JOURNAL OF VIROLOGY, Nov. 1988, p. 4224–4231
0022-538X/88/114224-08$02.00/0

Vol. 62, No. 11

# Structural Analysis of a Hepatitis B Virus Genome Integrated into Chromosome 17p of a Human Hepatocellular Carcinoma

YI-ZHONG ZHOU,[1]† BETTY L. SLAGLE,[1] LAWRENCE A. DONEHOWER,[1] PETER VANTUINEN,[2] DAVID H. LEDBETTER,[2] AND JANET S. BUTEL[1]*

*Department of Virology and Epidemiology[1] and Institute for Molecular Genetics,[2] Baylor College of Medicine, Houston, Texas 77030*

Received 13 June 1988/Accepted 30 July 1988

**Hepatitis B virus (HBV) is clearly a factor in the development of hepatocellular carcinoma, but its mechanism of action remains obscure. One possibility is that the HBV integration event alters the expression of a nearby growth-regulatory cellular gene. A 9-kilobase (kb) DNA fragment containing an HBV insert plus flanking cellular sequences was cloned from a hepatoma specimen from Shanghai, People's Republic of China. Restriction mapping of the insert revealed a large inverted repeat structure consisting of both viral sequences (encompassing all of the core and pre-S regions and portions of the X and S genes) and at least 3 kb of unique cellular sequences. The virus-cell junction mapped 11 nucleotides from the DR1 region, in a position within the HBV X gene and included in the cohesive overlap region. A probe generated from 1.0 kb of the flanking cellular DNA mapped the viral insert to chromosome 17 in the region designated 17p11.2-17p12, which is near the human proto-oncogene p53. Sequence data from a portion of the flanking cellular DNA revealed a stretch of approximately 70 base pairs that showed highly significant homology with a conserved region of a number of functional mammalian DNAs, including the human autonomously replicating sequence 1 (*ARS1*).**

Epidemiologic studies have demonstrated a strong association between chronic hepatitis B virus (HBV) infection and the development of human hepatocellular carcinoma (HCC). In areas of high risk such as Africa and Southeast Asia (1), HCC samples often contain HBV DNA sequences integrated into the high-molecular-weight tumor DNA (12). Integration of viral DNA is not a required step in the replicative cycle of the virus, so the presence of integrated HBV sequences in many HCCs is suggestive of a causal relationship between viral DNA integration and subsequent tumor development.

The precise mechanism by which HBV integration might contribute to the formation of a hepatoma is unclear. Current data support several different models. One hypothesis proposes viral integration next to an important cellular gene; in previous studies investigators observed integrations that resulted in hybrid virus-cell transcripts (31, 47) as well as integration within a putative steroid receptor gene (9). However, these findings represent a minority of hepatomas studied. More commonly observed are HBV integration-associated chromosomal abnormalities, including large deletions (14, 35), duplications (28, 42, 51), and chromosomal translocations (14, 42). Since changes in DNA structure are frequently associated with cancer, it is essential to characterize additional HBV insertions and surrounding cellular DNA sequences to discern patterns that might explain HBV involvement in the formation of different subsets of tumors.

Previous work in our laboratory focused on the analysis of a group of 17 HCC specimens obtained from Shanghai, People's Republic of China. Of 17 patients, 15 showed serologic evidence of past or present infection with HBV, and tumors from 13 of those 15 HBV-positive individuals contained integrated HBV sequences (52). We have now further characterized one tumor from this group. A 9-kilobase (kb) DNA fragment containing HBV DNA plus flanking cellular DNA was cloned from Chinese tumor no. 7. Restriction mapping demonstrated that the entire 9-kb insert consisted of a large inverted repeat structure containing both viral and flanking cellular DNAs. A probe generated from the flanking cellular DNA was used to map the HBV insert to the short arm of chromosome 17 in region 17p11.2-17p12. Sequence data from a portion of the flanking cellular DNA revealed a region of approximately 70 base pairs (bp) conserved in the noncoding portions of a number of functional mammalian genes.

## MATERIALS AND METHODS

**Tissue samples and preparation of tumor DNA.** Hepatoma tissue from Shanghai, People's Republic of China, was collected and stored as described previously (52). DNA was purified from hepatoma no. 7 (obtained from patient no. 5) by conventional procedures as described previously (52).

**Probe DNA.** The HBV DNA probes were generated from a cloned head-to-tail dimer of HBV (subtype *adw*) obtained from William Robinson (40). The subgenomic probes consisted of the following: S gene (1.2-kb *Bam*HI fragment); C gene (0.4-kb *Bgl*II fragment); X region (0.6-kb *Bam*HI-*Bgl*II fragment, containing a small amount of 5′ C gene-specific DNA); and X gene (277-bp *Hha*I fragment [52], extending from bp 1401 to 1678 on the *adr* map [17, 45]). All DNAs were verified by agarose gel electrophoresis, excised, electroeluted, and concentrated with an Elutip-d column (Schleicher & Schuell, Inc., Keene, N.H.). DNAs were labeled to a high specific activity with [$^{32}$P]dCTP (ICN Radiochemicals, Irvine, Calif.) and commercially available labeling kits (nick translation reagent kit [Bethesda Research Laboratories, Inc., Gaithersburg, Md.] or oligolabeling kit [Pharmacia, Inc., Piscataway, N.J.]).

**Cloning, subcloning, and sequencing of viral insert DNA.** High-molecular-weight DNA purified from tumor no. 7 was digested with *Eco*RI and fractionated on a 10 to 40% sucrose gradient, and the HBV-containing fractions were identified by Southern blot hybridization with an HBV S gene-specific

* Corresponding author.
† Present address: Shanghai Institute of Biochemistry, Academia Sinica, Shanghai 200031, People's Republic of China.

probe. HBV-positive fractions were pooled, and the DNA was ligated into lambda vector EMBL4 (Stratagene Cloning Systems, La Jolla, Calif.). Vector DNA was packaged with a Gigapack packaging kit (Stratagene Cloning Systems). HBV-positive plaques were identified by colony hybridization with an HBV S gene-specific probe. Six positive clones were obtained from $3 \times 10^6$ plaques. Positive plaques were picked, purified three times, and analyzed. Recombinant plaque no. 231 was subcloned into a pBR322-derived vector, pWR33 (13). Plasmid pT7/9.0 was used for the studies reported here.

Two fragments were subcloned and subjected to DNA sequencing analysis. The 2.1-kb *Bgl*II fragment, containing HBV plus flanking cellular DNA, was inserted into plasmid pSP73 (Promega Biotec, Madison, Wisc.), which contained SP and T7 promoters adjacent to the insertion site. Oligonucleotides complementary to these promoters (synthesized by O.C.S. Laboratories, Denton, Tex.) were annealed to a denatured plasmid template and extended by using dideoxy sequencing protocols with the Sequenase sequencing kit (U.S. Biochemical Corp., Cleveland, Ohio). The 1.0-kb *Eco*RI-*Bgl*II fragment, containing flanking cellular DNA, was subcloned into pWR33 (described above), and the DNA sequence was determined by primer extension and dideoxy sequencing protocols.

**Cell hybrids.** A somatic cell hybrid regional mapping panel for human chromosome 17 has been previously described (46). Hybrids were constructed from cells derived from patients with a variety of constitutional deletions or translocations of chromosome 17. Spacing of breakpoints represented in this panel is especially close on the short arm of 17, in which we estimate that from 2 to 7 megabases of DNA reside between neighboring breakpoints. Hybrid HO-11 (P. vanTuinen, W. B. Dobyns, D. C. Rich, K. M. Summers, T. J. Robinson, Y. Nakamura, and D. H. Ledbetter, Am. J. Human Genet., in press) contains a deletion in 17 [del(17)(p13.100)] from a fetal demise with a DiGeorge sequence (F. Greenberg, K. B. Courtney, R. A. Wessels, J. Huhta, R. J. Carpenter, D. C. Rich, and D. H. Ledbetter, Am. J. Med. Genet., in press).

**Colony and Southern blot hybridizations.** Colony hybridization was performed by standard procedures (23). Southern blot prehybridization and hybridization conditions were as described previously (52). Mapping studies were done with cellular DNAs bound to nylon filters (Bethesda Research Laboratories). Prehybridization incubations were done for 18 h at 68°C in 1× SSPE (180 mM NaCl, 10 mM sodium phosphate, 10 mM EDTA [pH 7.0]) containing 1% sodium dodecyl sulfate, 1% nonfat dry milk, and 500 μg of calf thymus DNA per ml. Hybridizations were performed for 18 h at 68°C in prehybridization solution containing $10^5$ cpm of $^{32}$P-labeled probe DNA per ml. Bound probe DNAs were stripped from the nylon filters by procedures described by Bethesda Research Laboratories so that sample DNAs could be retested with different labeled probes.

## RESULTS

**Identification and restriction mapping of the cloned 9-kb HCC DNA fragment containing integrated HBV sequences.** The DNA of hepatoma no. 7 contains two inserts of HBV sequences integrated into high-molecular-weight DNA (52). *Eco*RI was used to generate the library of genomic tumor DNA because this enzyme does not cleave HBV subtype *adr* (49), which is the most prevalent type in the People's Republic of China, and thus would yield DNA fragments containing the HBV genome as well as flanking cellular DNA sequences. Restricted DNA was packaged, and subsequent lambda plaques were screened by colony hybridization with a $^{32}$P-labeled HBV genomic probe. Six positive plaques were identified, and all contained insert DNA of 9 kb. One was subcloned into plasmid DNA for subsequent analysis. Numerous attempts to clone the second insert DNA from tumor no. 7 were unsuccessful. A 6.7-kb fragment containing HBV sequences was detected in lambda plaques, but the HBV DNA was invariably lost during subcloning procedures.

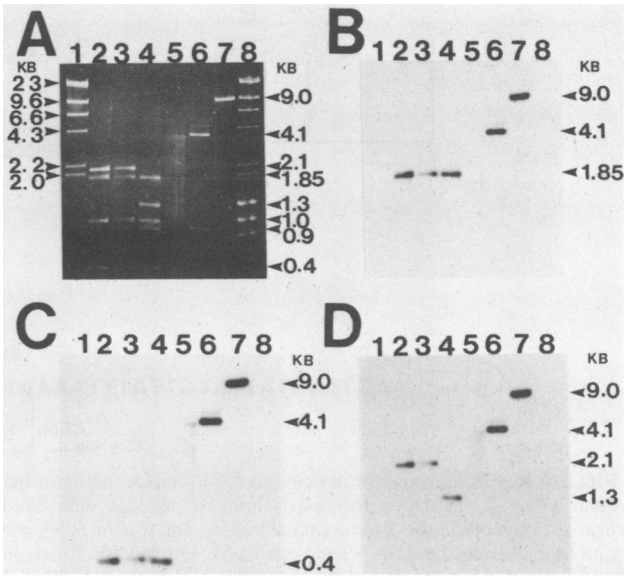


FIG. 1. Southern blot hybridization analysis of the 9.0-kb insert DNA from HCC no. 7. The 9-kb insert DNA was analyzed undigested (lanes 7) or after digestion with *Bgl*II (lanes 2 and 3), *Bgl*II and *Hpa*I (lanes 4), *Kpn*I (lanes 5), or *Hin*dIII (lanes 6). After agarose gel electrophoresis, separated DNA was analyzed by Southern blot hybridization as described in Materials and Methods. Shown are the ethidium bromide-stained gel (A) and autoradiograms after hybridization with HBV subgenomic probes: S gene (B), C gene (C), and X region (D). Molecular weight markers consisted of *Hin*dIII-digested lambda DNA alone (lanes 1) or mixed with *Hae*III-digested ϕX174 replicative-form DNA (lanes 8). The DNA in lanes 5 was degraded in this experiment and did not yield informative results. KB, Kilobases.

To facilitate interpretation of the potential functional significance of the HBV DNA integrated in tumor no. 7, we defined the extent of HBV sequences within the total cloned 9-kb DNA. For these mapping studies, the cloned DNA fragment was released from the plasmid DNA by *Eco*RI digestion and was purified. The 9-kb insert was then digested with various restriction enzymes, the restricted DNA fragments were separated by agarose gel electrophoresis, and Southern blot hybridizations were performed as described in Materials and Methods. A representative result is shown in Fig. 1. Of the four *Bgl*II fragments detected (Fig. 1A, lanes 2 and 3), the 2.1-kb fragment was HBV X region specific (Fig. 1D), the 1.85-kb fragment was HBV S gene specific (Fig. 1B), and the 0.4-kb fragment was HBV C gene specific (Fig. 1C). The 1.0-kb *Bgl*II fragment was cellular DNA since it did not hybridize to any of the HBV probes. Similar analyses were performed with enzymes *Hin*dIII, *Taq*I, *Xba*I, and *Xho*I (data for last three not shown), followed by Southern blot hybridizations with the various HBV probes. The organization of the integrated HBV sequences was

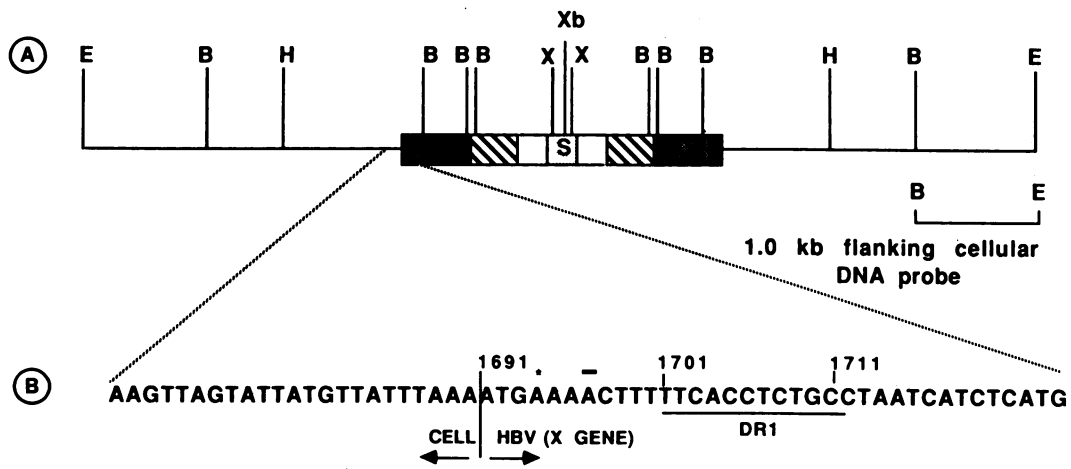


FIG. 2. Restriction map of the cloned 9.0-kb DNA fragment from Chinese hepatoma no. 7. (A) Structure of integrated HBV and flanking cellular DNA. HBV DNA (boxes) is flanked on each side by duplicated cellular DNA (line). HBV genes identified by Southern blot hybridization include the X gene (black boxes), the C gene (grey boxes, labeled "C"), the gap between C and pre-S (hatched boxes), the pre-S region (open boxes), and the S gene (grey box, labeled "S"). Enzymes used include *Eco*RI (E), *Bgl*II (B), *Hpa*I (H), *Xho*I (X), and *Xba*I (Xb). The following enzymes did not cut the 9.0-kb DNA: *Bam*HI, *Pst*I, *Sst*I, *Sst*II, *Not*I, *Eco*R5, *Bcl*I, *Nde*I, *Sca*I, *Pvu*2, *Sfi*, *Cla*I, and *Sma*I. The 1.0-kb *Bgl*II-*Eco*RI fragment, used for chromosome mapping, is identified at the right. Note that because all flanking cellular sequences were duplicated in this inverted repeat structure, an identical 1.0-kb *Bgl*II-*Eco*RI probe may be generated from the left side of the 9.0-kb insert. (B) Nucleotide sequence of virus-cell junction. Numbers above nucleotides refer to base assignments of the HBV subtype *adr* map (17). Two differences were noted between our sequence and a previously published map: the asterisk denotes A in place of C, and Ā denotes an extra A not found by Kobayashi and Katsuro (17). DR1 refers to one of the two direct repeat regions present in HBV.

deduced in accordance with published maps for HBV subtype *adw* (since *adw* was used to generate the DNA probes [45, 52]) and for HBV subtype *adr* (since it was presumed to be the subtype present in the hepatoma [17]).

The most striking feature of the restriction map (Fig. 2A) was that the entire 9-kb DNA consisted of a large inverted repeat structure proceeding outward from near the center of the cloned fragment (near the *Xba*I site). Of the total 9 kb of DNA, approximately 3 kb was HBV specific (1.5 kb was unique) and the remaining 6 kb was cellular (3 kb was unique). All flanking cellular sequences from the left side of the viral insert were repeated on the right side of the insert. The virus-virus junction lay within the S gene; at least 127 bp of the S gene were preserved, since the recognition site for *Xba*I (*adr* bp 121 [17]) was retained. Hybridization with a 277-bp X gene-specific probe (bp 1401 to 1678 [52]) was negative (data not shown), suggesting that the virus-cell junction mapped near the region of HBV containing direct repeat 1 (DR1; map positions 1701 to 1712).

To more precisely analyze the virus-cell junction, we subcloned the 2.1-kb *Bgl*II fragment and subjected it to nucleotide sequence analysis as described in Materials and Methods. The virus-cell junction was within the HBV X gene and was located 11 nucleotides from the DR1 region (Fig. 2B). Of the 178 HBV nucleotides sequenced, 174 (98%) were identical to those of another *adr* isolate (17). Of the four differences, two were found near the virus-cell junction: a substitution at bp 1694 (Fig. 2B, asterisk) and an additional nucleotide at bp 1697 (Fig. 2B, designated Ā). Although the sequence data shown were derived from the left virus-cell junction, the same sequence would be expected from the right virus-cell junction since both viral and cellular sequences were duplicated in the inverted repeat structure.

**Assignment of cellular DNA sequences that flank the HBV insert to chromosome 17p11.2-17p12.** The site in normal cellular DNA into which HBV integrates may be an important factor in the mechanism by which HBV contributes to the development of a hepatoma. Therefore, the 1.0-kb *Bgl*II-*Eco*RI fragment of cellular DNA that flanked the HBV insert (Fig. 2A) was used as a probe to identify the chromosomal origin of the 9-kb cloned DNA.

A collection of human-rodent cell hybrids containing segregated human chromosomes was used for chromosome mapping studies. High-molecular-weight DNA from human, hamster, or mouse cells as well as from the human-rodent cell hybrids was digested with *Eco*RI. The DNA was then separated by agarose gel electrophoresis, blotted to a nylon filter, and hybridized with a $^{32}$P-labeled 1.0-kb flanking cellular DNA probe. Under high-stringency hybridization conditions, a single band was identified in human DNA and in cell hybrids C.1A (Fig. 3, arrowhead) and MR2.2 (data not shown). This result indicated that the flanking DNA probe detected a single-copy gene and did not contain repetitive DNA sequences. The probe did not react with the mouse and hamster cell DNA controls under the hybridization conditions used. The results of similar analyses performed on additional cell hybrids are summarized in Table 1. Identical results were obtained with a 1.0-kb *Eco*RI-*Bgl*II probe generated from either the right half or the left half of the cloned 9-kb DNA (Fig. 2A). Taken together, the results indicated that the flanking cellular DNA was from human chromosome 17. Hybridization of the probe to DNA from cell hybrid MH-22-6, which contains exclusively chromosome 17, confirmed this assignment.

The availability of additional human-rodent cell hybrids that contained various deletions within chromosome 17 allowed more precise mapping of the viral insert to a specific region of the chromosome. Small terminal deletions or translocations of 17p did not affect the reactivity of the 1.0-kb probe (e.g., cell hybrid MH-74) (Table 2). Deletions inward to 17p13.100 (i.e., cell hybrid HO-11) remained reactive to the 1.0-kb probe. However, hybrids not containing region 17p11→pter (e.g., SA-5, MH-18) failed to hybridize to the probe. Hybrid HB-9, which has an interstitial

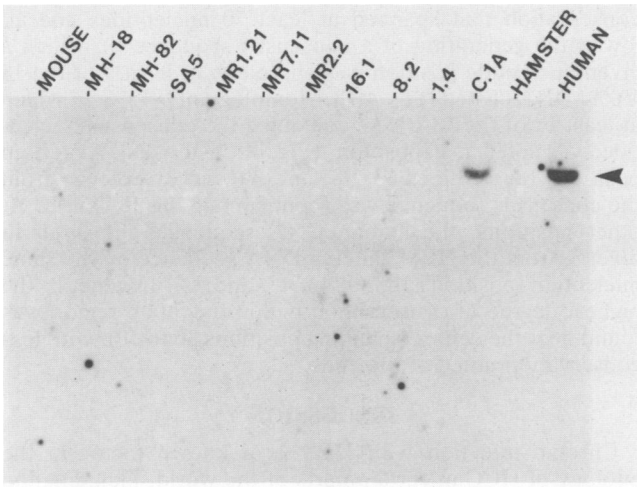


FIG. 3. Detection of 1.0-kb flanking cellular DNA probe sequences in different human-rodent cell hybrids. The 1.0-kb flanking cellular DNA probe (Fig. 2) was screened against a variety of rodent-human cell hybrids that have segregated human chromosomes. DNA from mouse, hamster, and human cells as well as from several cell hybrids was digested with *Eco*RI and subjected to agarose gel electrophoresis and Southern blot analysis with a $^{32}$P-labeled 1.0-kb probe. The DNA was degraded in the experiment with hybrid MH-82; therefore, the negative result was not informative. Under conditions of high-stringency hybridization, a single band was identified in human DNA and in the hybrid C.1A (arrowhead). A band also was detected in the hybrid MR2.2 upon longer exposure (data not shown). No bands were detected in mouse or hamster DNA.

deletion lacking most of band p11.2, hybridized positively to the probe. Thus, the probe maps to a region that includes all of p12 and may include very small portions of distal p11.2 or proximal p13.1 or both. The absence of the long arm of chromosome 17 (e.g., cell hybrid 88-H-5) did not affect hybridization with the flanking DNA probe. Thus, the 1.0-kb probe maps in the region 17p11.2-17p12.

It is of interest that the human proto-oncogene p53 maps to a nearby region of chromosome 17p. However, the pattern of reactivity of the 1.0-kb probe can be distinguished from that of a p53 cDNA probe; the p53 probe maps in the region distal to the HO-11 breakpoint (17p13.100; Table 2) (vanTuinen et al., in press), whereas the 1.0-kb probe maps proximally.

**Identification of a consensus sequence within the flanking cellular DNA.** We next considered whether the HBV insert we had cloned was integrated within an identifiable gene. The 1.0-kb flanking DNA probe was used to demonstrate the presence of complementary sequences in preparations of normal liver whole-cell RNA (data not shown), suggesting that sequences within the 1.0-kb DNA were from a portion of the chromosome being actively expressed. To further investigate the nature of these sequences, we subcloned the 1.0-kb *Bgl*II-*Eco*RI fragment and subjected it to DNA sequence analysis. In total, four portions of flanking cellular DNA were sequenced (extending inward from both ends of the 2.1-kb *Bgl*II fragment and from both ends of the 1.0-kb *Bgl*II-*Eco*RI fragment). None of the four sequences analyzed contained discernible open reading frames (data not shown). The DNA sequences were then used to search GenBank for identification of possible homology with other GenBank entries. Three of the four DNA sequences showed no homology with any GenBank sequences. However, the

TABLE 1. Assignment of 1.0-kb flanking cellular DNA to human chromosome 17

| Cell hybrid | Reactivity of 1.0-kb probe[a] | Presence[b] of human chromosome: | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | X |
| Human | + | | | | | | | | | | | | | | | | | | | | | | | |
| Mouse | − | | | | | | | | | | | | | | | | | | | | | | | |
| Hamster | − | | | | | | | | | | | | | | | | | | | | | | | |
| MH-22-6 | + | − | − | − | − | − | − | − | − | − | − | + | − | − | − | + | + | + | + | − | − | − | − | − |
| C.1A | + | − | − | + | + | + | + | − | − | − | + | + | + | − | + | + | + | + | + | − | + | + | + | + |
| 1.4 | − | − | − | + | + | − | − | − | − | − | − | − | − | − | − | − | + | − | − | + | − | + | − | + |
| 8.2 | − | − | + | + | + | − | − | + | + | − | + | + | + | − | + | + | + | − | − | − | − | − | − | + |
| 16.1 | − | − | + | + | + | + | + | − | − | − | − | + | − | + | − | − | − | − | + | − | − | + | − | + |
| MR7.11 | − | − | − | − | − | − | − | − | − | − | − | + | − | + | − | + | − | − | − | − | − | − | + | + |
| MR1.21 | − | − | − | − | − | + | + | + | − | − | − | − | + | − | + | − | − | − | − | − | − | − | − | + |
| 11.1.2.2 | − | − | − | + | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | − | + |
| 1.11 | − | − | − | − | − | − | + | − | + | − | − | + | − | − | − | − | − | − | − | − | − | − | − | + |
| MR2.2 | + | − | − | + | − | − | + | − | − | − | − | + | − | + | − | − | − | −[c] | − | − | + | − | + | + |
| No. (%) discordant[d] | | 3 (30) | 5 (50) | 5 (50) | 5 (50) | 3 (30) | 4 (40) | 4 (40) | 5 (50) | 3 (30) | 3 (30) | 5 (50) | 4 (40) | 3 (30) | 4 (40) | 4 (40) | 5 (50) | 0 (0) | 3 (30) | 5 (50) | 2 (20) | 4 (40) | 2 (20) | 8 (80) |

[a] +, Presence of band on autoradiogram; −, absence of band on autoradiogram.
[b] +, Presence of chromosome; −, absence of chromosome.
[c] Hybrid MR2.2 was not found to contain chromosome 17 by a cytological exam. However, chromosome 17 has previously been detected by DNA hybridization techniques.
[d] A discordant result is defined as a positive probe reaction in the absence of a specific chromosome or vice versa.

TABLE 2. Regional mapping of 1.0-kb flanking cellular DNA in chromosome 17

| Cell hybrid | Southern blot hybridization to: | | Portion of chromosome 17 retained |
|---|---|---|---|
| | 1.0-kb probe[a] | p53 probe[b] | |
| Human | + | + | Intact |
| Mouse | − | − | |
| Hamster | − | − | |
| MH-74 | + | + | p13.3 → q25.3 |
| FW-1 | + | + | p13.108 → qter |
| JW-4 | + | + | p13.105 → qter |
| HO-11 | + | − | p13.100 → qter |
| SA-5 | − | − | p11.2 → qter |
| HB-9 | + | + | pter → p11.2::p11.2 → qter |
| MH-18 | − | − | p11.2 → qter |
| MH-5 | − | − | p11.2 → qter |
| 88-H-5 | + | + | pter → p11.2 |
| LS-1 | − | − | cen → qter |
| SP-3 | − | − | q11.2 → qter |
| ND-1 | − | − | q21.1 → qter |
| MH-41 | − | − | q23 → qter |

[a] +, Presence of band on autoradiogram; −, absence of band on autoradiogram.
[b] Data are from vanTuinen et al. (in press).

remaining DNA sequence showed interesting similarities to GenBank sequences.

A DNA sequence derived from the *Eco*RI end of the 1.0-kb *Bgl*II-*Eco*RI flanking cellular DNA (Fig. 4A) showed highly significant homology with sequences located in the noncoding or flanking regions of a number of functional human genes. Noncoding regions from 16 human genes as well as from the human autonomously replicating sequence 1 (*ARS1*) (29) showed significant homology, as measured by our search program (22), demonstrating >6 standard deviations of similarity as compared with the expected score. Alignment of the HBV flanking sequence with the homologous sequences from the 17 DNAs demonstrated a striking conservation that spanned at least 70 nucleotides and allowed the generation of a consensus sequence (in which a given nucleotide position was shared by at least 9 of the 18 [50%] DNAs listed [Fig. 4B]). Homology of >75%, in which at least 14 of the 18 DNAs contained the consensus nucleotide, was noted at 30 positions (Fig. 4C, asterisks). At certain positions (nucleotides 33, 38, and 39), no divergence from the consensus sequence was found among the 18 DNAs. At other positions, the human DNA sequences appeared to diverge from the HBV flanking DNA sequence by the same nucleotide (positions 16, 40, 42, 47, and 52). In general, the highest degree of conservation within the 70-bp region was found near the center (spanning positions 30 to 50), with less conservation noted at the ends.

## DISCUSSION

Chronic infection with HBV is a known factor in the etiology of HCC in certain parts of the world, including the People's Republic of China, and confers a risk of >200-fold (1, 26). The analysis of viral DNA sequences integrated into the tumor DNA may help elucidate a molecular mechanism(s) by which HBV contributes to hepatoma formation. The present report describes the analysis of an integrated HBV genome and its flanking cellular sequences cloned from a Chinese HCC specimen.

Analysis of several virus-cell junctions from HBV-positive hepatomas and hepatoma-derived cell lines has revealed that the two 11-bp direct repeat (DR) regions are often a preferred site within the viral genome in which integration with cellular DNA may occur (4, 8, 20, 35, 39, 50, 51). Consistent with these previous findings, we have determined that the virus junction from the cloned 9-kb insert described in this paper maps to bp 1691 on the HBV subtype *adr* map (17). This site is 11 bp from DR1 and is within the cohesive overlap region. Although the precise mechanism for HBV integration remains unclear, a growing number of virus junctions have been mapped in or near the DRs, so it is likely that conditions favorable for integration include some characteristic feature of the DR-cohesive overlap region.

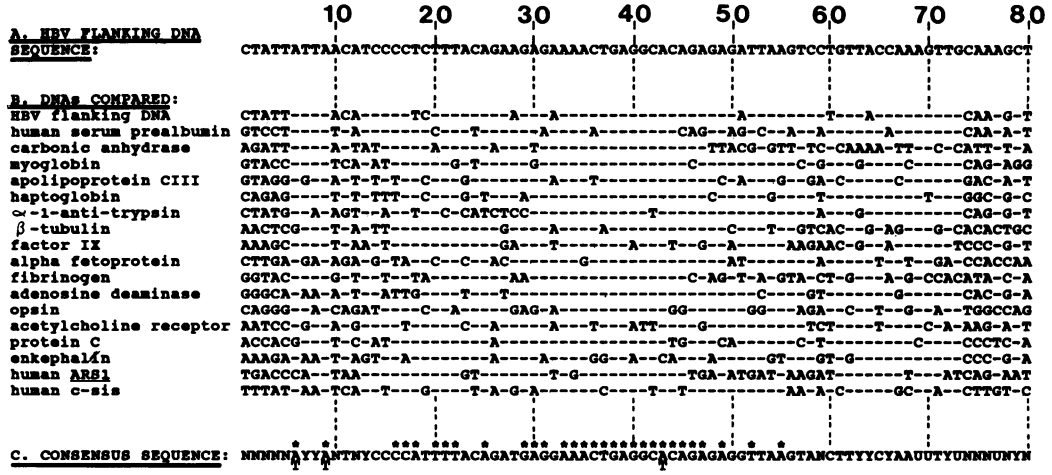


FIG. 4. Identification of a consensus sequence within HBV flanking cellular DNA. (A) HBV flanking DNA sequence derived from near the *Eco*RI recognition site of the 1.0-kb *Bgl*II-*Eco*RI fragment of DNA (Fig. 2A). (B) Comparison of human DNA sequences to the consensus sequence. HBV flanking DNA was used to screen GenBank for DNAs containing homologous regions. The DNAs were aligned and compared with the consensus sequence (see panel C). Dashes identify homology with the consensus sequence at a given nucleotide position. (C) Consensus sequence. Comparison of the 18 DNAs listed in panel B led to the generation of a consensus sequence in which a given position was occupied by the consensus nucleotide in at least 9 of the 18 DNAs (50%). Homologies of >75% at a given position are noted by an asterisk. N, Any of the four nucleotides; Y, pyrimidine; U, purine.

Previous studies have shown that HBV sequences integrated into hepatoma DNA may exist as an intact genome (11), as subgenomic forms (4, 6, 7, 14, 28, 35, 39, 42, 51), or even as complicated inverted structures (7, 14, 28, 42, 50). Consistent with these previous descriptions of complex forms of viral DNA, the cloned insert recovered from an HCC and characterized in this report consists of a large inverted repeat structure in which the duplication of DNA involves not only viral sequences but also flanking cellular sequences. The entire HBV C gene is present, although the 9-kb insert lacks the tissue-specific transcriptional enhancer sequences located approximately 450 bp upstream from the C gene (3, 16, 38). Also missing from the insert are the previously identified glucocorticoid responsive elements (43) and all but 26 bp from the 3′ end of the X gene. Since the X gene is known to possess a *trans*-activating function (41, 44; T.-H. Lee and J. S. Butel, unpublished data), this particular viral insert would lack that activity. The duplication of flanking cellular sequences extends outwards at least 3 kb, to the *Eco*RI site used for cloning the 9-kb fragment. Similarly, large duplications of cellular sequences have been observed at the site of HBV integration in other hepatomas (28, 42), and the effect that that event would have on the cell clearly would depend on the nature of the cell sequences duplicated.

HBV inserts from individual tumors have been localized to different human chromosomes, including 3 (6), 4 (33), 6 (14), and 11p (35), at a chromosome 17q25;18 translocation (14, 34), at a chromosome X;17 translocation (42), at a chromosome 5;9 translocation (42), and at chromosome 17p11.2-17p12 (this report). If HBV integration within an important chromosomal location is a key event in transformation, then disruption of DNA in a variety of locations may contribute to the development of a tumor. Although we found rearrangement of cellular sequences near the site of integration (a duplication of at least 3 kb), our experiments have not addressed the possibility of a chromosomal deletion. A deletion at the site of HBV integration in chromosome 11p, with the resultant loss of heterozygosity in that region of the chromosome, has been described (H. P. Wang and C. E. Rogler, Cytogenet. Cell Genet., in press).

The activation of proto-oncogenes has been suggested as one mechanism by which HBV may cause tumors, although definitive data are lacking (12). Assignment of the 9-kb insert DNA to human chromosome 17p immediately raises the possibility of involvement of the nearby proto-oncogene p53. Numerous reports have shown that p53 may be involved in transformation of a variety of cell types (5, 10, 18, 21, 24, 32). In situ hybridization techniques have demonstrated that human proto-oncogene p53 maps to chromosome 17p13 (15, 25, 27). This would place the HBV flanking cellular DNA within the same small region of 17p as p53, a region of DNA estimated to span from 2 to 7 megabases (46). However, when Southern blot hybridization analysis was used to analyze the cell hybrids that contain chromosome 17p deletions, the p53 gene and the 1.0-kb probe mapped to opposite sides of the HO-11 breakpoint at 17p13.100 (Table 2). Although the two DNAs were separated by the chromosomal breakpoint, additional experiments are needed to define the precise distance between the HBV insert and the p53 gene.

Detailed analysis of the HBV flanking cellular DNA sequences revealed a region of DNA that was approximately 70 bp in length and that was highly homologous to an intron-associated sequence present in the noncoding regions (predominantly introns, but also 5′- and 3′-flanking regions) of a number of functional human genes. The HBV flanking DNA sequence was aligned with the homologous regions of the other DNA sequences, and a consensus sequence was generated. The consensus sequence shows the highest homology in the core region, while progressively less homology is observed nearer the ends. The consensus sequence contains several features that distinguish it from the short interspersed nucleotide elements (SINEs) previously described (37, 48). Whereas the consensus sequence is present only once in the genes listed (Fig. 4B), SINEs such as human *Alu* sequences may be present many times in multiple introns of the same gene (37). Additionally, all SINEs possess characteristic 3′ poly(A) sequences and DRs at their termini (37). Searches beyond the 70-bp consensus sequence reported here failed to identify such structures. Finally, the homology among SINEs (such as *Alu* sequences) indicates that conserved regions alternate with divergent regions (37) and that a core region of homology is lacking.

The possible function of the consensus sequence remains unknown. Although GenBank contains sequence data primarily from well-characterized genes and it could be argued that the consensus sequence may also be located within nonfunctional DNA and therefore be of little or no significance, several observations are consistent with the idea that the consensus sequence is functional. The core region of high homology is similar to that seen with transcriptional regulatory elements. While most such regulatory elements are located in the 5′-noncoding region of genes, exceptions do exist. The β-globin gene contains sequences in the second intron (IVS-2) required for efficient transcriptional expression of that gene (2, 19). However, IVS-2 bears no significant homology with the consensus sequence reported here.

One intriguing piece of evidence suggestive of a functional role for the consensus sequence is its homology with a segment of the human *ARS1* DNA. *ARS1* is a sequence of human DNA that allows replication of *Saccharomyces cerevisiae* integrative plasmids as autonomously replicating elements in *S. cerevisiae* cells (29). Within *ARS1* is a 325-bp segment that is necessary for maximal expression of the autonomous replication phenotype, analogous to an enhancer. Interestingly, the consensus sequence is located in the center of this 325-bp segment with enhancerlike activity. However, it is not known if the consensus sequence represents the active portion of the replicative enhancer. A similar association of the consensus sequence with potential DNA replicative sequences is suggested by its presence in African green monkey cell DNA, within a few hundred base pairs of a sequence exhibiting similarity to the simian virus 40 origin of replication (36). Further experiments are needed to more directly assess the functional significance of the consensus sequence.

Current data suggest that integration of HBV DNA may contribute to liver tumor formation in at least three general ways. First, the expression of cellular genes may be directly affected by the integration event itself (6, 9, 31, 47). Second, the chromosomal abnormalities frequently associated with HBV integration sites may have important but as-yet-undefined effects on normal cell growth. Third, the *trans*-activating capability of the HBV X gene (41, 44; Lee and Butel, unpublished data) may activate the expression of cellular genes. It is noteworthy that the *trans*-activating *tat* gene of human T-lymphotropic virus type 1 is tumorigenic in transgenic mice (30).

It is appearing more likely that the crucial cellular targets involved in triggering liver cell malignant transformation may differ from tumor to tumor. Analysis of additional integrated HBV sequences and the flanking cellular DNA

will be a necessary prerequisite to a more complete understanding of the role(s) of HBV in liver cancer.

## ACKNOWLEDGMENTS

This study was supported in part by Public Health Service research grants CA37257 (J.S.B.) from the National Cancer Institute and HD20619 (D.H.L.) and HD06814 (P.V.) from the National Institute of Child Health and Human Development and by National Research Service award CA09197.

## LITERATURE CITED

1. **Beasley, R. P., L.-Y. Hwang, C.-C. Lin, and C.-S. Chien.** 1981. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22,707 men in Taiwan. Lancet **ii**:1129–1133.
2. **Behringer, R. R., R. E. Hammer, R. L. Brinster, R. D. Palmiter, and T. M. Townes.** 1987. Two 3′ sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. Proc. Natl. Acad. Sci. USA **84**:7056–7060.
3. **Bulla, G. A., and A. Siddiqui.** 1988. The hepatitis B virus enhancer modulates transcription of the hepatitis B virus surface antigen gene from an internal location. J. Virol. **62**:1437–1441.
4. **Choo, K. B., M. S. Liu, P. C. Chang, S. M. Wu, M. W. Su, C. C. Pan, and S. H. Han.** 1986. Analysis of six distinct integrated hepatitis B virus sequences cloned from the cellular DNA of a human hepatocellular carcinoma. Virology **154**:405–408.
5. **de Fromentel, C. C., F. May-Levin, H. Mouriesse, J. Lemerle, K. Chandrasekaran, and P. May.** 1987. Presence of circulating antibodies against cellular protein p53 in a notable proportion of children with B-cell lymphoma. Int. J. Cancer **39**:185–189.
6. **Dejean, A., L. Bougueleref, K. H. Grzeschik, and P. Tiollais.** 1986. Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. Nature (London) **322**:70–72.
7. **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 PLC/PRF/5. Proc. Natl. Acad. Sci. USA **80**:2505–2509.
8. **Dejean, A., P. Sonigo, S. Wain-Hobson, and P. Tiollais.** 1984. Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat. Proc. Natl. Acad. Sci. USA **81**:5350–5354.
9. **de Thé, H., A. Marchio, P. Tiollais, and A. Dejean.** 1987. A novel steroid thyroid hormone receptor-related gene inappropriately expressed in human hepatocellular carcinoma. Nature (London) **330**:667–670.
10. **Eliyahu, D., A. Raz, P. Gross, D. Givol, and M. Oren.** 1984. Participation of p53 cellular tumor antigen in transformation of normal embryonic cells. Nature (London) **312**:646–649.
11. **Fowler, M. J. F., H. C. Thomas, and J. Monjardino.** 1986. Cloning and analysis of integrated hepatitis B virus DNA of the *adr* subtype derived from a human primary liver cell carcinoma. J. Gen. Virol. **67**:771–775.
12. **Ganem, D., and H. E. Varmus.** 1987. The molecular biology of the hepatitis B viruses. Annu. Rev. Biochem. **56**:651–693.
13. **Guo, L., and R. Wu.** 1982. Sequencing of proteins by DNA sequencing, p. 169–182. *In* T. Lo, T. Liu, and C. Li (ed.), Biochemical and biophysical studies of proteins and nucleic acids. Elsevier Science Publishing, Inc., New York.
14. **Hino, O., T. B. Shows, and C. E. Rogler.** 1986. Hepatitis B virus integration site in hepatocellular carcinoma at chromosome 17;18 translocation. Proc. Natl. Acad. Sci. USA **83**:8338–8342.
15. **Isobe, M., B. S. Emanuel, D. Givol, M. Oren, and C. M. Croce.** 1986. Localization of gene for human p53 tumour antigen to band 17p13. Nature (London) **320**:84–85.
16. **Jameel, S., and A. Siddiqui.** 1986. The human hepatitis B virus enhancer requires *trans*-acting cellular factor(s) for activity. Mol. Cell. Biol. **6**:710–715.
17. **Kobayashi, M., and K. Katsuro.** 1984. Complete nucleotide sequence of hepatitis B virus DNA of subtype *adr* and its conserved gene organization. Gene **20**:227–232.
18. **Koeffler, H. P., C. Miller, M. A. Nicolson, J. Ranyard, and R. A. Bosselman.** 1986. Increased expression of p53 protein in human leukemia cells. Proc. Natl. Acad. Sci. USA **83**:4035–4039.
19. **Kosche, K. A., C. Dobkin, and A. Bank.** 1985. DNA sequences regulating human β-globin gene expression. Nucleic Acids Res. **13**:7781–7793.
20. **Koshy, R., S. Koch, A. 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.
21. **Lane, D. P., and L. V. Crawford.** 1979. T antigen is bound to a host protein in SV40-transformed cells. Nature (London) **278**:261–263.
22. **Lawrence, C. B., D. A. Goldman, and R. T. Hood.** 1986. Optimized homology searches of the gene and protein sequence data banks. Bull. Math. Biol. **48**:569–583.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Masuda, H., C. Miller, H. P. Koeffler, H. Battifora, and M. J. Cline.** 1987. Rearrangement of the p53 gene in human osteogenic sarcomas. Proc. Natl. Acad. Sci. USA **84**:7716–7719.
25. **McBride, O. W., D. Merry, and D. Givol.** 1986. The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). Proc. Natl. Acad. Sci. USA **83**:130–134.
26. **Melnick, J. L.** 1985. Approaching the control of primary liver cancer by means of a virus vaccine. Prog. Med. Virol. **32**:22–38.
27. **Miller, C., T. Mohandas, D. Wolf, M. Prokocimer, V. Rotter, and H. P. Koeffler.** 1986. Human p53 gene localized to short arm of chromosome 17. Nature (London) **319**:783–784.
28. **Mizusawa, H., M. Taira, K. Yaginuma, M. Kobayashi, E. Yoshida, and K. Koike.** 1985. Inversely repeating integrated hepatitis B virus DNA and cellular flanking sequences in the human hepatoma-derived cell line huSP. Proc. Natl. Acad. Sci. USA **82**:208–212.
29. **Montiel, J. F., C. J. Norbury, M. F. Tuite, M. J. Dobson, J. S. Mills, A. J. Kingsman, and S. M. Kingsman.** 1984. Characterization of human chromosomal DNA sequences which replicate autonomously in *Saccharomyces cerevisiae*. Nucleic Acids Res. **12**:1049–1068.
30. **Nerenberg, M., S. H. Hinrichs, R. K. Reynolds, G. Khoury, and G. Jay.** 1987. The *tat* gene of human T-lymphotropic virus type 1 induces mesenchymal tumors in transgenic mice. Science **237**:1324–1329.
31. **Ou, J., and W. J. Rutter.** 1985. Hybrid hepatitis B virus-host transcripts in a human hepatoma cell. Proc. Natl. Acad. Sci. USA **82**:83–87.
32. **Parada, L. F., H. Land, R. A. Weinberg, D. Wolf, and V. Rotter.** 1984. Cooperation between gene encoding p53 tumour antigen and *ras* in cellular transformation. Nature (London) **312**:649–651.
33. **Pasquinelli, C., F. Garreau, L. Bougueleref, E. Cariani, K. H. Grzeschik, V. Thiers, O. Croissant, M. Hadchouel, P. Tiollais, and C. Brechot.** 1988. Rearrangement of a common cellular DNA domain on chromosome 4 in human primary liver tumors. J. Virol. **62**:629–632.
34. **Rogler, C. E., O. Hino, and C. Su.** 1987. Molecular aspects of persistent woodchuck hepatitis virus and hepatitis B virus infection and hepatocellular carcinoma. Hepatology (Baltimore) **7**:74S–78S.
35. **Rogler, C. E., M. Sherman, C. Y. Su, D. A. Shafritz, J. Summers, T. B. Shows, A. Henderson, and M. Kew.** 1985. Deletion in chromosome 11p associated with a hepatitis B integration site in hepatocellular carcinoma. Science **230**:319–322.
36. **Saffer, J. D., and M. I. Lerman.** 1983. Unusual class of *Alu* sequences containing a potential Z-DNA segment. Mol. Cell. Biol. **3**:960–964.
37. **Schmid, C. W., and C.-K. J. Shen.** 1986. The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates, p. 323–358. *In* R. J. MacIntyre (ed.), Molecular evolutionary genetics. Plenum Publishing Corp., New York.
38. **Shaul, Y., W. J. Rutter, and O. Laub.** 1985. A human hepatitis B virus enhancer element. EMBO J. **4**:427–430.

39. **Shih, C., K. Burke, M. Chou, J. B. Zeldis, C. Yang, C. Lee, K. J. Isselbacher, J. R. Wands, and H. M. Goodman.** 1987. Tight clustering of human hepatitis B virus integration sites in hepatomas near a triple-stranded region. J. Virol. **61**:3491–3498.
40. **Siddiqui, A., P. L. Marion, and W. S. Robinson.** 1981. Ground squirrel hepatitis virus DNA: molecular cloning and comparison with hepatitis B virus DNA. J.Virol. **38**:393–397.
41. **Spandau, D. F., and C.-H. Lee.** 1988. *trans*-Activation of viral enhancers by the hepatitis B virus X protein. J. Virol. **62**:427–434.
42. **Tokino, T., S. Kukushige, T. Nakamura, T. Nagaya, T. Murotsu, K. Shiga, N. Aoki, and K. Matsubara.** 1987. Chromosomal translocation and inverted duplication associated with integrated hepatitis B virus in hepatocellular carcinomas. J. Virol. **61**:3848–3854.
43. **Tur-Kaspa, R., R. D. Burt, Y. Shaul, and D. A. Shafritz.** 1986. Hepatitis B virus DNA contains a glucocorticoid-responsive element. Proc. Natl. Acad. Sci. USA **83**:1627–1631.
44. **Twu, J.-S., and R. H. Schloemer.** 1987. Transcriptional *trans*-activating function of hepatitis B virus. J. Virol. **61**:3448–3453.
45. **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. N. Fields, R. Jaenisch, and C. F. Fox (ed.), Animal virus genetics. Academic Press, Inc., New York.
46. **vanTuinen, P., D. Rich, K. M. Summers, and D. H. Ledbetter.** 1987. Regional mapping panel for human chromosome 17: application to neurofibromatosis type 1. Genomics **1**:374–381.
47. **von Loringhoven, A. F., S. Koch, P. H. Hofschneider, and R. Koshy.** 1985. Co-transcribed 3′ host sequences augment expression of integrated hepatitis B virus DNA. EMBO J. **4**:249–255.
48. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986. Non-viral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.
49. **Wu, X., Y. Zhou, Z. Feng, Z. Li, and S. Xia.** 1983. Cloning and restriction mapping of human HBV genome serotype adr. Sci. Sin. Ser. B **XXVI**:954–960.
50. **Yaginuma, K., H. Kobayashi, M. Kobayashi, T. Morishima, K. Matsuyama, and K. Koike.** 1987. Multiple integration site of hepatitis B virus DNA in hepatocellular carcinoma and chronic active hepatitis tissues from children. J. Virol. **61**:1808–1813.
51. **Yaginuma, K., M. Kobayashi, E. Yoshida, and K. Koike.** 1985. Hepatitis B virus integration in hepatocellular carcinoma DNA: duplication of cellular flanking sequences at the integration site. Proc. Natl. Acad. Sci. USA **82**:4458–4462.
52. **Zhou, Y., J. S. Butel, P. Li, M. J. Finegold, and J. L. Melnick.** 1987. Integrated state of subgenomic fragments of hepatitis B virus DNA in hepatocellular carcinoma from mainland China. JNCI **79**:223–231.