JOURNAL OF VIROLOGY, Sept. 1984, p. 776–787
0022-538X/84/090776-12$02.00/0

Vol. 51, No. 3

Cloning and Analysis of Integrated Hepatitis Virus Sequences from a Human Hepatoma Cell Line

YOSEF SHAUL, MARILYN ZIEMER, PABLO D. GARCIA, ROBERT CRAWFORD, HENRY HSU, PABLO VALENZUELA, AND WILLIAM J. RUTTER*

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Received 4 January 1984/Accepted 9 May 1984

We report here the isolation by molecular cloning and the analysis by heteroduplex and restriction enzyme mapping of seven distinct DNA fragments containing hepatitis B virus (HBV) sequences from genomic DNA of the PLC/PRF/5 human liver carcinoma cell line (the Alexander cell). No intact full-length HBV genomes were present. Three different patterns of organization of HBV fragments were detected. These included two linear fragments without detectable rearrangement, three other HBV fragments with internal deletions, and two HBV fragments containing long inverted duplications. HBsAg sequences are preferentially included in the integrated fragment, whereas the core gene is preferentially eliminated. Several of the integrated HBV fragments might act as templates for the synthesis of functional HBsAg mRNA, whereas only one clone could produce a full core antigen transcript.

Hepatitis B virus (HBV) is the etiological agent for a major form of hepatitis particularly prevalent in Asia and Africa. Epidemiological studies have revealed a striking correspondence between the areas of high incidence of HBV infection and the frequency of hepatocellular carcinoma (HCC). These observations led to the postulate that HBV and HCC are causally related (29). A number of subsequent observations are consistent with this hypothesis. (i) There are several animal viruses with biochemical and biophysical characteristics similar to HBV. One of these, the woodchuck hepatitis virus, induces HCC as a sequel to chronic hepatitis (27). (ii) HBV carrier patients are several hundredfold more at risk to HCC than control patients (1). (iii) HCC tissues often produce HBsAg. (iv) HBV sequences are frequently integrated into the host genome in HCC cells (4, 5, 7, 8, 12, 26) and in the DNA of carrier patients (3, 16).

These results support the further hypothesis that the integration of HBV sequences is associated with oncogenesis. We have examined the structure of the integrated HBV sequences in a primary liver carcinoma cell line (PLC/PRF/5, the Alexander cell) isolated from an African male with persistent HBV infection (18). Alexander cells synthesize and secrete 22-nm particles with the immunochemical characteristics of HBsAg (19); these cells also induce HBsAg-positive tumors in nude mice (11). However, core antigen (14), the HBeAg (30), viral DNA polymerase, and free virus (Dane particles) (12) are not detected. Genomic analysis shows that Alexander cells contain at least six copies of full or partial HBV genomes (7, 12, 20). We have isolated these integrated HBV sequences by molecular cloning in bacteria and determined their major structural features.

MATERIALS AND METHODS

Construction of the Alexander cell library. Alexander cell DNA was prepared as described previously (12) and partially digested with *Sau*3A. Fragments of 12 to 19 kilobases (kb) were isolated on a sucrose gradient (10 to 40% sucrose in 0.6 M NaCl, 1 mM EDTA, 10 mM Tris-hydrochloride [pH 7.6]; 39,000 rpm; 17 h; 4°C; Beckman SW41 rotor) and cloned into Charon 28 or 30 arms as described by Blattner et al. (2).

Construction of phage probes HSM and HSD. Two probes were constructed for use in heteroduplex mapping. The constructions were designed to contain HBV sequences in opposite orientations relative to the λ phage arms so that at least one of the probes would be able to hybridize to each of the Alexander library clones. Simian virus 40 (SV40) fragments were introduced between the HBV and phage arms to provide a region of known length which would always be single stranded in the heteroduplexes and therefore define the exact portion of the HBV genome in the isolated clones. One-half microgram of the *Eco*RI insert of HBV DNA from clone HBV3.2, 0.3 μg of SV40 digested with *Bam*HI and *Eco*RI, and 1 μg of Charon 28 arms containing *Bam*HI ends were ligated and packaged in vitro and used to infect DP50SupF by the method of Blattner et al. (2). Plaques containing HBV sequences were identified by hybridization. HSM and HSD were identified from the restriction maps of the hybridizing plaques.

Restriction mapping and genomic Southern blotting. Restriction enzyme digests of cloned DNA were electrophoresed on 1% agarose gels, blotted to nitrocellulose by the method of Southern (28), and hybridized to cloned HBV3.2 DNA (31) labeled with ^{32}P by nick translation (24). Restriction maps were determined by partial digestion or by comparing blots from single and double enzyme digestions. Detailed maps were prepared by analyzing fragments of the λ clones after subcloning into the plasmid vector pBR322. Specific fragments of HBV DNA were prepared as hybridization probes by digestion of cloned DNA with the appropriate enzymes and isolation of the specific fragments from agarose gels before nick translation. For genomic Southern blots, Alexander cell DNA was prepared, digested with *Eco*RI, electrophoresed, blotted, and hybridized as described previously (12).

Heteroduplex analysis. Heteroduplexes were prepared and mounted for electron microscopy by the formamide procedure of Davis et al. (10). Samples were examined in a Philips 300 electron microscope. Single-stranded M13 DNA and the length of the phage arms were used as single- and double-stranded DNA size markers, respectively. The DNA sizes in all of the figures derive from measurements of at least 10 different heteroduplexes. The standard errors ranged be-

* Corresponding author.

tween 5 and 10%, higher in general for single-stranded regions, especially in clones 6, 10, and 27 in which a complicated secondary structure is seen.

RESULTS

Molecular cloning of integrated HBV. Alexander cell DNA was cloned as a partial *Sau*3A digest because previous experiments had shown that limit digests with *Eco*RI, *Bam*HI, or *Hin*dIII contained some fragments too large to efficiently clone into λ phage. By using a partial digest of a frequent cutting restriction enzyme, we hoped to obtain the full repertoire of integrated HBV sequences. Alexander cell DNA was partially digested by *Sau*3A, and fragments of 12 to 19 kb were cloned into Charon 28 and 30 λ phage vectors. Eleven distinct clones were identified from 10^6 phage by hybridization with ^{32}P-labeled cloned HBV DNA. Restriction mapping analysis of the cloned DNAs revealed that clones 6, 10, and 14 as well as clones 5, 19, and 27 are overlapping. Thus, only seven distinct clones were identified from this Alexander cell genomic DNA library.

In earlier studies, we had shown that the Alexander cell line contains at least six distinct integrated copies of HBV DNA (12). Figure 1A shows a genomic (Southern) blot of Alexander cell DNA digested with *Eco*RI. We have compared the bands which hybridize to the HBV probe with hybridizing *Eco*RI fragments from the λ clones of our library (Fig. 1B) and can account for most of the Alexander cell *Eco*RI fragments. Four bands (A to D) apparently do not have counterparts in our set of clones. However, there are four clones, 8, 13, 15, and 27, in which one of the *Eco*RI sites bordering HBV sequences was not mapped due to interruption by the phage arms: clones 13 and 27 contain all of the HBV sequence which is present at these integration sites because cellular sequences flank each end of the HBV DNA (see below). In the two remaining clones, the HBV sequence is ligated directly to λ DNA so that part of the HBV sequence at these integration sites is not represented in our library. Therefore, we believe that we have accounted for all of the *Eco*RI fragments in the Southern blot and are lacking only one end of two integrated HBV fragments. The remaining five integration sites are fully represented. Cloning artifacts are unlikely since the *Eco*RI fragments of the genomic blot match the *Eco*RI fragments of the cloned DNAs.

Structural analysis of the clones. To identify specific HBV sequences in the clones, each cloned DNA was *Eco*RI-digested, and the fragments were resolved by electrophoresis on an agarose gel, blotted to nitrocellulose paper, and hybridized to ^{32}P-labeled cloned HBV genomic or subgenomic fragments (Fig. 1B). All of the clones contain at least one *Eco*RI site within the HBV sequence, and all hybridized to HBsAg and pre-HBsAg-specific probes (probes a and d). Only clones 13 and 14 and the related clones, 5, 19, and 27, hybridized to the core gene-specific probe (probe c).

These clones were further analyzed by electron microscopy of heteroduplexes and by mapping with restriction endonucleases. For heteroduplex analysis, we constructed Charon 28 phage probes which contain the entire HBV genome (HBV3.2; 31) flanked by SV40 fragments of unequal length (Fig. 2). The SV40 DNA is always single stranded in the heteroduplex and thus marks the ends of the phage arms. Since the integrated HBV might be positioned in two possible orientations, two different probes containing the HBV sequence in opposite orientation relative to the phage arms were employed. One of the constructions, named HSD, contained two tandem HBV sequences, so that integrated

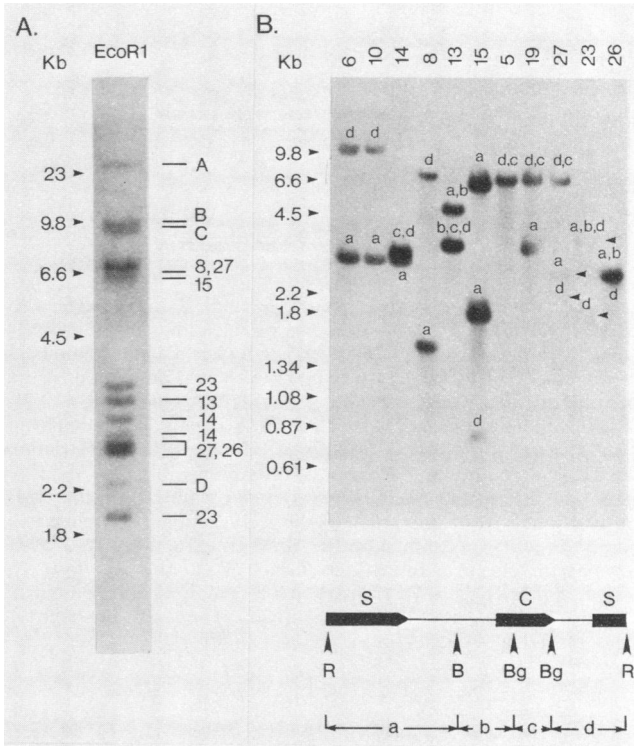

FIG. 1. (A) Assignment of the virus sequences in the Alexander genome to isolated clones. Alexander genomic DNA was digested by *Eco*RI enzyme, electrophoresed on a 1% agarose gel, blotted to nitrocellulose paper, and hybridized with ^{32}P-HBV3.2 DNA. Numbers at the right indicate the clones which contain corresponding *Eco*RI fragments. See text for details. Numbers to the left indicate the size in kilobases of standard DNAs. (B) *Eco*RI digestion pattern of the HBV sequence of the 11 genomic clones isolated from the Alexander cell library. Phage DNAs were cleaved by *Eco*RI enzyme, electrophoresed in a 1% agarose gel, blotted onto nitrocellulose paper, and hybridized with ^{32}P-labeled HBV3.2 DNA (28). In parallel, four similar blots were prepared, and each was hybridized with various probes. These probes are designated by the letters a, b, c, and d and were prepared by *Eco*RI (R), *Bam*HI (B), and *Bgl*II (Bg) digestion of HBV3.2 DNA, followed by gel purification of fragments and nick translation with ^{32}P. The restriction sites which define each probe and the HBV gene it represents (S, HBsAg; C, core) are shown at the bottom of the figure. The bands that hybridized to each probe and were detected after overnight exposure are indicated by the corresponding letter (a to d). The clone number is indicated above each lane. The bands of the clones 27 and 23 are very faint in this specific autoradiogram and are marked by arrows. Positions and sizes of *Hin*dIII-digested λ DNA and *Hae*III-digested φX DNA used as size markers are indicated. The band in the last lane (clone 26) represents two comigrating DNA fragments.

full-length and tandem repeated HBV sequences could be analyzed. With these probes it was possible to measure both the lengths and the approximate endpoints of the integrated viral sequences relative to the HBV transcription map. Standard errors were between 5 and 10%.

In some cases, the full HBV sequence could not be mapped by heteroduplex analysis with HSM because the hybrid regions could not extend beyond the *Eco*RI site of the HBV DNA (see Fig. 2B). These regions were mapped by restriction enzyme analysis. A composite restriction map for the HBV sequences in the Alexander cell was constructed by comparing the maps of the seven clones (Fig. 2A). This map most closely resembles that of HBV subtype ayw (13). Of 18 restriction sites monitored, 9 are present in both maps

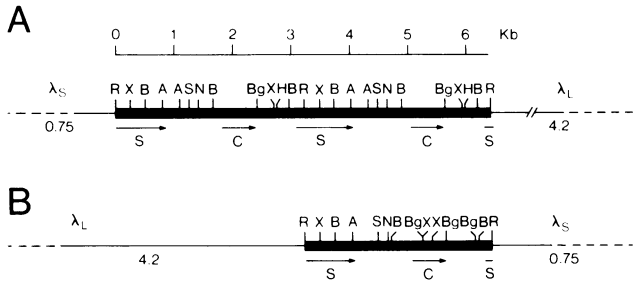

FIG. 2. Restriction maps of the heteroduplex probes, HSD (A) and HSM (B). HBV3.2 DNA is shown by thick lines, SV40 DNA is shown by thin lines, and λ DNA is shown by dashed lines. Numbers indicate the size in kilobases of SV40 DNA. The locations and direction of transcription of HBV genes, HBsAg (S) and core (C), are indicated by arrows. Note that these genes are in opposite orientation relative to the λ arms in HSM and HSD. A restriction map of our prototype obtained by comparing all of the HBV restriction sites in the Alexander cell clones is shown on the HSD construction. A restriction map of the HBV genome sequenced by Galibert et al. (13) is shown on the HSM construction. Restriction enzyme symbols are as follows: B, *Bam*HI; R, *Eco*RI; Bg, *Bgl*II; A, *Acc*I; S, *Sph*I; X, *Xba*I; H, *Hin*dIII; and N, *Nco*I. λ_L and λ_S indicate the phage long and short arms, respectively. The scale above part A marks the HBV sequence in kilobases.

and 9 are present in only one or the other. The restriction sites of the composite map are indicated relative to the HBV transcription map in Fig. 2A. The map of the sequenced HBV genome (13) is presented in Fig. 2B for comparison. Interestingly, some of the restriction sites in the composite Alexander cell map are absent from certain clones due to sequence heterogeneity (see below).

The application of these methods revealed that the HBV genome is present in three patterns in the Alexander cell line. These include linear sequences, HBV fragments containing long inverted duplications, and fragments with internal deletions. A schematic representation of the HBV sequences present in each clone is aligned with a dimer of the HBV transcription map in Fig. 3. Data for this figure were derived from the detailed analysis of each clone presented in Fig. 4 to 11 and discussed below.

Linear integration of the HBV genome. In two clones, 14 and 26, most of the HBV genome is integrated as a linear fragment without any rearrangement detectable by either electron microscopic analysis or restriction mapping. Restriction maps of the related clones 6, 10, and 14 are shown in Fig. 4. In clones 6 and 10, the short arms of the λ phages are ligated directly to HBV sequences so that only a portion of the integrated HBV is present in these clones. Clone 10 was used for analysis by electron microscopy because clone 14 contains more highly reiterated cellular DNA sequences, which interfere with the heteroduplex analysis. Electron micrographs of heteroduplexes with clone 14 showed many small hairpin structures in the DNA flanking the HBV hybrid region (data not shown). Further analysis revealed that this clone contains at least seven repeats from the Alu family (see discussion below). Figure 5A shows a typical electron micrograph of the heteroduplex between HSD (lower line in the schematic of Fig. 5A) and clone 10 (upper line). A 1.5-kb hybrid of HBV DNA is bordered on the left by a 3.3-kb, single-stranded loop-out structure and the phage short arm and on the right by open loops of 8.2 and 7.1 kb and the phage long arm. The 3.3-kb loop-out structure is due to HSD DNA comprised of 0.75-kb SV40 sequence and 2.55-kb HBV DNA. Thus, clone 14 consists of the phage short arm, which is directly ligated to 1.5-kb linear HBV sequence, followed by 8.2-kb human genomic DNA and the phage long arm. The HBV portion starts ca. 2.55 kb from the *Eco*RI site in the probe, that is, at the 3′ end of the core gene, and extends 1.5 kb to the 3′ portion of the HBsAg gene (Fig. 5B).

The region of clone 14 which corresponds to sequences to the left of this region determined by electron microscopy was mapped by restriction analysis. The *Bgl*II site at the 3′ end of the core gene is present in the appropriate position in clone 14 (Fig. 5B). However, this clone does not hybridize to the pre-core-gene-specific probe (probe b in Fig. 1B). We conclude that the HBV sequence terminates in the core gene. Therefore, this clone contains 1.5-kb linear HBV DNA which begins within the core gene (excluding the putative promoter region; 23), continues through the 5′ portion of the HBsAg gene and promoter (6, 23; D. N. Standring, W. J. Rutter, H. E. Varmus, and D. Ganem, submitted for publication), and ends within the 3′ sequence of the HBsAg gene. Thus, neither the HBsAg nor core genes are complete in clone 14.

The heteroduplex formed between clone 26 and HSM shows a linear hybrid region of 1.7-kb HBV DNA (Fig. 6A). An open loop structure comprised of two single-stranded DNAs of 3.3 and 4.2 kb is seen between the phage long arm and the 1.7-kb hybrid region. The 4.2-kb DNA should contain mostly or entirely the HSM SV40 sequence; therefore, the left end of the hybrid maps to the SV40-HBV junction in HSM, and the right end extends 1.7 kb further to very near the 5′ end of the core gene (Fig. 6B). The HBV sequence beyond the left end of this hybrid is not mapped by electron microscopic analysis. However, when the sequence in clone 26 to the left of the region mapped by the electron microscopic analysis is compared with the composite map (Fig. 2A), the sequence of restriction sites *Eco*RI, *Bam*HI, *Hin*dIII, and *Xba*I, which extend from the HBsAg gene toward the 3′ end of the core gene, can be identified. We conclude that in clone 26, the HBV sequence at the left end extends at least to the *Xba*I site located between the 3′ end

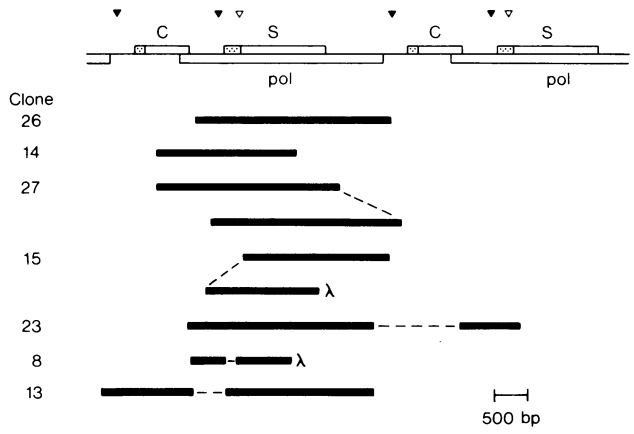

FIG. 3. Schematic representation of the HBV sequences present in all of the Alexander cell clones. The HBV sequence present in each clone is aligned with the HBV transcription map (top line) indicating the coding sequences for the mature HBsAg (S), core (C), and DNA polymerase (*pol*) genes (open boxes) and the precursor coding regions for HBsAg and core genes (stippled boxes). The promoter sites determined in vitro (solid triangles) and in vivo (open triangles) are indicated for the HBsAg and core genes. Dashed lines indicate positions in which nonconsecutive sequences of HBV DNA are joined in the Alexander cell. λ indicates positions at which the HBV sequence is ligated directly to phage DNA in the λ clones.

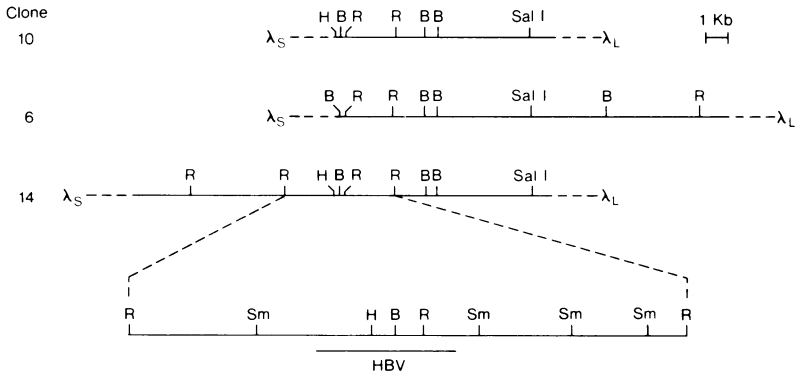

FIG. 4. Restriction endonuclease sites and location of HBV sequences in the overlapping clones 10, 6, and 14. A restriction map of the two *Eco*RI fragments containing the HBV sequence is shown. The region that hybridized to ^{32}P-HBV3.2 DNA probe is shown by the line designated HBV. Note that the phage short arm is ligated directly to HBV DNA in clones 6 and 10. Enzyme symbols are as in the legend to Fig. 2, with the addition of Sm for *Sma*I.

of the core gene and the 5′ end of the HBsAg gene. The HBV sequences continue through the HBsAg gene and extend on the right up to the 5′ end of the core gene. Thus, this HBV fragment consists of the viral genome minus the core gene.

Some integrated HBV sequences contain internal deletions. Three clones, 8, 23, and 13, contain HBV sequences with internal deletions (Fig. 7 to 9). A typical electron micrograph of the heteroduplex formed between HSD and clone 8 shows a hybrid of 1.4 kb next to a loop-out structure of 6.5 kb adjoining the phage long arm (Fig. 7A). This loop-out structure allows us to map the 1.4-kb hybrid from the 3′ end of the HBsAg gene to the middle of the core gene (Fig. 7C); this map is consistent with the fact that this clone hybridizes to HBsAg and pre-HBsAg-specific probes (probes a and d, Fig. 1B). The 6.5-kb loop-out structure also indicates that the phage long arm DNA is directly ligated to the integrated HBV sequences; therefore, we do not know the full dimensions of the integrated virus. The 1.4-kb hybrid region is interrupted by an "eye"-shaped loop structure of ca. 200 base pairs (bp) (Fig. 7A). This loop maps to the 5′ end of the HBsAg gene. In most of the observed heteroduplexes, this small loop hybridized to a portion of HBV sequence in HSD (Fig. 7B), indicating that the loop contains HBV sequence. Thus, clone 8 contains 14.7-kb human genomic DNA and 1.4-kb HBV sequence, which extends from the middle of the core gene to the 3′ end of the HBsAg gene, and is interrupted in the pre-HBsAg region by 200 bp of HBV DNA which is not colinear with the surrounding sequences (Fig. 7C).

The integrated HBV sequence in clone 23 was mapped with HSM as a heteroduplex probe. A hybrid region of 2.2-kb HBV DNA is interrupted by an internal loop-out structure of 1.1 kb (Fig. 8A). The measurements of the single-stranded DNAs flanking this hybrid allow us to map the integrated HBV sequence as shown in Fig. 8B; we conclude that the 1.1-kb loop-out structure is due to a deletion in the integrated virus. This deletion removes the entire core gene; this is consistent with the absence of hybridization to core-specific probe (probe c, Fig. 1B). In all of the electron micrographs examined, an additional small loop-out structure of ca. 100 bp is seen ca. 200 bp downstream of the big deletion (Fig. 8A). Sequence analysis has shown that this loop is due to a deletion in clone 23 (P. D. Garcia, Y. Shaul, and W. J. Rutter, manuscript in preparation). The right end of the integrated virus maps to the pre-HBsAg gene. The left end is not mapped by the HSM probe due to the limitation of the monomeric HBV DNA. However, restriction mapping analysis shows that the *Bgl*II site located 800 bp upstream from the unique *Eco*RI site in the Alexander cell HBV DNA (Fig. 2A) is present in clone 23 at the analogous position (Fig. 8B). In addition, this region hybridized to the HBV probe (Fig. 8B). Thus, clone 23 contains at least 800 bp of HBV sequence upstream from the HBsAg coding region mapped by electron microscopy. This clone does not hybridize to a core-specific probe (probe c, Fig. 1B) and therefore does not extend far beyond this *Bgl*II site. The final map of clone 23 reveals that the sequence upstream from the HBsAg gene, containing the *Bgl*II and *Xba*I sites, is directly repeated at the ends of the integrated HBV DNA. In summary, clone 23 contains 1.1- and 12-kb human genomic DNA separated by ca. 3-kb HBV sequence extending from the 3′ end of the core gene to the 5′ end of the HBsAg gene. The HBV DNA contains a deletion of 1.1 kb from the core gene and sequences 5′ to the core gene.

The HBV sequence in clone 13 was mapped by formation of heteroduplexes with HSD. In the electron micrograph, a hybrid region of 2.5 kb is flanked by open regions of 2.8 and 6.6 kb next to the phage long arm and 2.2 and 12.2 kb next to the phage short arm (Fig. 9A). The 6.6- and 2.2-kb DNAs contain the HSD sequence; therefore, both ends of the 2.5-kb hybrid region map close to the 5′ end of the core gene (Fig. 9B). The two 5′ ends of the core gene, however, are 3.2 kb apart in a dimer of prototype virus, whereas the length of the hybrid region in clone 13 is only 2.5 kb. The 0.7-kb difference and the presence of a loop-out structure of 0.65 kb in the 2.5-kb hybrid region show that there is a deletion in the HBV sequence. This conclusion is confirmed by the restriction enzyme cleavage patterns. The *Eco*RI-*Bgl*II fragment is 250 bp in clone 13, whereas the analogous fragment in the composite genome is 800 bp (compare Fig. 3A and 9B). This suggests that a segment of ca. 550 to 650 bp, extending from the 3′ end of the core gene to the 5′ end of the HBsAg gene, is deleted. Thus, the core and the surface genes may be incomplete in this clone. In summary, clone 13 contains 12.2- and 2.8-kb human genomic DNA flanking a monomeric HBV fragment which starts in sequences upstream from the core gene and ends virtually in the same position, but the fragment lacks the noncoding region between the 3′ end of the core gene and the 5′ end of the HBsAg gene.

Tandem integration of HBV sequences. The heteroduplex formed between HSD and clone 27 reveals a hybrid of 650 bp adjacent to a stem structure of 1.6 kb (Fig. 10A). A restriction enzyme analysis of clone 27 shows that the long inverted repeat which forms the stem in the electron microscopic structure is comprised of sequences containing the restriction sites characteristic of the HBsAg gene (compare Fig. 2B and 10B). Most of the 4.5-kb HBV sequence detected by

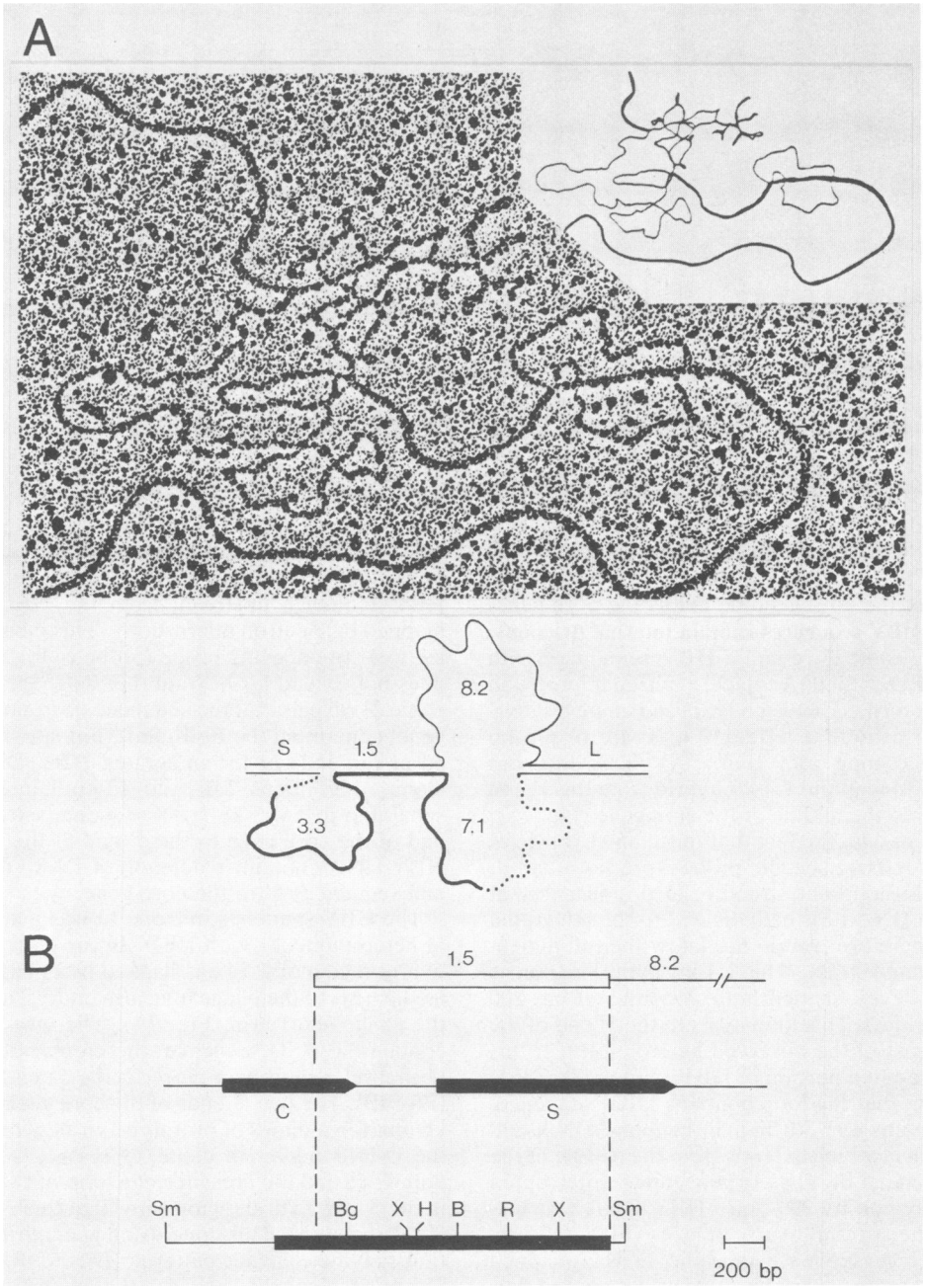

FIG. 5. Mapping of the HBV sequence in the overlapping clones 10 and 14. (A) Electron micrograph of a heteroduplex between clone 10 and the HSD probe. Insert is the interpretive drawing of the electron micrograph. Thin and thick lines represent single-stranded and double-stranded DNA, respectively. The schematic heteroduplex structure is shown below the electron micrograph. The HSD probe is drawn with dotted and thick lines representing the estimated location of the SV40 and HBV sequences, respectively. Numbers indicate the average sizes in kilobases of 15 different measurements. L and S indicate the phage long and short arms, respectively. (B) The restriction map of the *Sma*I DNA fragment from Fig. 4 (line 3, this figure) is aligned with the transcription map of HBV (line 2) and a schematic representation of the heteroduplex (line 1). Line 1, Open bars represent double-stranded regions of the heteroduplex. Thin lines signify single-stranded regions. Lengths of these regions are given in kilobases. Line 2, Sequences coding for the structural genes for HBsAg (S) and core (C) and the direction of transcription are indicated by thick arrows. Line 3, HBV DNA is represented by the thick line; host DNA is represented by the thin lines. Restriction enzyme symbols are as in the legends to Fig. 2 and 4. Vertical dashed lines indicate positions of alignment.

restriction mapping is contained in this stem structure. Both of the repeats comprising the stem structure extend from the 3′ end of the HBsAg gene through this gene to the *Xba*I site located just 5′ of the HBsAg gene. The repeat at the left of Fig. 10 extends through the intergenic region to the *Bgl*II site at the 3′ end of the core gene. This is the region detected as the 0.65-kb hybrid in the electron micrograph (Fig. 10A); it maps from 0.7 to 1.35 on the HBV map in HSD (Fig. 2A), i.e., at the 3′ portion of the core gene (Fig. 10B). This is consistent with the observation that the large *Eco*RI fragment present in clone 27 (Fig. 10B) hybridizes to core-specific probe (probe c, Fig. 1B). The right repeat does not appear to extend far beyond the *Xba*I site; this defines the end of the duplication and the beginning of the stem struc-

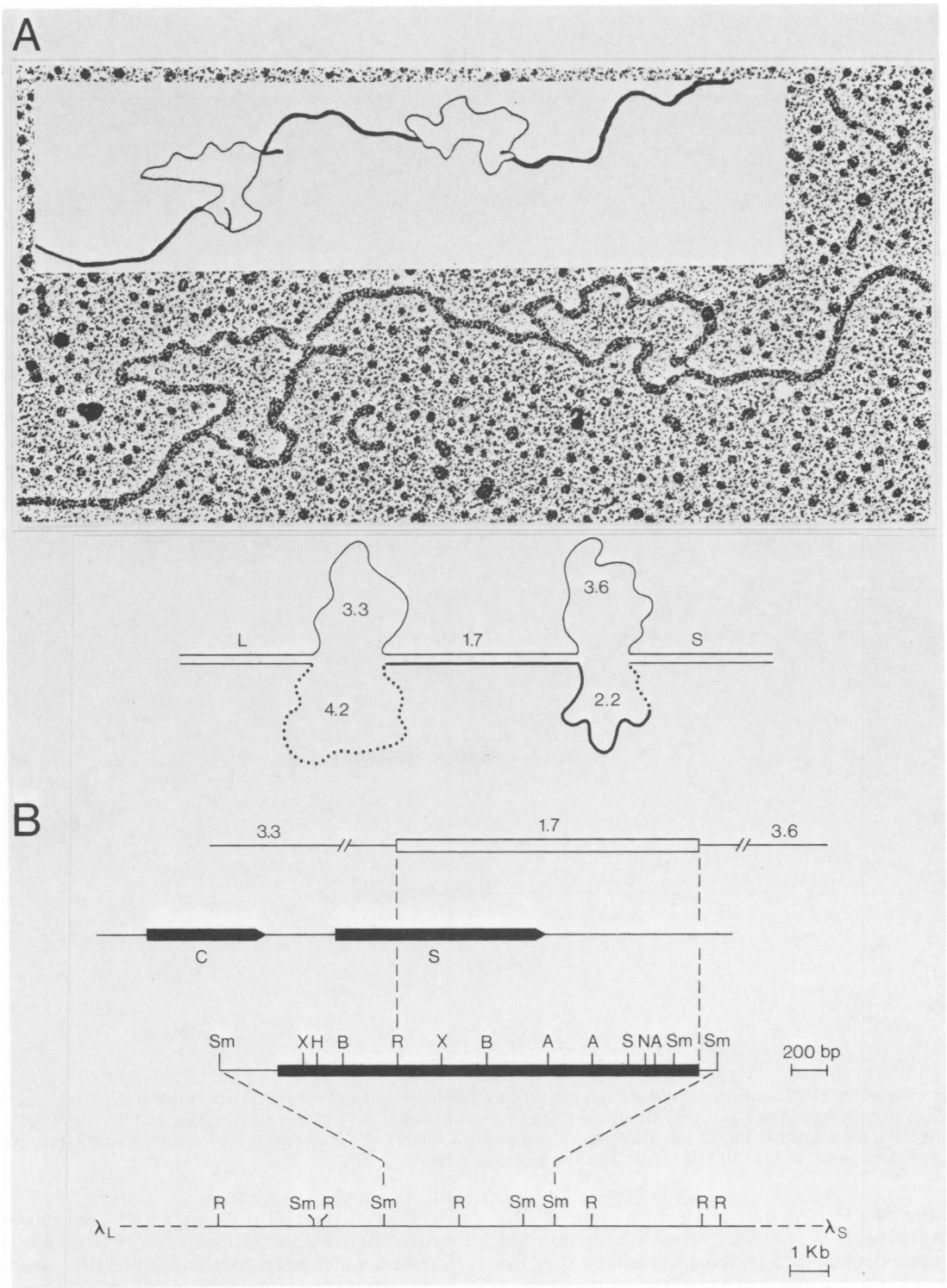

FIG. 6. Mapping of the HBV sequence in clone 26. (A) Electron micrograph of a representative heteroduplex between clone 26 and HSM and its interpretive drawing. See the legend to Fig. 5A for details. (B) The restriction map of the λ clone (line 4) is aligned with the map of two tandem *Sma*I fragments containing the HBV sequence (line 3), the transcription map of HBV (line 2), and a schematic representation of the heteroduplex (line 1). Line 4: thin lines, Alexander cell DNA; dashed lines, phage DNA. λ_L and λ_S indicate the phage long and short arms, respectively. Restriction enzyme symbols are as in the legends to Fig. 2 and 4. See the legend to Fig. 5B for an explanation of the symbols used in lines 1 to 3.

ture. The right repeat extends from the 3′ end of this site in the region between the core and HBsAg genes toward the *Nco*I and *Bam*HI sites near the 5′ end of the core gene. The left repeat contains only the sites up to the *Acc*I site at the 3′ end of the HBsAg gene. Therefore, the junction between the left and right repeats is in the *Bam*HI-*Acc*I fragment next to the *Nco*I site, and the loop structure seen in the electron micrograph is formed by the region between the *Acc*I sites at

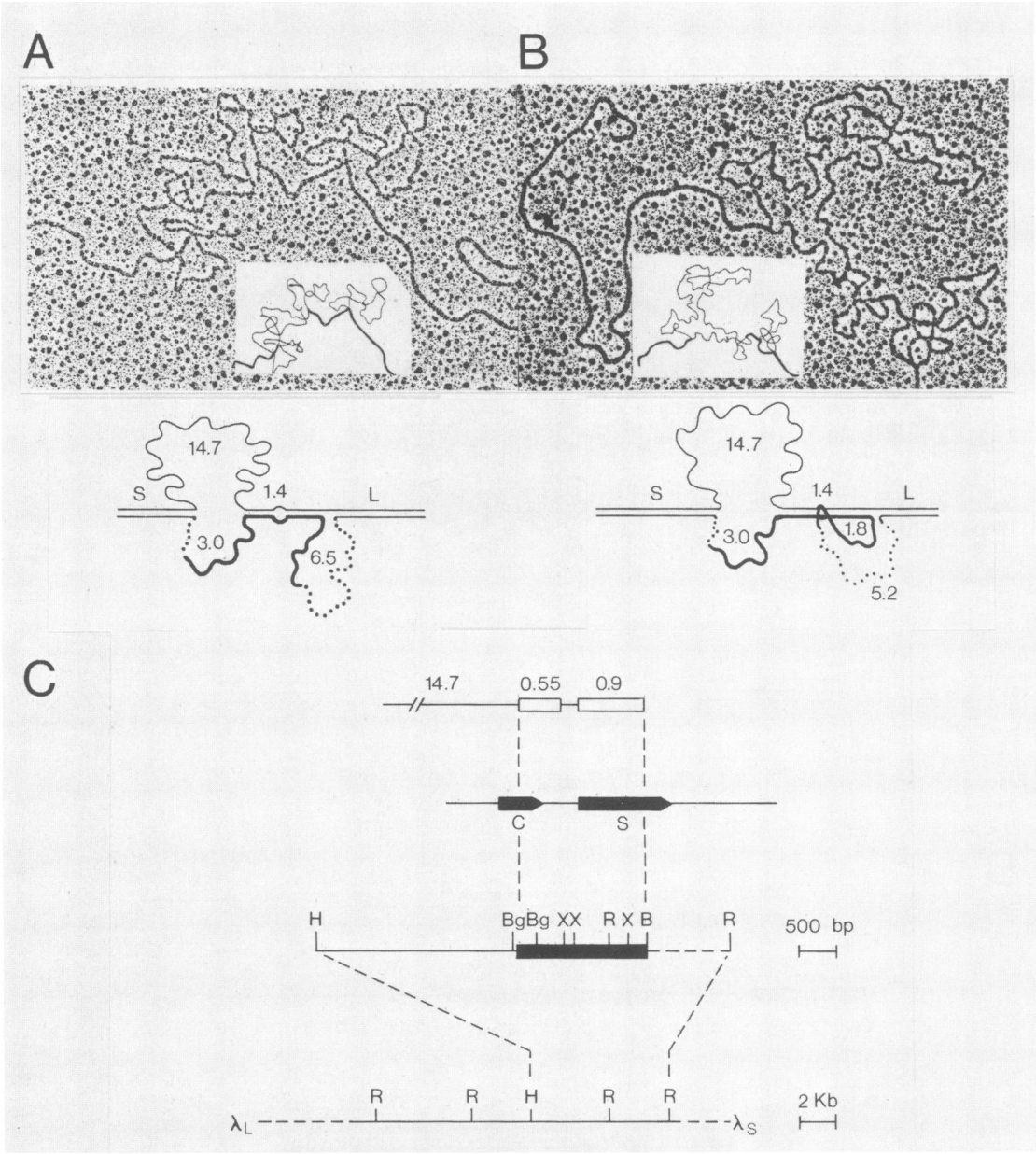

FIG. 7. Mapping of the HBV sequence in clone 8. (A and B) Electron micrograph of the two types of heteroduplexes observed between clone 8 and HSD. See the legend to Fig. 5A for details. (C) The restriction map of the λ clone (line 4) is aligned with the *Hin*dIII-*Eco*RI fragment containing the HBV sequence (line 3), the transcription map of HBV (line 2), and a schematic representation of the heteroduplex shown in A (line 1). See the legends to Fig. 5B and 6B for an explanation of symbols.

the ends of the two HBsAg genes. Thus, in clone 27 the entire HBsAg gene is represented twice as an inverted repeat, one copy containing additional sequences 5′ of the gene and the other extending past the 3′ end of the gene.

Clone 15 does not hybridize with the HBV region of the heteroduplex probes in either orientation. Thus, the HBV sequences in this clone are located in the stem loop structure (Fig. 11). The two external *Eco*RI fragments hybridize to a HBsAg-specific probe (probe a, Fig. 1B), and the central fragment hybridizes to a pre-HBsAg-specific probe (probe d, Fig. 1B; Fig. 11B). A characteristic pattern of *Eco*RI, *Xba*I, and *Acc*I restriction sites covering the HBsAg gene in the composite HBV is presented twice in inverted fashion in the fragments which hybridize to HBsAg gene-specific probe. Therefore, we conclude that the stem structure seen in the electron micrograph is due to an inverted repeat composed of two HBsAg gene fragments (Fig. 11B). The 850-bp *Eco*RI fragment which hybridizes to the pre-HBsAg-specific probe is ca. 300 bp larger than the loop structure seen in electron micrographs (550 bp); this implies that the inverted region extends at least 150 bp beyond the *Eco*RI site in the HBsAg gene. This *Eco*RI fragment contains the restriction sites characteristic of the pre-HBsAg gene region (*Eco*RI, *Bam*HI, *Hin*dIII, and *Xba*I; Fig. 2A). Thus, the left repeat extends to the *Xba*I site, whereas the right repeat terminates ca. 150 bp past the *Eco*RI site.

Host flanking sequences. During the electron microscopic analysis of the isolated clones, we found that the host flanking sequences of most of the clones display a complicated secondary structure. This can easily be seen as hairpin

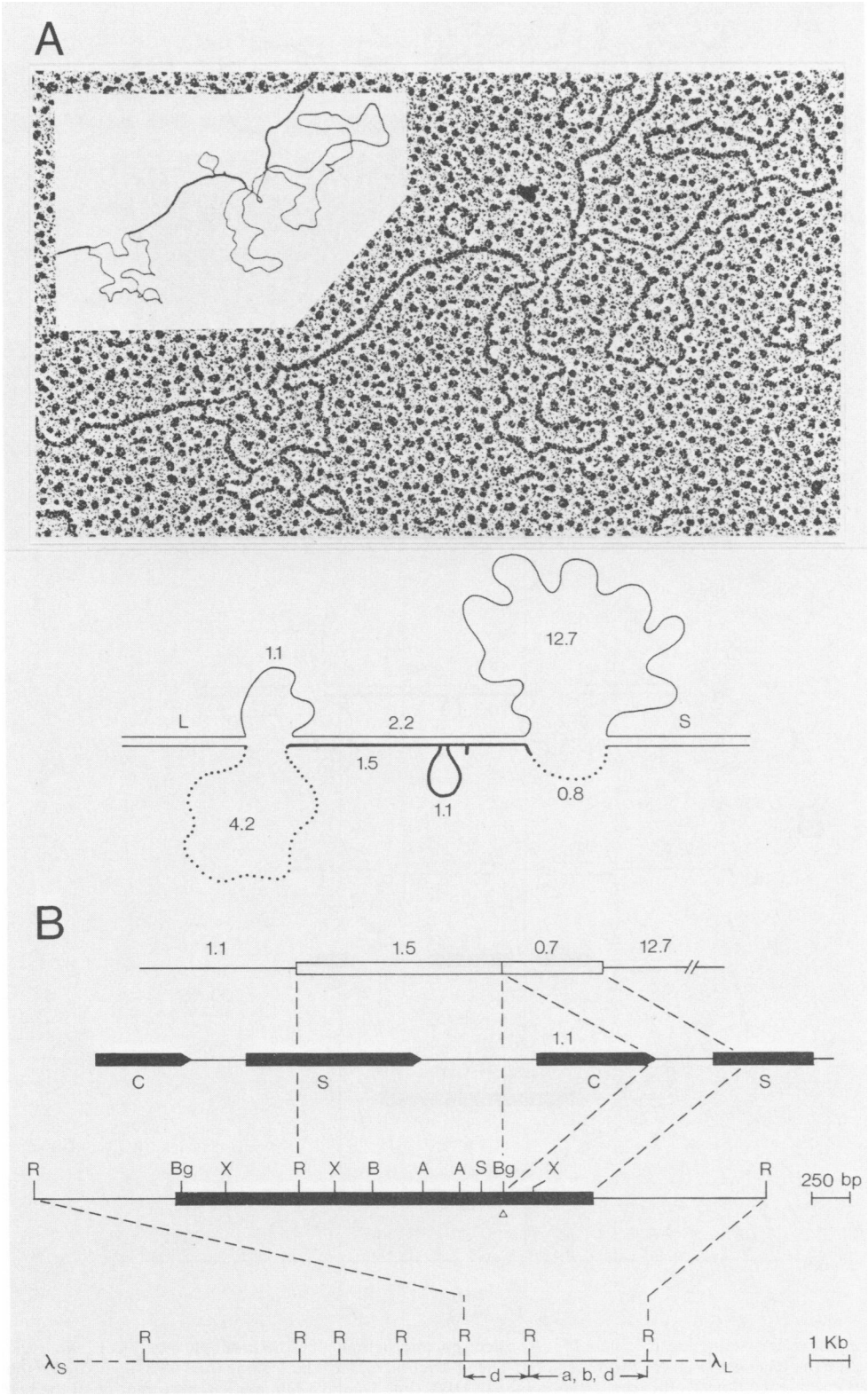

FIG. 8. Mapping of the HBV sequence in clone 23. (A) Electron micrograph of a heteroduplex between clone 23 and HSM and its interpretive drawing. See the legend to Fig. 5A for details. The small deletion of ca. 100 bp is located ca. 200 bp to the right of the big deletion. (B) The restriction map of the λ clone (line 4) is aligned with the two tandem *Eco*RI fragments which contain the HBV sequence (line 3), the transcription map of HBV (line 2), and a schematic representation of the heteroduplex (line 1). See the legends to Fig. 5B and 6B for an explanation of the symbols. The open triangle in line 3 indicates the position of the large deletion. The bars and letters under line 4 indicate the *Eco*RI fragments and the probes to which they hybridized in Fig. 2.

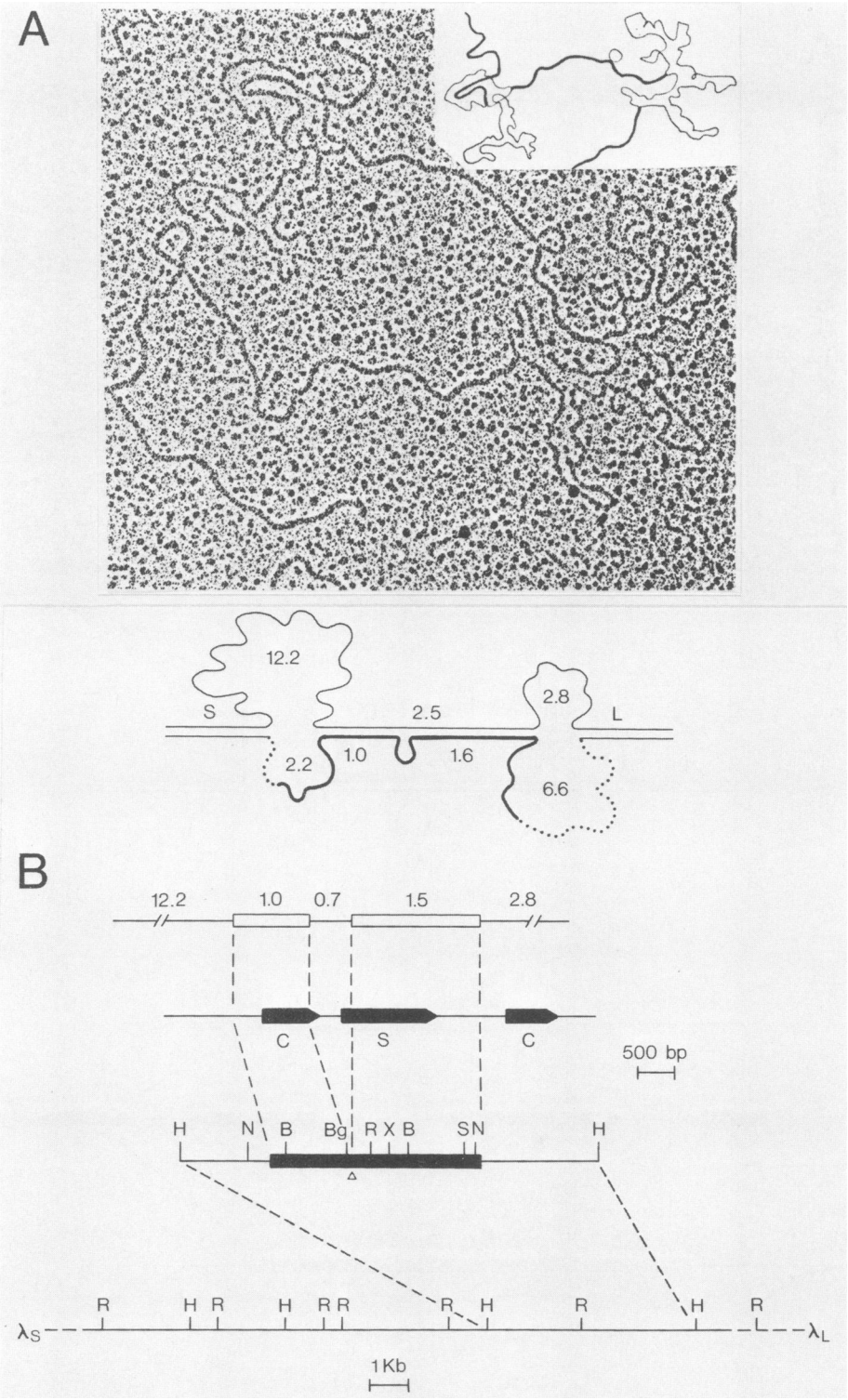

FIG. 9. Mapping of the HBV sequence in clone 13. (A) Electron micrograph of a heteroduplex between clone 13 and HSD and its interpretive drawing. See the legend to Fig. 5A for details. (B) The restriction map of the λ clone (line 4) is aligned with the *Hin*dIII fragment which contains the HBV sequence (line 3), the transcription map of HBV (line 2), and a schematic representation of the heteroduplex (line 1). See the legends to Fig. 5B and 6B for an explanation of the symbols.

structures in the host flanking sequences of clones 10 and 27 (Fig. 5 and 10, respectively). It appears that the HBV DNA is usually integrated into the host genome at regions that contain highly repeated sequences. To confirm this, we have tested the host flanking sequences for the presence of interspersed repeated sequences of the *Alu* family. Genomic clones were digested with *Alu*I, electrophoresed on an acrylamide gel, blotted to nitrocellulose paper, and hybridized to ^{32}P-labeled cloned DNA from the *Alu*I repeat family (Blur 8; 15) (data not shown). We found that several clones

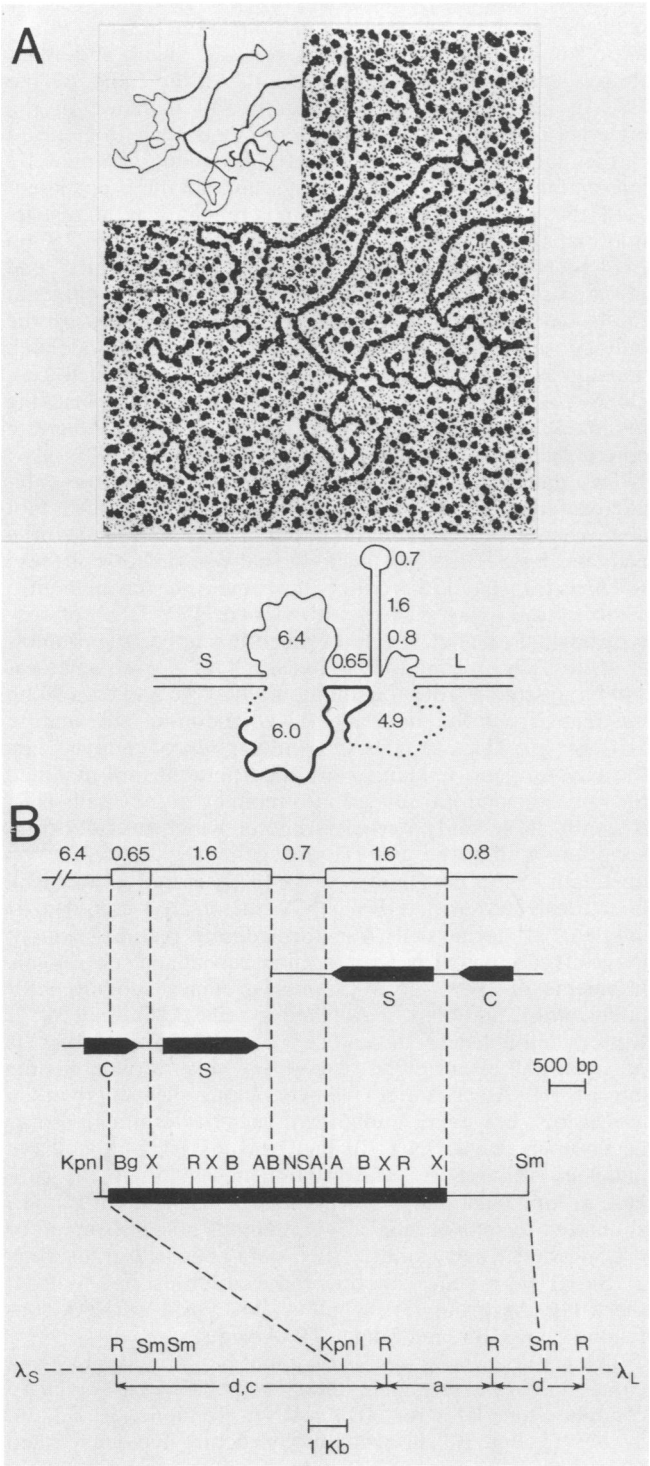

FIG. 10. Mapping of the HBV sequence in clone 27. (A) Electron micrograph of a heteroduplex between HSD and clone 27 and its interpretive drawing. See the legend to Fig. 5A for details. (B) The restriction map of the λ clone (line 5) is aligned with the *Kpn*I-*Sma*I fragment which contains the HBV sequence (line 4), the transcription map of HBV (lines 2 and 3), and a schematic representation of the heteroduplex (line 1). See the legends to Fig. 5B and 6B for an explanation of the symbols. The bars and letters under line 5 indicate the *Eco*RI fragments and the probes to which they hybridized in Fig. 2.

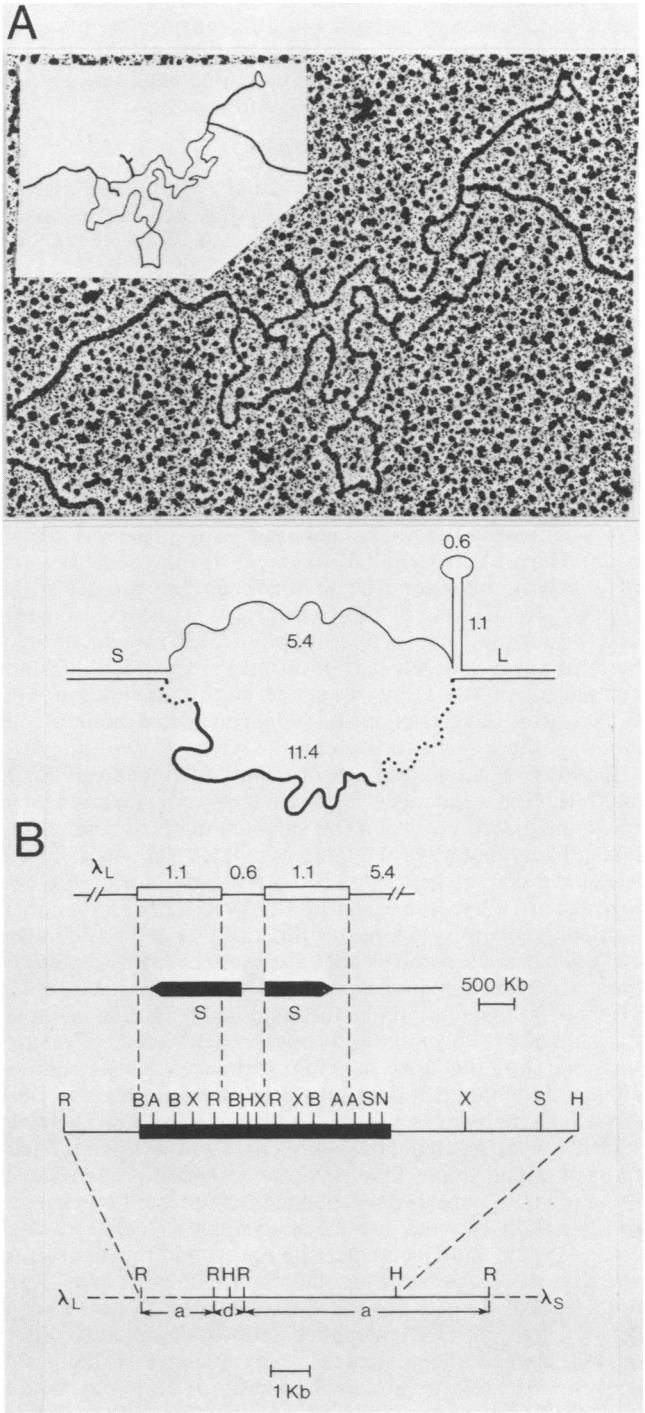

FIG. 11. Mapping of the HBV sequence in clone 15. (A) Electron micrograph of a heteroduplex between HSD and clone 15 and its interpretive drawing. See the legend to Fig. 5A for details. (B) The restriction map of the λ clone (line 4) is aligned with the *Eco*RI-*Hin*dIII fragment which contains the HBV sequence (line 3), the transcription map of HBV (line 2), and a schematic representation of the heteroduplex (line 1). See the legends to Fig. 5B and 6B for an explanation of the symbols. The bars and letters under line 4 indicate the *Eco*RI fragments and the probes to which they hybridized in Fig. 2.

contain a relatively large number of *Alu* repeats: clone 14 contains at least seven repeats, clones 26 and 27 contain four repeats each, clones 13, 15, and 23 contain two repeats each, and clone 8 contains at least one *Alu* repeat. These values are most probably an underestimation of the real numbers. A detailed restriction map of the entire clone would be needed to determine the exact number of *Alu* repeats.

DISCUSSION

We report here the isolation and analysis of seven distinct clones containing HBV sequences integrated into genomic DNA of the PLC/PRF/5 (Alexander) cell. While this manuscript was being written, two other laboratories reported the isolation of similar clones from Alexander cell DNA containing HBV sequences. Dejean et al. (10a) have reported the analysis of a clone containing an inverted repeat of the HBsAg gene region. A comparison of the HBV structure and restriction maps indicates that this clone is the same as our clone 27. Koshy et al. (16a) have identified three clones containing HBV DNA from a *Hin*dIII digest of Alexander cell DNA. One of these clones, containing a 6.0-kb *Hin*dIII fragment which hybridizes to HBV DNA, is like our clone 13. The restriction maps indicate that a second clone, containing a 10.7-kb *Hin*dIII fragment, is represented by the right half of our clone 26 (extending rightward from the *Hin*dIII site near the 5′ end of the HBV sequences; Fig. 6). The third clone, containing a 10.5-kb *Hin*dIII fragment, may be the same as the left half of our clone 15 which does not contain all of the HBV sequence at this integration site. Comparison of restriction maps indicates that none of our clones is the same as this clone.

In some viruses, e.g., SV40 and adenovirus (9, 25), multiple viral sequences may result by amplification of a single integrated virus and the surrounding host sequences. It is unlikely that all of the integrated HBV sequences in the Alexander cell are generated by amplification for the following reasons. First, the integrated HBV DNA is very stable. Ten independent subclones of this cell line exhibit identical HBV integration patterns (12). Furthermore, this pattern of integration has remained constant in DNA prepared over a 2-year period. Second, unlike the amplified SV40 and adenovirus sequences, no similarity is observed between the restriction maps of the host flanking sequences in the various clones. If amplification does occur, it must involve only short host sequences not detected by restriction mapping. Third, the integrated HBV sequences show different patterns of organization. Finally, there is sequence heterogeneity in the integrated HBV sequences as reflected by changes in restriction enzyme sites. For example, *Hin*dIII cleaves clones 14, 15, and 26 between the *Eco*RI and the *Xba*I sites, whereas the corresponding regions in clones 23 and 27 are not cleaved. This sequence heterogeneity has been confirmed by recent DNA sequence data which revealed differences between the clones in some regions of the HBV sequence (Garcia et al., in preparation). It is not known whether the changes are due to infection by multiple viruses or caused by mutagenesis subsequent to infection.

The study of the seven clones allows us to define some features of the integration of HBV. No intact full-length HBV copies are integrated into the Alexander cell genome; rather, three different patterns of HBV organization are observed in the clones; these include linear DNA, HBV sequences containing deletions of 100 to 1,000 bp, and large (1- to 1.6-kb) inverted repeats of HBV DNA (Fig. 3). We have now sequenced across all junctions between HBV and host DNA and across all deletions of HBV DNA. Our analysis confirms the conclusions drawn from electron microscopic and restriction mapping analysis and will be published in detail in future communications (Garcia et al., in preparation; M. Ziemer, P. Garcia, Y. Shaul, and W. J. Rutter, manuscript in preparation). All or most of the HBsAg gene sequence is preferentially included in the integrated HBV sequences of every clone. Only in clone 14 is the coding portion of the HBsAg gene disrupted by integration (Fig. 3). All of the clones contain the 5′ portion of the HBsAg gene. This observation is reflected in the restriction maps of the clones; all but clone 8 display the 250-bp *Eco*RI-*Xba*I fragment (Fig. 5 to 11) derived from the 5′ end of the HBsAg gene. Interestingly, the equivalent region and the B-gene region of the woodchuck hepatitis virus are the only conserved sequences among two different woodchuck hepatomas (21). In contrast, the core gene is excluded from clones 26, 15, 23, and 8, and the 5′ end of this gene is lacking in clones 14 and 27. Only clone 13 could contain the entire core gene.

Two distinct sizes (2.3 and 3.0 kb) of cytoplasmic polyadenylated RNA have been detected by Northern blot analysis in Alexander cells (7; unpublished data). Since these cells produce HBsAg, it is likely that one or both of these RNAs is translated to produce this protein. Using an in vitro transcription assay primed with cloned HBV DNA of subtype adw2, Rall et al. (23) have recently mapped the promoter of the HBsAg gene to a site near a TATAA sequence ca. 440 bp upstream from the unique *Eco*RI site and ca. 600 bp upstream from the initiating AUG codon of the mature HBsAg (23). This pre-HBsAg promoter appears active when the HBsAg gene sequences are incorporated in a modified SV40 vector and introduced into monkey kidney cells (17). Recently, a second internal promoter which initiates transcription within the pre-HBsAg gene region ca. 25 bp upstream of the *Eco*RI site has been described in a rat cell line transfected with HBV DNA, in infected chimpanzee liver (6), in murine cells transformed with recombinants of the Rous sarcoma virus long terminal repeat and subgenomic fragments of HBV DNA (Standring et al., submitted for publication), and also in Alexander cells (J. Ou and W. J. Rutter, unpublished data). The pre-HBsAg promoter is present in all of our clones except possibly 13, whereas the internal HBsAg promoter is in every clone. The length of the transcripts, however, is unknown because termination may occur in the host DNA. In fact, the in vivo studies have indicated that HBsAg transcription terminates in the core gene around nucleotides 1900 to 2000 (6, 22; Standring et al., submitted for publication). This region is absent from all of the Alexander cell clones. Thus, only abnormal transcripts of the HBsAg gene can be produced; transcription from these HBsAg promoters would yield hybrid mRNAs containing both HBV and human DNA sequences.

The active HBsAg mRNA(s) could be derived only from clones having a promoter and 850-bp HBV DNA. These candidate templates for HBV mRNA are clones 13, 23, 26, 27, and 15 (Fig. 3). The secondary structure in clones 27 and 15 is unlikely to be conducive to effective transcription. In contrast, clones 13, 23, and 26 contain the full HBsAg gene sequence in a simple intact linear structure. The pre-HBsAg promoter with the TATAA sequence is only present in clones 26, 27, 23, and 15. Of these, clone 23 contains a deletion in the pre-HBsAg gene region; clones 15 and 27 are subject to the effects of secondary structure mentioned above. Thus, clone 26 is the most likely candidate for HBsAg production from the promoter identified in vitro.

Core antigen coding sequences are absent from all clones except clone 13, and even in this clone, 3′ sequences of the core gene may be absent. The promoter identified in vitro

around nucleotide 1690 (23) is present. Thus, transcription from clone 13 could produce either the full core-coding sequence or a fusion product containing 3′ sequences from the HBsAg gene region. Sequence analysis shows that the core gene is not intact (Ziemer et al., in preparation). Thus, we predict that the Alexander cell cannot produce a normal core antigen molecule.

Inspection of the HBV DNA in the seven clones reveals that none of the clones contains sequences which code for the complete presumptive DNA polymerase gene (Fig. 3). This explains the absence of this activity from the Alexander cell.

The further characterization of these HBV sequences and their flanking host sequences may give insight into the process of integration and its effects on the host cell.

ACKNOWLEDGMENTS

We thank D. N. Standring for helpful discussions and L. Spector for preparation of the manuscript.

This research was supported by Public Health Service grant AI 19744 from the National Institutes of Health.

LITERATURE CITED

1. **Beasley, R. P., L. Y. Hwang, C. C. Lin, and C. S. Chein.** 1981. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22,707 men in Taiwan. Lancet **ii**:1129–1133.
2. **Blattner, F. R., A. E. Blechl, K. Denniston-Thompson, H. E. Faber, J. E. Richard, J. L. Slightom, P. W. Tucker, and O. Smithies.** 1978. Cloning human fetal gammaglobulin and mouse alpha-type globin DNA: preparation and screening of shotgun collections. Science **202**:1279–1284.
3. **Brechot, C., M. Hadchouel, J. Scotto, M. Fonck, F. Potet, G. N. Vyas, and P. Tiollais.** 1981. State of hepatitis B virus DNA in hepatocytes of patients with hepatitis B surface antigen-positive and -negative liver diseases. Proc. Natl. Acad. Sci. U.S.A. **78**:3906–3910.
4. **Brechot, C., C. Pourcel, M. Hadchouel, A. Dejean, A. Louise, J. Scotto, and P. Tiollais.** 1982. State of hepatitis B virus DNA in liver diseases. Hepatology **2**:27S–35S.
5. **Brechot, C., C. Pourcel, A. Louise, B. Rain, and P. Tiollais.** 1980. Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. Nature (London) **286**:533–535.
6. **Cattaneo, R., H. Will, N. Hernandez, and H. Schaller.** 1983. Signals regulating hepatitis B surface antigen transcription. Nature (London) **305**:336–338.
7. **Chakraborty, P., N. Ruiz-Opazo, D. Shouval, and D. Shafritz.** 1980. Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg producing human hepatocellular carcinoma cell line. Nature (London) **286**:531–533.
8. **Chen, D.-S., B. H. Hoyer, J. Nelson, R. H. Purcell, and J. Gerin.** 1982. Detection and properties of hepatitis B viral DNA in liver tissues from patients with hepatocellular carcinoma. Hepatology **2**:42S–46S.
9. **Clayton, C. E., and P. W. S. Rigby.** 1981. Cloning and characterization of the integrated viral DNA from three lines of SV40 transformed mouse cells. Cell **25**:547–559.
10. **Davis, R., M. Simon, and N. Davidson.** 1971. Electron microscopic heteroduplex methods for mapping regions of base sequence homology. Methods Enzymol. **21**:413–428.

10a. **Dejean, A., C. Brechot, P. Tiollais, and S. Wain-Hobson.** 1983. Characterization of integrated hepatitis B viral DNA cloned from a human hepatoma and the hepatoma-derived cell line PRF/PLC/5. Proc. Natl. Acad. Sci. U.S.A. **80**:2505–2509.

11. **Desmyter, S., M. B. Ray, A. F. Bradburne, and J. Alexander.** 1978. Human HBsAg-positive hepatoma in nude athymic mice. p. 459–463. *In* G. N. Vyas, S. N. Cohen, and R. Schmid (ed.), Viral hepatitis. Franklin Institute Press, Philadelphia, Pa.
12. **Edman, J. C., P. Gray, P. Valenzuela, L. B. Rall, and W. J. Rutter.** 1980. Integration of hepatitis B virus sequences and their expression in a human hepatoma cell. Nature (London) **286**:535–538.
13. **Galibert, F., E. Mandart, F. Fitoussi, P. Tiollais, and P. Charnay.** 1979. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in E. coli. Nature (London) **281**:646–650.
14. **Gerin, J. L., and J. W. K. Shih.** 1978. Structure of HBsAg and HBcAg. p. 147–153. *In* G. N. Vyas, S. N. Cohen, and R. Schmid, (ed.) Viral hepatitis. Franklin Institute Press, Philadelphia, Pa.
15. **Houck, C. H., F. P. Rinehart, and W. Schmid.** 1979. A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. **132**:289–306.
16. **Kam, W., L. Rall, E. Smuckler, R. Schmid, and W. J. Rutter.** 1982. Hepatitis B viral DNA in liver and serum of asymptomatic carriers. Proc. Natl. Acad. Sci. U.S.A. **79**:7522–7526.

16a. **Koshy, R., S. Koch, A. F. von Loringhoven, R. Kahmann, K. Murray, and P. H. Hofschneider.** 1983. Integration of hepatitis B virus DNA: evidence for integration in the single-stranded gap. Cell **34**:215–223.

17. **Laub, O., L. B. Rall, M. Truett, Y. Shaul, D. N. Standring, P. Valenzuela, and W. J. Rutter.** 1983. Synthesis of hepatitis B surface antigen in mammalian cells: expression of the entire gene and the coding region. J. Virol. **48**:271–280.
18. **Macnab, G. M., J. J. Alexander, E. M. Lecatsas, D. Bey, and J. M. Urbanowicz.** 1976. Hepatitis B surface antigen produced by a human hepatoma cell line. Br. J. Cancer **34**:509–515.
19. **Marion, P. L., F. H. Salazar, J. J. Alexander, and W. S. Robinson.** 1979. Polypeptides of hepatitis B virus surface antigen produced by a hepatoma cell line. J. Virol. **32**:796–802.
20. **Marion, P. L., F. H. Salazar, J. J. Alexander, and W. S. Robinson.** 1980. State of hepatitis B viral DNA in a human hepatoma cell line. J. Virol. **33**:795–806.
21. **Ogston, C. W., G. J. Jonak, C. E. Rogler, S. M. Astrin, and J. Summer.** 1982. Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from hepatocellular carcinomas of woodchucks. Cell **29**:385–394.
22. **Pourcel, C., A. Louise, M. Gervais, N. Chenciner, M.-F. Dubois, and P. Tiollais.** 1982. Transcription of the hepatitis B surface antigen gene in mouse cells transformed with cloned viral DNA. J. Virol. **42**:100–105.
23. **Rall, L. B., D. N. Standring, O. Laub, and W. J. Rutter.** 1983. Transcription of hepatitis B virus by RNA polymerase II. Mol. Cell. Biol. **3**:1766–1773.
24. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
25. **Sambrook, J., R. Green, J. Stringer, T. Mitchison, S. L. Hu, and M. Botchan.** 1980. Analysis of the sites of integration of viral DNA sequences in rat cells transformed by adenovirus or SV40. Cold Spring Harbor Symp. Quant. Biol. **44**:569–584.
26. **Shafritz, D. A., and M. C. Kew.** 1981. Identification of integrated hepatitis B virus DNA sequences in human hepatocellular carcinomas. Hepatology **1**:1–8.
27. **Snyder, R. L., and J. Summers.** 1980. Woodchuck hepatitis virus and hepatocellular carcinoma. p. 447–457. *In* M. Essex, E. Todaro, and H. Zur Hausen (ed.), Viruses in naturally occurring tumors. Cold Spring Harbor Conference on Cell Proliferation, vol. 7. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
29. **Szmuness, W.** 1978. Hepatocellular carcinoma and the hepatitis B virus: evidence for a causal association. Prog. Med. Virol. **24**:40–69.
30. **Takahashi, K., Y. Akahane, T. Gotanda, T. Mishiro, M. Imai, Y. Miyakawa, and M. Mayumi.** 1979. Demonstration of hepatitis B e antigen in the core of Dane particles. J. Immunol. **122**:275–279.
31. **Valenzuela, P., M. Quiroga, J. Zaldivar, P. Gray, and W. J. Rutter.** 1980. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. p. 57–70. *In* B. Fields, R. Jaenisch, and C. F. Fox (ed.), Animal virus genetics. Academic Press, Inc., New York.