DNA which are not normally contiguous are indicated (//). The central 3,495 bp were rearranged sequences totally derived from standard HR-1 *Bam*HI-W.

repeat in the formation of a single stem-and-loop structure (Fig. 5). Thus, the palindrome of *Hin*dIII het 4.7 was not only a perfect palindrome outside the central unique region, it was nearly a palindrome within this region. The structural organization of the center of the palindrome was the result of a single recombination of *Bam*HI-W sequences from one end of this stem-and-loop structure with the complementary sequences from the opposite end (Fig. 6).

The total base composition of *Hin*dIII het 4.7, with 63% G+C content, was comparable to that of the entire B95-8 EBV genome, with 60% G+C content. However, the base composition of the 387-bp stem-and-loop structure was 76% G+C content.

Comparison of *Bam*HI-W DNA sequences from B95-8 EBV, standard HR-1 EBV, and HR-1 clone 5 het DNA. Standard HR-1 *Bam*HI-W differed from the previously reported sequences for B95-8 *Bam*HI-W (1, 8, 24). A total of 15 1-bp differences existed between the *Bam*HI-W sequences of HR-1 and B95-8 EBV (Fig. 7). Two individual base-pair insertions existed in the HR-1 *Bam*HI-W sequence compared with the B95-8 *Bam*HI-W sequence. The net result was that HR-1 *Bam*HI-W was 2 bp larger than B95-8 *Bam*HI-W (3,074 bp versus 3,072 bp, respectively).

All the base-pair differences present in HR-1 *Bam*HI-W compared with B95-8 *Bam*HI-W were also present in those sequences of *Bam*HI-W found in het DNA. In addition, two 1-bp differences existed between HR-1 *Bam*HI-W and that portion of *Bam*HI-W found in het DNA. These two 1-bp changes were not in a probable coding region, nor were they part of probable regulatory sequences. No insertions or deletions existed between *Bam*HI-W sequences of standard HR-1 and het DNA. These findings are consistent with the derivation of het DNA from the standard HR-1 genome and also suggest that some differences may have been introduced during the generation or propagation of the defective genome. Alternatively, differences may exist between *Bam*HI-W sequences between individual genomes or between individual repeats of internal repeat 1 within the same EBV genome. The existence of 1-bp changes between standard HR-1 and het DNA is also seen in other regions of het DNA (H. B. Jenson and G. Miller, in preparation).

Possible open reading frames in *Bam*HI-W DNA sequences of standard HR-1 EBV and HR-1 clone 5 het DNA. The nomenclature for EBV open reading frames proposed by

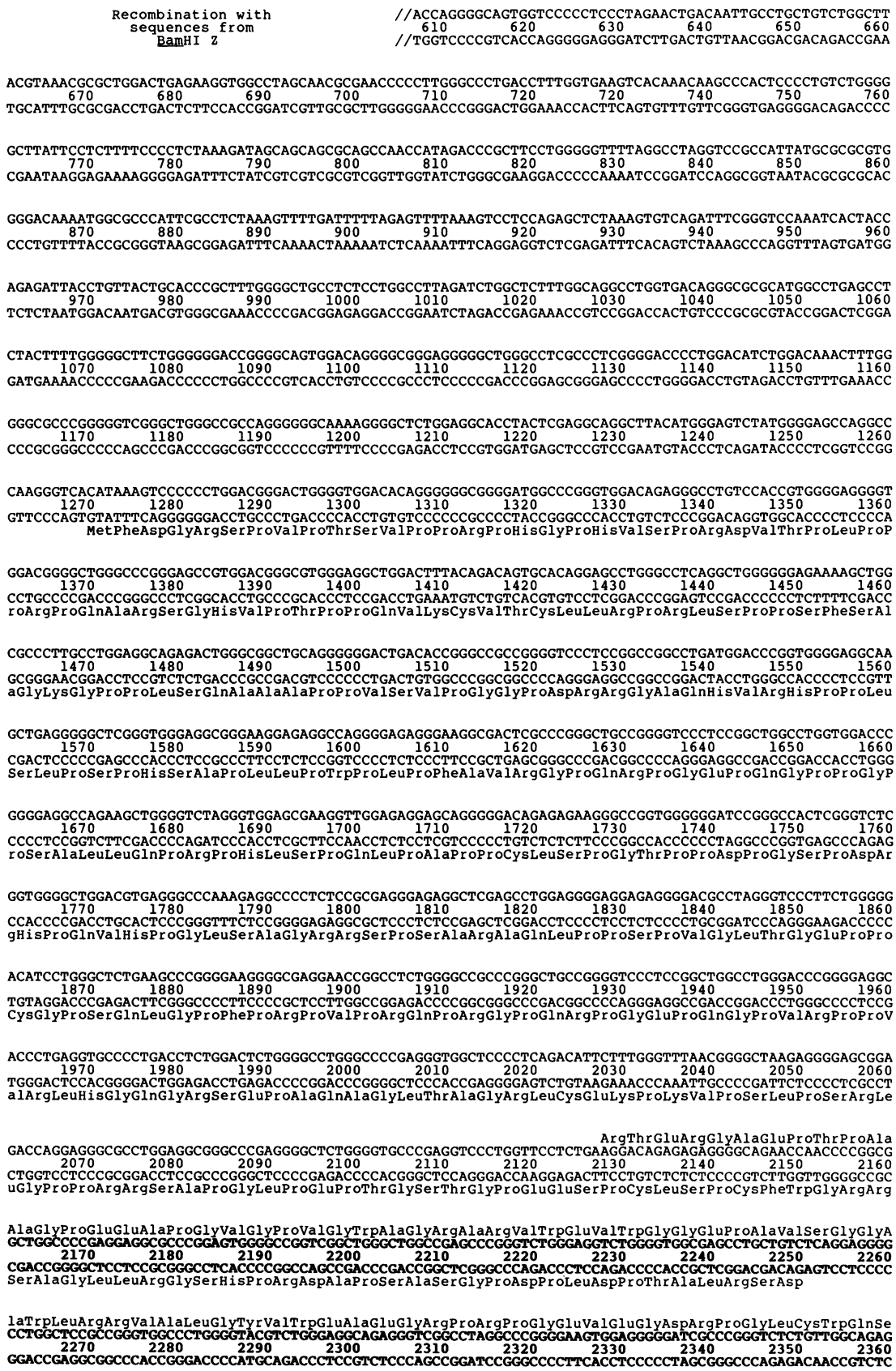


FIG. 3. DNA sequence of the central 3,495 bp of the palindrome of *Eco*RI het 16. The numbering begins at the left *Hin*dIII site of *Hin*dIII het 4.7, the central *Hin*dIII 4.7-kbp DNA fragment of this palindrome (Fig. 2). The central 387 bp of the palindrome, from base pairs 2161 to 2547, are shown in boldface type. Amino acid translations of the two variants of the standard HR-1 BWRF1 are shown. The rightward reading frame is shown above the top strand of DNA (HR-1 het BWRF1), and the leftward reading frame is shown under the bottom strand of DNA (HR-1 het BWLF1).

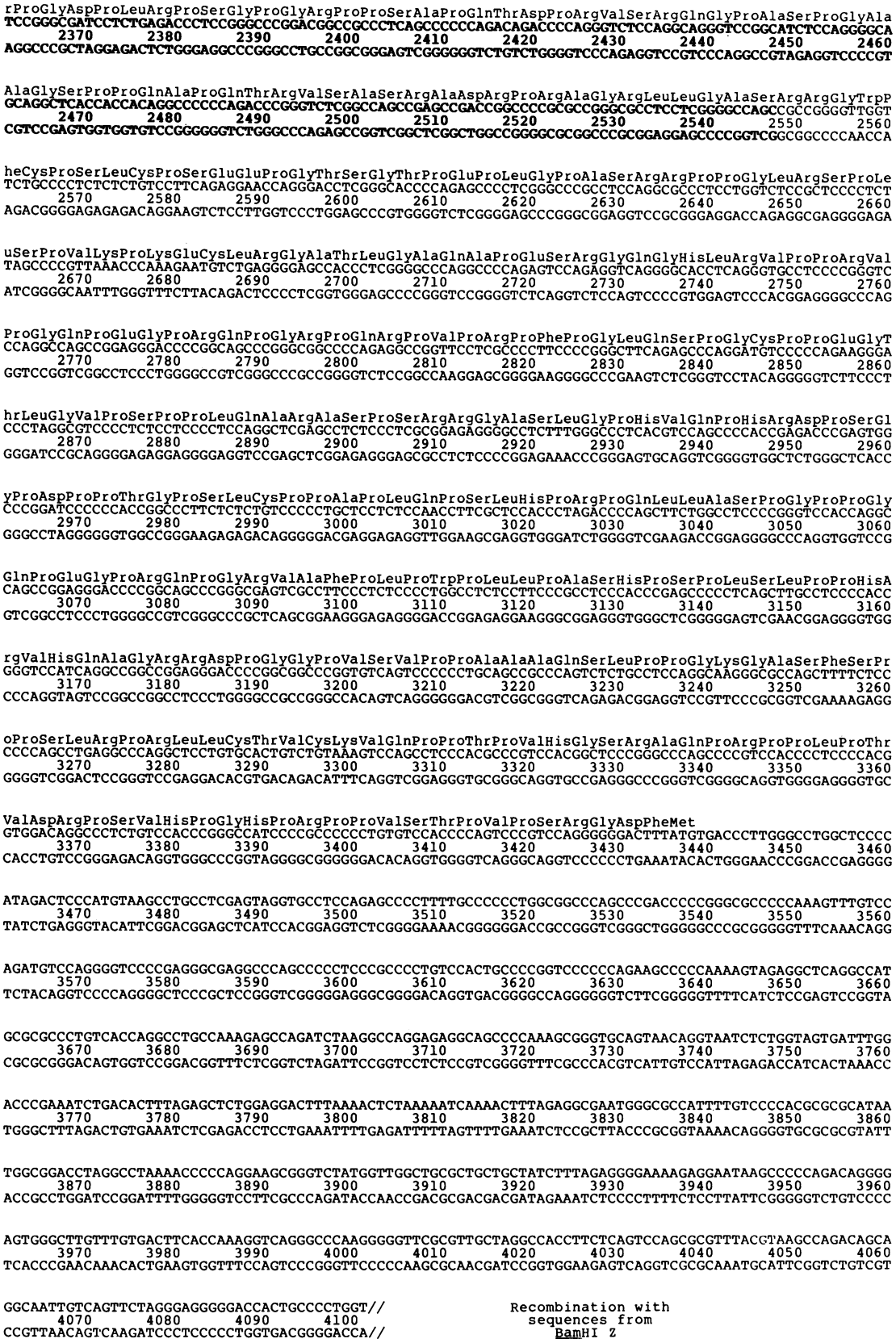


Fig. 3—*Continued.*

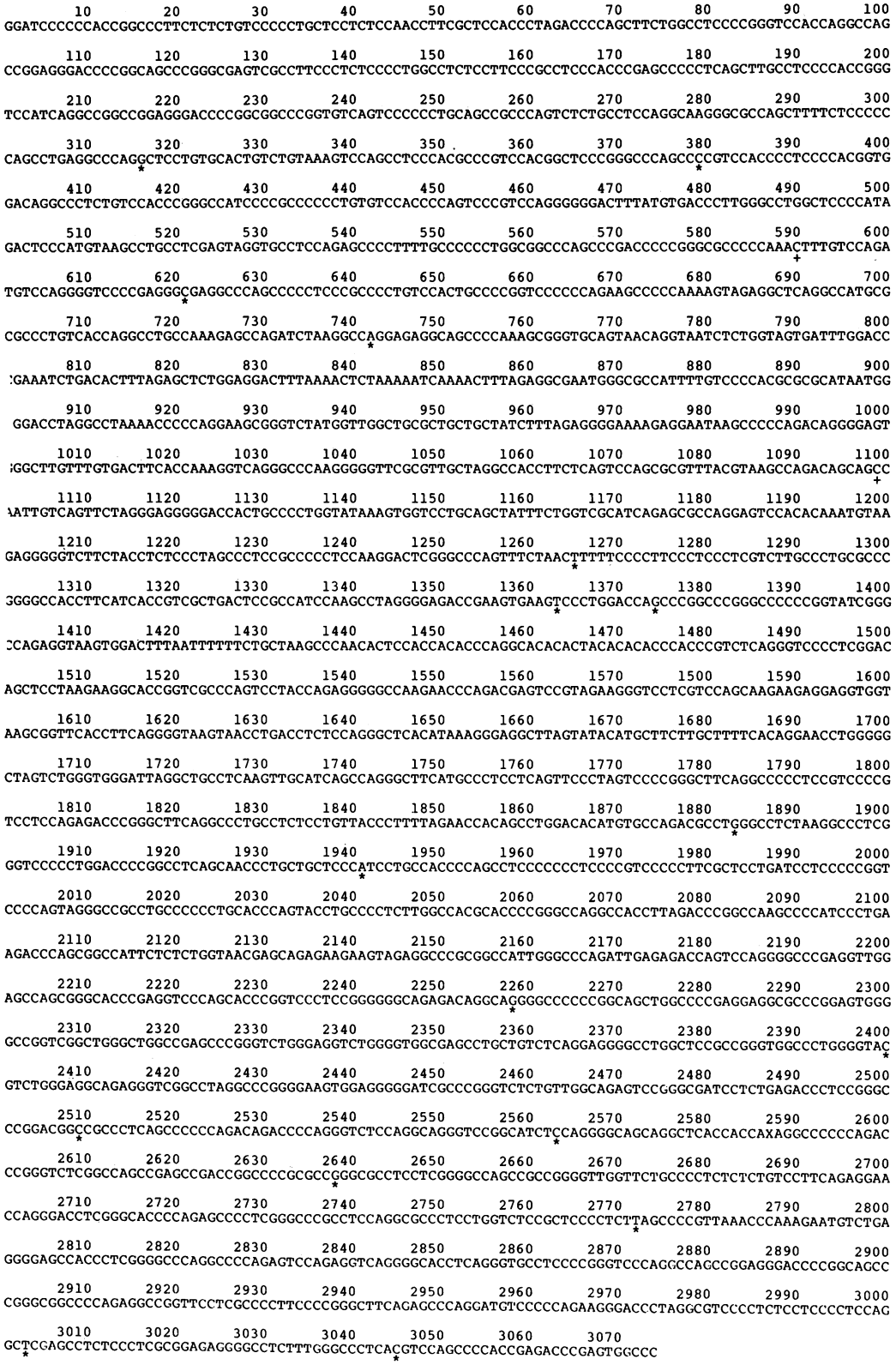
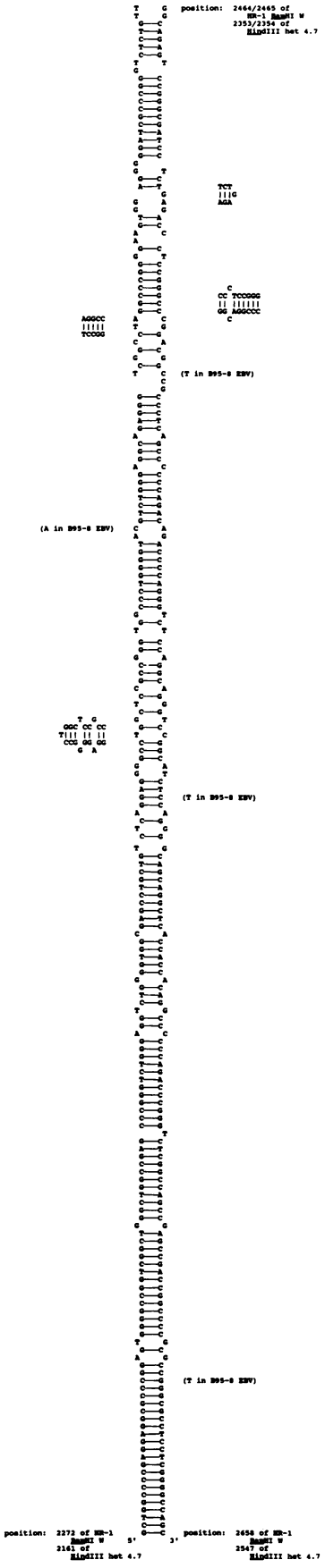


FIG. 4. DNA sequence of standard HR-1 *Bam*HI-W. Differences with the B95-8 *Bam*HI-W sequence (1) are indicated (*). These included 2 1-bp insertions and 15 1-bp changes in the HR-1 *Bam*HI-W sequence as compared with the B95-8 *Bam*HI-W sequence. The two 1-bp changes of standard HR1 *Bam*HI-W compared with the sequences of *Bam*HI-W found in het DNA are indicated (+). The 387-bp region of dyad symmetry extended from base pairs 2272 to 2658. The point of recombination between sequences of *Bam*HI-W to sequences of *Bam*HI-Z found in het DNA was at base pair 1138 of the HR-1 *Bam*HI-W sequence.



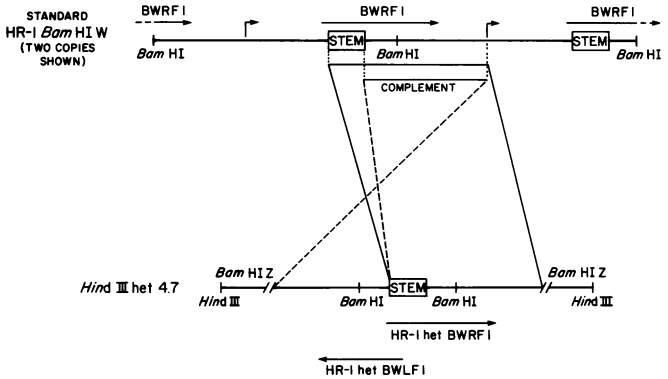


FIG. 6. Novel reading frames generated by rearrangement of regions of standard HR-1 *Bam*HI-W forming the center of the palindrome of het DNA. There were two variations of the BWRF1 open reading frame compared with the BWRF1 of standard HR-1 *Bam*HI-W. One open reading frame was in the rightward direction, HR-1 het BWRF1, and one was in the leftward direction, HR-1 het BWLF1. Both of these had unique amino-terminal regions compared with those of standard HR-1 BWRF1. A symmetrical recombination of sequences (//) from *Bam*HI-W to sequences of *Bam*HI-Z was present on each side of the palindrome. The palindromic sequence continued through this rearrangement to the *Hin*dIII site and beyond. Possible promoters (↱) are indicated. The rectangles labeled STEM define the limits of the 387-bp region of dyad symmetry.

Baer et al. (1) was extended to the standard HR-1 genome and het DNA. Standard HR-1 open reading frames are preceded by HR-1, and HR-1 het DNA open reading frames are preceded by HR-1 het.

The largest open reading frame present in B95-8 *Bam*HI-W is BWRF1 (1). Possibly significant differences existed between this open reading frame in standard HR-1 compared with that of B95-8 EBV (Fig. 7). Of the 15 1-bp differences between HR-1 and B95-8 *Bam*HI-W, 1 resulted in a change of the termination codon immediately preceding B95-8 BWRF1 to the codon for tyrosine. This resulted in the 5′ extension of the standard HR-1 BWRF1 open reading frame compared with that of B95-8 BWRF1. But before the next expected upstream termination codon is reached in that frame, a 1-bp insertion in standard HR-1 *Bam*HI-W compared with that in B95-8 *Bam*HI-W resulted in a frameshift and further 5′ extension of this open reading frame. Thus, the HR-1 BWRF1 was 1,404 bp long, 22% larger than the 1,149 bp of BWRF1 in B95-8 EBV.

The central region of *Eco*RI het 16 derived from sequences of standard HR-1 *Bam*HI-W contained two large possible open reading frames, one in the rightward direction and one in the leftward direction (Fig. 6 and 7). These two open reading frames were designated HR-1 het BWRF1 and HR-1 het BWLF1, respectively. The change of the termination codon immediately preceding B95-8 BWRF1 to a tyrosine

FIG. 5. The 387-bp region of dyad symmetry present in HR-1 *Bam*HI-W, indicating possible stem-and-loop formation. Base pairing is indicated by lines between nucleotides. Some secondary loops may also form, which are shown at the respective sides of the primary loop. These secondary loops are in two regions which display the least potential for formation of a single stem-and-loop structure. The four 1-bp differences compared with the B95-8 DNA sequence (1) are indicated. There were no DNA sequence differences between parental HR-1 and het DNA in this region.

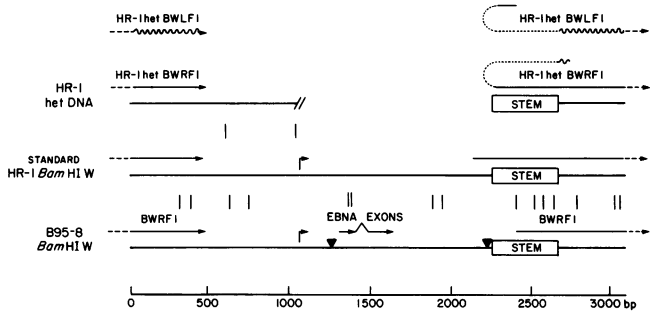


FIG. 7. Diagram of DNA differences in the *Bam*HI-W sequences from B95-8 *Bam*HI-W (1), standard HR-1 *Bam*HI-W, and HR-1 clone 5 het DNA. The sequence is represented along a horizontal line beginning at the *Bam*HI site. The rectangle labeled STEM defines the limits of the 387-bp region of dyad symmetry present in *Bam*HI-W of B95-8 EBV, standard HR-1 EBV, and HR-1 clone 5 het DNA. Only a portion of standard HR-1 *Bam*HI-W sequences are represented in HR-1 clone 5 het DNA, from the left end of the 387-bp region (STEM) to the recombination point with sequences of *Bam*HI-Z (//). In het DNA, the sequences to the left of STEM are the inverted repeat of the sequences to the right of STEM. Open reading frames, a possible promoter (⌐), and the exons mapped to the potentially polycistronic latent mRNAs of B95-8 EBV are represented above the sequences. Regions of het DNA open reading frames from the sequences of the palindrome located to the left of the STEM region are indicated (∿). Vertical lines between HR-1 *Bam*HI-W and B95-8 *Bam*HI-W, and HR-1 *Bam*HI-W and het DNA represent 1-bp differences. The two individual base-pair deletions in B95-8 *Bam*HI-W compared with HR-1 *Bam*HI-W are indicated (▼). No deletions or insertions were found between standard HR-1 *Bam*HI-W and sequences of *Bam*HI-W in het DNA.

codon is also seen in het DNA. As in standard HR-1 *Bam*HI-W, this change extended the open reading frame. In het DNA, the 5′ end of the HR-1 het BWRF1 open reading frame extended to the end of 387-bp stem-and-loop region. However, the palindromic rearrangement of het DNA deleted the region of standard HR-1 *Bam*HI-W containing the 5′ end of the HR-1 BWRF1 open reading frame. Thus, the single genomic rearrangement which resulted in the generation of the palindrome also resulted in different amino termini for both of these het DNA open reading frames compared with each other and with the standard HR-1 BWRF1. HR-1 het BWRF1 is 1,311 bp long, with 33 bp of the 5′ end unique to this open reading frame. HR-1 het BWLF1 is 981 bp long, with 90 bp of the 5′ end unique.

DISCUSSION

HR-1 clone 5 cells are enriched for one member of a family of defective EBV genomes present in the parental HR-1 cell line. The defective viral genome present in clone 5 cells contains several genomic rearrangements, including large duplications of the standard EBV genome (23). A particularly intriguing rearrangement is found in an *Eco*RI fragment of 16 kbp, designated *Eco*RI het 16, which is organized as a palindrome of *Bam*HI subfragments surrounding a central 1.2-kbp *Bam*HI subfragment. In addition to the palindromic rearrangement, two additional symmetric rearrangements are present on each side of *Eco*RI het 16, each joining sequences of the EBV genome which are not normally contiguous. One of these rearrangements, found in *Hin*dIII het 4.7, has joined sequences of *Bam*HI-W with sequences of *Bam*HI-Z. This rearrangement created a het DNA fragment which has been shown to be capable of disrupting latency (10; E. Grogan, H. Jenson, J. Countryman, L. Heston, L. Gradoville, and G. Miller, Proc. Natl. Acad. Sci. USA, in press).

Before the DNA was sequenced, *Eco*RI het 16 was known only to be a palindrome of DNA restriction endonuclease subfragments. The center of this extremely large palindrome present in defective EBV DNA has now been sequenced to determine the extent to which the palindrome is exact, to document its exclusive derivation from the standard HR-1 EBV genome, to identify possible open reading frames or regulatory sequences, and to provide clues to the molecular events responsible for the generation of the palindrome. The nucleotide sequence of the corresponding region of the standard HR-1 genome has been studied for comparison to identify the differences between standard HR-1 EBV and HR-1 het DNA at the level of the DNA sequence. This comparison is crucial in understanding the unique biologic activities associated with het DNA of HR-1 clone 5, such as its capacity to disrupt latency and to replicate independently (27, 28).

Problems in cloning a large palindrome into M13. Before it was sequenced, linear *Hin*dIII het 4.7 DNA was circularized before sonication to minimize overrepresentation of the ends of the linear fragment in the M13 library (2, 12). After accumulation of DNA sequence data, two unusual points became apparent. First, there were no M13 clones which contained an intact *Hin*dIII site. This was likely due to the in vitro construction of palindromic sequences around the *Hin*dIII site produced by self-ligation and the subsequent inability to clone these fragments into M13 vectors. Thus, there was overrepresentation of the linear ends of *Hin*dIII het 4.7 in the M13 library. Second, very few random clones contained DNA sequences from the unique central 341 bp. In fact, no clone was ever found to contain an insert which crossed from one side of the palindrome through the central 341 bp and into the complementary sequences on the other side. These two points exemplify the difficulty in cloning palindromic DNA into the M13 vectors. The DNA sequence through the central 341 bp was confirmed by sequencing specific M13 clones obtained by screening the random M13 library with a specific probe for the unique central region, clone turnaround (2) of random clones which had one end in the center of the palindrome, and subsequently cloning restriction endonuclease fragments of *Hin*dIII het 4.7 which were known to have one end in the unique central region.

Comparison of the large open reading frame in *Bam*HI-W DNA sequences from B95-8 EBV, standard HR-1 EBV, and HR-1 clone 5 het DNA. The largest open reading frame in B95-8 *Bam*HI-W, designated BWRF1 (1), is 1,149 nucleotides long. A product from this reading frame has not yet been characterized. Because of the 1-bp difference that changed the termination codon preceding BWRF1 in B95-8 to a tyrosine codon in HR-1 and a 1-bp insertion in HR-1 *Bam*HI-W that shifted the reading frame, HR-1 BWRF1 was extended 255 nucleotides at its amino-terminal end compared with that of B95-8 BWRF1 (Table 1).

The palindromic rearrangement of het DNA in this region resulted in the formation of two novel open reading frames in het DNA derived from standard HR-1 BWRF1. One of these open reading frames was in the rightward direction (HR-1 het BWRF1), and one was in the leftward direction (HR-1 het BWLF1). Because the normal 5′ end of standard HR-1 BWRF1 was not present in the het DNA genome, each of these open reading frames in het DNA had a unique amino-terminal end compared with each other and with standard HR-1 BWRF1 (Fig. 7 and Table 1). These two het DNA open

TABLE 1. Comparison of the *Bam*HI-W large open reading frame from B95-8 EBV, standard HR-1 EBV, and HR-1 het DNA

Virus stock	*Bam*HI-W open reading frame	Length (nucleotides)	Length (amino acids)	5'-End comparisons	3'-End comparisons
B95-8	B95-8 BWRF1	1,149	383		
Parental HR-1, HR-1 clone 5, HR-1 clone 16	HR-1 BWRF1	1,404	468	255-Nucleotide extension compared with B95-8 BWRF1	
Parental HR-1, HR-1 clone 5	HR-1 het BWRF1	1,311	437	Unique 33 nucleotides compared with HR-1 BWRF1	Comparable to 1,278 nucleotides of HR-1 BWRF1
Parental HR-1, HR-1 clone 5	HR-1 het BWLF1	981	327	Unique 90 nucleotides compared with HR-1 BWLF1	Comparable to 891 nucleotides of HR-1 BWRF1

reading frames also differed from each other in the number of amino acids of the carboxy-terminal end in common with standard HR-1 BWRF1. Thus, although HR-1 het BWRF1 and HR-1 het BWLF1 are largely derived from the single large open reading frame found in standard HR-1 *Bam*HI-W, these open reading frames are unique from each other and from the parent.

Exons of latent mRNAs from *Bam*HI-W sequences. Two separate exons from *Bam*HI-W, of 66 and 132 nucleotides, have been identified as part of a potentially polycistronic message in B95-8 EBV encoding EBNA1 (39). These exons in *Bam*HI-W have also been found in cDNA clones encoding EBNA1 and EBNA2 from the IB4 cell line (34), as well as additional latent messages in B95-8 cells (5) and Raji cells (6). These two exons are part of a group of latent transcripts which encode nuclear products variably designated EBNA4 (33a), EBNA5 (14), or LP (34).

In standard HR-1, the first of these two exons has the same DNA sequence as in B95-8. However, two of the 1-bp differences between B95-8 *Bam*HI-W and HR-1 *Bam*HI-W lie in the second exon (Fig. 7). Neither of these 1-bp changes appeared to alter a donor or acceptor site for mRNA splicing. However, both 1-bp changes resulted in translation of a different amino acid. By comparison of B95-8 to HR-1, a glycine was changed to a valine, and an arginine was changed to a threonine.

The region of *Bam*HI-W contributing to these latent messages was not present in *Eco*RI het 16. The recombination of sequences of *Bam*HI-Z to sequences of *Bam*HI-W in clone 5 het DNA was at thymidine residue 1 in the consensus eucaryotic promoter TATAAA which is located upstream of the EBV latent transcripts which have been mapped to *Bam*HI-W (34, 39, 43). This recombination therefore has resulted in the exclusion of this putative promoter and of the exons of latent mRNAs from *Eco*RI het 16.

Possible models for generation of the palindrome. The palindromic structure of het DNA in HR-1 clone 5, which extends for at least 16 kbp and possibly 56 kbp or more (23), is the largest palindromic structure identified in any biologic system. From the DNA sequence (Fig. 3) it appears that a single rearrangement within the outside 23 bp of the central 387-bp stem-and-loop region resulted in the generation of the palindrome. The exact point of this recombination cannot be precisely localized because the outside 23 bp of the 387-bp region are exact inverted repeats. However, it is evident from the DNA sequence that the rearrangement must have occurred between base pairs 2272 to 2294 and 2636 to 2658 of standard HR-1 *Bam*HI-W.

Since extensive regions of dyad symmetry were present even within the central unique region, a plausible hypothesis is that this region of dyad symmetry in standard *Bam*HI-W played a seminal role in the genesis of the palindrome. The palindrome could be the result of an error in DNA replication, perhaps involving a single-stranded DNA intermediate which duplexed to the opposite parent strand and was subsequently replicated.

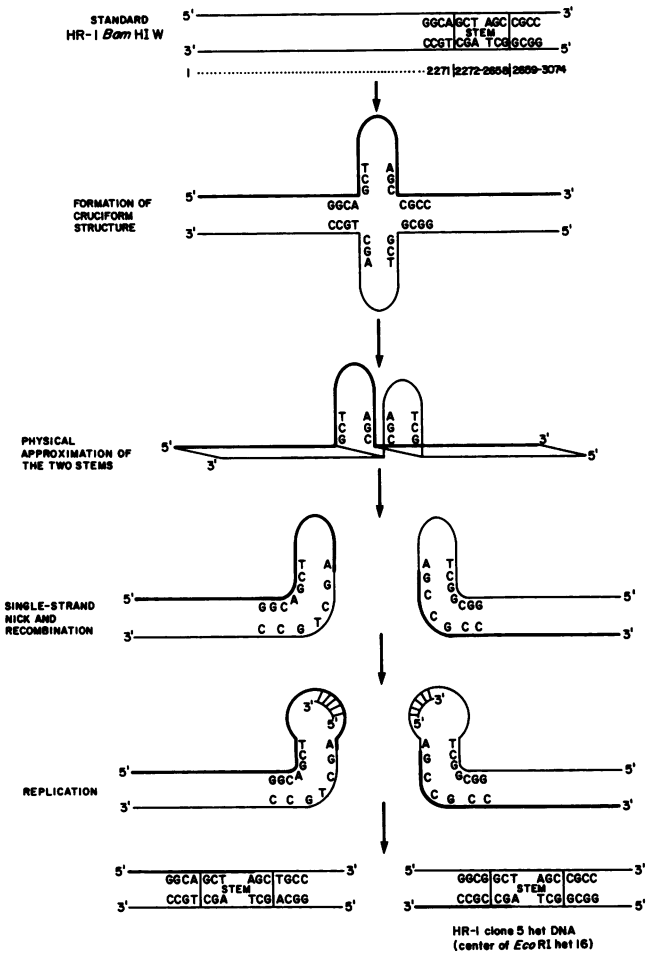


FIG. 8. One possible model for the generation of the palindrome of het DNA by recombination. The extensive dyad symmetry in standard HR-1 *Bam*HI-W would permit the formation of a cruciform structure. These stems could undergo recombination via a single-strand nick followed by strand exchange. After replication, possibly initiated within the single-stranded bubble, potentially two different palindromes are formed. One of these palindromic structures is *Eco*RI het 16. These two final products are identical to those shown in Fig. 9.

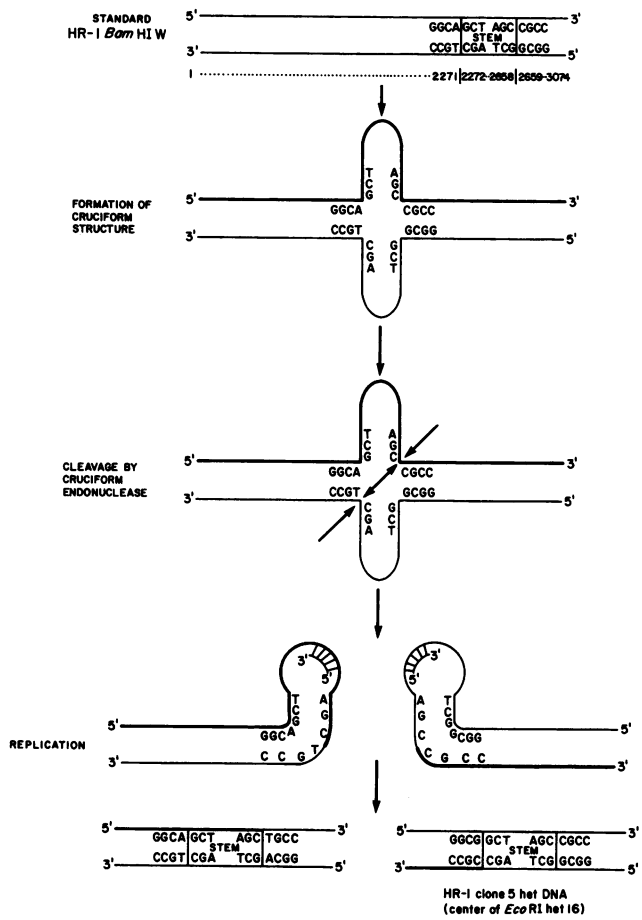


FIG. 9. One possible mode for generation of the palindrome by cleavage of the extruded cruciform structure. Cruciform endonucleases can cleave across either diagonal, but the two resulting products are the same after cleavage in either direction. After replication, possibly initiated within the single-stranded bubble, potentially two different palindromes are formed. One of these palindromic structures is *Eco*RI het 16. These two final products are identical to those shown in Fig. 8.

Other mechanisms which could have generated the palindrome include the capability of the dyad symmetry of standard *Bam*HI-W to form a cruciform structure in the native EBV genome. Palindromic sequences have been shown to be extruded in vitro in negatively supercoiled DNA (17, 25, 31). Such cruciform structures are stabilized by the release of torsional stress (17, 22, 25). After cruciform formation of the dyad symmetry in standard *Bam*HI-W, the homologous double-stranded stems would be capable of approximation and a recombination event initiated by a nick in a single DNA strand followed by strand exchange (Fig. 8). The resulting loop-and-stem structure could then be used as a template for synthesis of the complementary strand. If an origin of replication is within this loop, it could serve as the site of initiation of DNA synthesis. Regardless of which of the two strands of the loops are initially exchanged, the two sets of palindromic products produced by such a recombination are the same.

Another closely related mechanism for the generation of the palindrome in which the region of dyad symmetry in standard *Bam*HI-W is key would involve the novel activity of a cruciform nuclease, such as has been identified in *Saccharomyces cerevisiae* (42, 44) and bacteriophage T4 (31). These nucleases exhibit specificity for the extruded cruciform structure and cleave diagonally across a cruciform junction (Fig. 9). The site of cleavage is within the stem, usually within a few base pairs from the base of the cruciform arms. Cleavage of the cruciform structure in either direction would have generated the same two sets of palindromic products.

Besides the palindromic rearrangement, *Eco*RI het 16 contains two other symmetrical rearrangements. One joins sequences of *Bam*HI-W with sequences of *Bam*HI-Z, and the second approximates sequences of *Bam*HI-S with sequences close to the *Bam*HI site between *Bam*HI-E and *Bam*HI-c' (Fig. 1) (23). The finding of symmetrical rearrangements on each side of the unique central region suggests that these lateral rearrangements in het DNA occurred before the single rearrangement responsible for formation of the center of the palindrome. Then during the formation of the palindrome these lateral rearrangements were symmetrically duplicated. Alternatively, the recombinant mechanisms responsible for these lateral rearrangements would have to be repeated identically on each side of the palindrome to generate the fragment designated *Eco*RI het 16. Were this the case, one would need to postulate site-specific recombination mechanisms.

Fragments of het DNA with homology to the 387-bp stem-and-loop region which are present in parental HR-1 cells but not in HR-1 clone 5 cells. In both models, the two possible products of recombination would have an identical central region surrounded by either one of the two different sides of the EBV genome which flank this region. One of the two products would be identical to *Eco*RI het 16, in which sequences from the right side of *Bam*HI-W flank the stem-and-loop region in a palindromic arrangement. This fragment would have a center *Bam*HI digestion fragment of 1,225 bp, which is seen in clone 5 het DNA as *Bam*HI het 1.2 (Fig. 1). The second product would have a center *Bam*HI digestion fragment of 4,921 bp. It is intriguing that a *Bam*HI digestion fragment of approximately 5 kbp with homology to the 387-bp stem-and-loop region is detected in parental HR-1 cells, but not in HR-1 clone 5 cells (Fig. 1). Thus, among the various defective EBV genomes harbored by the HR-1 cell line may be another genome with a large palindrome centered around this 387-bp region but flanked by sequences from the left side of *Bam*HI-W. In fact, several defective *Bam*HI digestion fragments found in the parental HR-1 cell line have homology to the stem-and-loop region (Fig. 1), suggesting that this region has a relatively high propensity for genomic rearrangement.

Possible origin of replication within this region. The het DNA genome is not stably associated with cells and does not remain latent (28). Because the defective clone 5 genome is encapsidated, replicates as a virus, and is maintained horizontally in tissue culture by cell-to-cell spread (27), it is clear that het DNA must contain one or more origins of viral DNA replication.

An EBV origin capable of plasmid maintenance, *oriP*, has been localized to the *Bam*HI C fragment (45, 46). Sequences from *Bam*HI-C have been found in rearranged fragments of het DNA (28). *OriP* may also be present in the het DNA genome on a standard-sized digestion fragment which would not be identified as het DNA per se. However, the ability of *oriP* to act as an origin of replication of DNA destined for incorporation into virions has not been demonstrated. The extensive dyad symmetry found in the 387-bp stem-and-loop

region of standard *Bam*HI-W suggests that this region may function as an origin of DNA replication. A small segment of this region shares homology with the simian virus 40 and BK virus origins of replication and has previously been suggested as an origin of replication for EBV (8). This would be compatible with, but not necessary for, possible models for generation of the palindrome in heterogeneous EBV DNA (Fig. 8 and 9).

ACKNOWLEDGMENTS

We thank Mark Biggin for advice on DNA sequencing.

H.B.J. was supported by Physician Scientist Award AI-00651 from the National Institutes of Health and by a Swebilius Cancer Research Award from Yale University. This work was supported in part by Public Health Service grants AI-07210, AI-22959, and CA-12055 from the National Institutes of Health and by grant MV173G from the American Cancer Society.

LITERATURE CITED

1. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
2. **Bankier, A. T., and B. G. Barrell.** 1983. Shotgun DNA sequencing, p. 1–34. *In* R. A. Flavell (ed.), Techniques in nucleic acid biochemistry. Elsevier Scientific Publishers Ireland Ltd., County Clare.
3. **Biggin, M., P. J. Farrell, and B. G. Barrell.** 1984. Transcription and DNA sequence of the *Bam*HI L fragment of B95-8 Epstein-Barr virus. EMBO J. **3**:1083–1090.
4. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
5. **Bodescot, M., O. Brison, and M. Perricaudet.** 1986. An Epstein-Barr virus transcription unit is at least 84 kilobases long. Nucleic Acids Res. **14**:2611–2620.
6. **Bodescot, M., B. Chambraud, P. Farrell, and M. Perricaudet.** 1984. Spliced RNA from the IR1-U2 region of Epstein-Barr virus: presence of an open reading frame for a repetitive polypeptide. EMBO J. **3**:1913–1917.
7. **Bornkamm, G. W., J. Hudewentz, U. K. Freese, and U. Zimber.** 1982. Deletion of the nontransforming Epstein-Barr virus strain P3HR-1 causes fusion of the large internal repeat to the DS_L region. J. Virol. **43**:952–968.
8. **Cheung, A., and E. Kieff.** 1982. Long internal direct repeat in Epstein-Barr virus DNA. J. Virol. **44**:286–294.
9. **Cho, M.-S., G. W. Bornkamm, and H. zur Hausen.** 1984. Structure of defective DNA molecules in Epstein-Barr virus preparations from P3HR-1 cells. J. Virol. **51**:199–207.
10. **Countryman, J., and G. Miller.** 1985. Activation of expression of latent Epstein-Barr herpesvirus after gene transfer with a small cloned subfragment of heterogeneous viral DNA. Proc. Natl. Acad. Sci. USA **82**:4085–4089.
11. **Davis, D. B., and D. T. Kingsbury.** 1976. Quantitation of the viral DNA present in cells transformed by UV-irradiated herpes simplex virus. J. Virol. **17**:788–793.
12. **Deininger, P. L.** 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. **129**:216–223.
13. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
14. **Dillner, J., B. Kallin, H. Alexander, I. Ernberg, M. Uno, Y. Ono, G. Klein, and R. A. Lerner.** 1986. An Epstein-Barr virus (EBV)-determined nuclear antigen (EBNA5) partly encoded by the transformation-associated *Bam* WYH region of EBV DNA: preferential expression in lymphoblastoid cell lines. Proc. Natl. Acad. Sci. USA **83**:6641–6645.
15. **Fresen, K.-O., M.-S. Cho, and H. zur Hausen.** 1978. Heterogeneity of Epstein-Barr virus. IV. Induction of a specific antigen by EBV from two transformed marmoset cell lines in Ramos cells. Int. J. Cancer **22**:160–165.
16. **Fresen, K.-O., B. Merkt, G. W. Bornkamm, and H. zur Hausen.** 1977. Heterogeneity of Epstein-Barr virus originating from P3HR-1 cells. I. Studies on EBNA induction. Int. J. Cancer **19**:317–323.
17. **Gellert, M., K. Mizuuchi, M. H. O'Dea, H. Ohmori, and J. Tomizawa.** 1978. DNA gyrase and DNA supercoiling. Cold Spring Harbor Symp. Quant. Biol. **43**:35–40.
18. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
19. **Henle, W., G. Henle, B. A. Zajac, G. Pearson, R. Waubke, and M. Scriba.** 1970. Differential reactivity of human serums with early antigens induced by Epstein-Barr virus. Science **169**:188–190.
20. **Heston, L., M. Rabson, N. Brown, and G. Miller.** 1982. New Epstein-Barr virus variants from cellular subclones of P3J-HR-1 Burkitt lymphoma. Nature (London) **295**:160–163.
21. **Hinuma, Y., M. Konn, J. Yamaguchi, D. J. Wudarski, J. R. Blakeslee, Jr., and J. T. Grace, Jr.** 1967. Immunofluorescence and herpes-type virus particles in the P3HR-1 Burkitt lymphoma cell line. J. Virol. **1**:1045–1051.
22. **Hsieh, T., and J. C. Wang.** 1975. Thermodynamic properties of superhelical DNAs. Biochemistry **14**:527–535.
23. **Jenson, H. B., M. S. Rabson, and G. Miller.** 1986. Palindromic structure and polypeptide expression of 36 kilobase pairs of heterogeneous Epstein-Barr virus (P3HR-1) DNA. J. Virol. **58**:475–486.
24. **Jones, M. D., and B. E. Griffin.** 1983. Clustered repeat sequences in the genome of Epstein-Barr virus. Nucleic Acids Res. **11**:3919–3937.
25. **Lilley, D. M., G. W. Gough, L. R. Hallam, and K. M. Sullivan.** 1985. The physical chemistry of cruciform structures in supercoiled DNA molecules. Biochimie **67**:697–706.
26. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA of double-digest restriction fragments. Gene **19**:269–276.
27. **Miller, G., L. Heston, and J. Countryman.** 1985. P3HR-1 Epstein-Barr virus with heterogeneous DNA is an independent replicon maintained by cell-to-cell spread. J. Virol. **54**:45–52.
28. **Miller, G., M. Rabson, and L. Heston.** 1984. Epstein-Barr virus with heterogeneous DNA disrupts latency. J. Virol. **50**:174–182.
29. **Miller, G., J. Robinson, L. Heston, and M. Lipman.** 1974. Differences between laboratory strains of Epstein-Barr virus based on immortalization, abortive infection and interference. Proc. Natl. Acad. Sci. USA **71**:4006–4010.
30. **Mills, D. R., and F. R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA **76**:2232–2235.
31. **Mizuuchi, K., B. Kemper, J. Hays, and R. A. Weisberg.** 1982. T4 endonuclease VII cleaves Holliday structures. Cell **29**:357–365.
32. **Rabson, M., L. Gradoville, L. Heston, and G. Miller.** 1982. Non-immortalizing P3J-HR-1 Epstein-Barr virus: a deletion mutant of its transforming parent, Jijoye. J. Virol. **44**:834–844.
33. **Rabson, M., L. Heston, and G. Miller.** 1983. Identification of a rare Epstein-Barr virus variant that enhances early antigen expression in Raji cells. Proc. Natl. Acad. Sci. USA **80**:2762–2766.

33a. **Rowe, D. T., P. J. Farrell, and G. Miller.** 1987. Novel nuclear antigens recognized by human sera in lymphocytes latently infected by Epstein-Barr virus. Virology **156**:153–162.

34. **Sample, J., M. Hummel, D. Braun, M. Birkenbach, and E. Kieff.** 1986. Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a probable transcriptional initiation site. Proc. Natl. Acad. Sci. USA **83**:5096–5100.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5476.
36. **Schmitt, J. J., and B. N. Cohen.** 1983. Quantitative isolation of DNA restriction fragments from low-melting agarose by Elutip-d affinity chromatography. Anal. Biochem. **133**:462–464.
37. **Skare, J., J. Farley, J. L. Strominger, K. O. Fresen, M. S. Cho, and H. zur Hausen.** 1985. Transformation by Epstein-Barr virus

requires DNA sequences in the region of *Bam*HI fragments Y and H. J. Virol. **55**:286–297.
38. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
39. **Speck, S., and J. Strominger.** 1985. Analysis of the transcript encoding the latent Epstein-Barr virus nuclear antigen I: a potentially polycistronic message generated by long-range splicing of several exons. Proc. Natl. Acad. Sci. USA **82**:8305–8309.
40. **Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10**:4731–4751.
41. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14**:217–231.
42. **Symington, L. S., and R. Kolodner.** 1985. Partial purification of an enzyme from *Saccharomyces cerevisiae* that cleaves Holliday junctions. Proc. Natl. Acad. Sci. USA **82**:7247–7251.
43. **van Santen, V., A. Cheung, M. Hummel, and E. Kieff.** 1983. RNA encoded by the IR1-U2 region of Epstein-Barr virus DNA in latently infected, growth-transformed cells. J. Virol. **46**:424–433.
44. **West, S. C., and A. Korner.** 1985. Cleavage of cruciform DNA structures by an activity from *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **82**:6445–6449.
45. **Yates, J., 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.
46. **Yates, J., N. Warren, and B. Sugden.** 1985. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. Nature (London) **313**:812–815.