JOURNAL OF VIROLOGY, Apr. 1987, p. 1180–1186
0022-538X/87/041180-07$02.00/0

Vol. 61, No. 4

Integration of Hepatitis B Virus: Analysis of Unoccupied Sites

ITZHAK BERGER AND YOSEF SHAUL*

Department of Virology, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 21 August 1986/Accepted 6 December 1986

Hepatitis B virus (HBV) sequences integrated in the PLC/PRF/5 cell line (Alexander cells), which was derived from a human primary liver carcinoma, were previously extensively studied. Here we describe the analysis of the unoccupied sites of two linearly integrated forms of HBV DNA, AL-14 and AL-26, that were characterized previously. No major cellular DNA rearrangements were seen at the integration sites except for small deletions of host sequences: 2 kilobases of DNA in AL-14 and 17 base pairs (bp) in AL-26. The unoccupied site of AL-26 was found to be missing 182 bp, which previously mapped next to the right end of the integration sites of several independent clones. These were believed to be of cellular origin, but we show here that these 182 bp are in fact from unusual HBV sequences. Surprisingly, a region of this newly detected HBV DNA sequence is more homologous to that of woodchuck HBV DNA. Our analysis shows that the normal counterparts of both AL-14 and AL-26 contain minisatellite-like repetitive sequences. Based on the data presented here and our previous finding of HBV DNA integration at satellite sequences, we propose that genomic simple repetitive sequences are hot spots for HBV DNA integration.

Hepatitis B virus (HBV) is considered to be a major cause of liver cancer (hepatocellular carcinoma; HCC), and integrated forms of HBV DNA are frequently found in HCC cells (for a review, see reference 24). The nature of the integrated HBV DNA was intensively studied in the PLC/PRF/5 (Alexander) cell line, which was derived from HCC tissue (10, 20). This cell line contains at least seven integrated copies of HBV DNA, none of which are intact (20). In most of these integrations, deletions, duplications, and inversions of HBV DNA were observed; and only two clones (AL-14 and AL-26) contained an uninterrupted linear form of HBV sequences. AL-26 and clones A-10.7 and A-10.5, which have been isolated by Koch et al. (9), were found to be almost identical in their nonviral sequences at the 3′ end of the integration site; and therefore, it was postulated that these clones are derived from the amplification of a primary integrated viral DNA with its immediate host sequences (8). Here we report the isolation and characterization of the unoccupied sites of AL-26 and AL-14. These analyses allow us to show that sequences that were previously believed to be of host origin are in fact those of unusual viral sequences. We also found that in both cases the viral DNA was integrated next to simple repetitive sequences of cellular DNA, which are partially homologous to the previously reported minisatellite sequences (7). We propose here that simple cellular repetitive sequences are hot spots for the integration of HBV DNA.

MATERIALS AND METHODS

Preparation of genomic DNA. DNA was prepared from normal blood cells, hepatoma cell lines, and primary HCCs as follows. Isolated nuclei or whole cells were suspended in 10 volumes of 0.1 M NaCl–40 mM Tris (pH 7.6)–40 mM EDTA. Then, 0.5% sodium dodecyl sulfate and 0.5 mg of proteinase K per ml were added, and the mixture was incubated at 60°C for 1 h. The resultant lysate was extracted with an equal volume of phenol-chloroform for 1 h and then extracted several times with an equal volume of chloroform. Nucleic acids in the aqueous phase were precipitated with 2 volumes of ethanol. DNA was dissolved in TE buffer (0.1 mM EDTA, 1 mM Tris [pH 8]) to a final concentration of 250 μg/ml. About 20 μg of DNA was used for each lane.

Restriction mapping and Southern blotting. Restriction enzyme digests of cloned DNA were electrophoresed on 1% agarose gels, blotted onto nitrocellulose by the method described by Southern (22), hybridized to probe DNA, and labeled with ^{32}P by nick translation (17, 25) or with the riboprobe system (14). The hybridizations were for 16 h at 42°C in 50% formamide–10% dextran sulfate. The filters were washed at 50°C in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate.

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 they were subcloned into the plasmid vectors pBR322 or pSP6. Specific fragments of the DNA of Alexander clones were prepared as hybridization probes by digestion of cloned DNA with the appropriate enzymes and isolation of the specific fragments from agarose gels before labeling.

Construction of a genomic DNA library. Lymphocyte DNA was digested completely with *Eco*RI enzyme. The region around the 2.8-kilobase (kb) fragment was recovered from a 1% agarose gel, ligated to λ*gt*11 arms, packaged in vitro, and used to infect *Escherichia coli* Y1088 (23). Positive clones were identified by ^{32}P labeling by previously published methods (2, 11, 12).

DNA sequencing and computer analysis. DNA sequencing of the M13 subcloned fragments was carried out by the dideoxy method (18).

Computer analysis of the DNA was performed on a computer (XT; International Business Machines) with the

* Corresponding author.

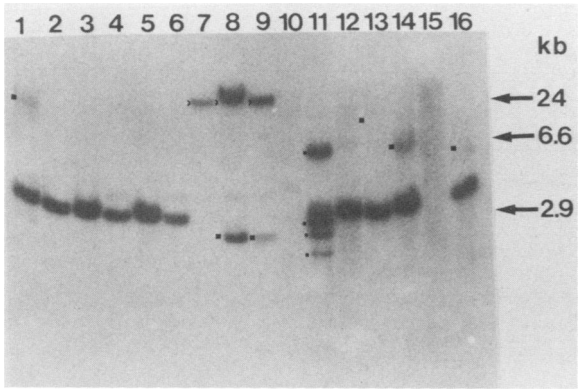

FIG. 1. Southern blot analysis of AL-14 unoccupied sites in normal and hepatoma cells. DNA samples (20 μg) were prepared from hepatoma cell lines Hep3B (lane 1), SK-Hep1 (lane 2), HepG2 (lane 3), Mahlavu (lane 4), and Alexander (lane 11); primary hepatocarcinomas (lanes 12 to 16); normal human blood cels (lanes 5 and 6); a mouse cell line (lane 10); and a monkey cell line (lane 7); lanes 8 and 9 contained the same DNA sample as lane 7 but with 1 and 0.2 equivalent copies, respectively, of *Eco*RI-digested AL-26. These DNA samples were digested with *Eco*RI electrophoresed in an agarose gel, Southern blotted, and hybridized to the probe shown in Fig. 2. The bands seen when HBV DNA was used as a probe are marked by dots; the arrowheads show the AL-26-related monkey sequences.

Micro-Genie program and on a personal computer (International Business Machines) with the Pustell program.

RESULTS

Analysis of the unoccupied site of AL-14 in normal and hepatoma DNA samples. Clone AL-14 contains linear integrated forms of HBV DNA, with the viral core and S genes and host flanking sequences occurring at both ends (20). A 2.6-kb *Eco*RI fragment, which contains about 1.4 kb of HBV DNA sequences and 1.2 kb of cellular DNA sequences corresponding to the left end of the virus-host junction, was shown not to contain repetitive DNA elements (data not shown) and, therefore, was used as a probe for analysis of genomic DNA from normal and several hepatoma cells (Fig. 1). A positive *Eco*RI band of 2.8 kb was seen in all the human DNA samples, and in addition, extra bands in some of the hepatoma samples were visible. By using more specific probes, we found that these extra bands are positive only to the HBV sequence and not to the host sequence in the AL-14 probe that was used (data not shown). This implies that no major DNA rearrangement occurred at this locus in the tested hepatoma cells. Monkey DNA contained a 24-kb *Eco*RI-positive band, while mouse DNA did not contain any homologous sequences. Thus, the cellular sequence is not evolutionarily conserved.

Determination of the sequence of the AL-14 unoccupied site. To analyze further the cellular counterpart of clone AL-14, we isolated the unoccupied site (normal allele) from a genomic library, which was prepared from normal cells, using the AL-14 probe. About 10^6 bacteriophages were screened, and 11 positive clones were isolated and analyzed. Detailed restriction map analysis of one of the clones, NHAL14, revealed that it contains host sequences corresponding to both ends of the integration site on a single 2.7-kb *Sma*I fragment (Fig. 2). In AL-14, which was isolated from Alexander cells, the size of the homologous *Sma*I fragment was only 2.4 kb. This 2.4-kb fragment was composed of 1.7 kb of HBV DNA and only 0.7 kb of host sequences. Thus, about 2 kb (2.7 to 0.7 kb) of cellular DNA was deleted as a consequence of the integration.

To determine the sequence of the unoccupied site, two restriction DNA fragments, the 0.7-kb *Eco*RI-*Bam*HI fragment and the 0.5-kb *Eco*RI-*Sma*I fragment, which contain the cellular sequences corresponding to the left and right ends of the integration site, respectively (Fig. 2), were isolated, cloned in the M13 vector, and sequenced. The sequences of HBV DNA and the left and the right ends of the virus-host junctions and the corresponding unoccupied site in NHAL14 are compared in Fig. 2. This comparison maps the left-end virus-host junction at nucleotides 2036 and 2037 and the right-end junction at nucleotides 489 to 491 on the HBV genome. The homology of the sequences of integrated and prototype HBV used by us was relatively poor at the 5′ end of the integration site. This was due, perhaps, to the sequence differences that are commonly observed among HBV subtypes. These mismatches among viral sequences at this region led to misinterpretation of the integration site, which was previously placed at nucleotide 2051 (E. Winocour, T. Chitlaru, K. Tsutsui, R. Ben-Levi, and Y. Shaul, *in Cancer Cells*, in press). There is 1 common base between HBV DNA and cellular human DNA at the 5′ end of the integration site, and there are 3 common bases at the 3′ end. Analysis of the unoccupied site of AL-26, as described below, also shows that there are 3 common bases at the integration site between the two DNAs (Fig. 3). This phenomenon seems to be a characteristic feature of nonhomologous recombination in general, as was suggested previously by us and others (3, 13; Winocour et al., in press).

Analysis of the unoccupied site of AL-26 in normal and hepatoma DNA samples. Similar to AL-14, clone AL-26 contains the linear form of integrated HBV DNA, with the viral S gene, a portion of the X gene, and cellular host sequences occurring at both ends (20). To analyze host flanking sequences corresponding to both ends of the integrated HBV DNA in normal and hepatoma cells, the AL-26 3.8-kb *Pst*I fragment was used as a probe. This fragment contains 0.7 kb of host sequences from the 5′ end of the integration site, 2.4 kb of integrated HBV DNA, and 0.7 kb of host sequences at the 3′ end (Fig. 3). This probe hybridized to a 2.8-kb *Eco*RI fragment in normal cells and in all the HCC samples that were tested (Fig. 4); again, the extra bands observed in some of the HCC DNA samples were due to the presence of the HBV sequences in the probe and in the tumor DNA samples (data not shown). Only monkey DNA, but not *Drosophila* DNA, hybridized with this probe, implying that these sequences are not evolutionarily conserved.

The AL-26 probe that we used contains a low number of repetitive sequences, as was seen by smears of hybridization background over the various lanes (Fig. 4). For the isolation of the unoccupied site of the DNA by this probe, we decided to prepare a partial genomic library that was enriched for the positive 2.8-kb *Eco*RI fragment. To do so we cloned gel-purified genomic *Eco*RI fragments of normal cellular DNA ranging from 2.6 to 3.0 kb in the λ*gt*11 vector. Two positive clones were isolated, and their restriction maps were determined and were found to be identical; thus, we used only one clone, NHAL26, for further analysis (Fig. 3).

The enzymes *Sma*I, *Pst*I, and *Xba*I cut NHAL26 at the equivalent position as that of the host flanking sequences of AL-26 (Fig. 3), except that the 0.4-kb *Sma*I fragment of NHAL26 was about 3.2 kb in AL-26. This difference is

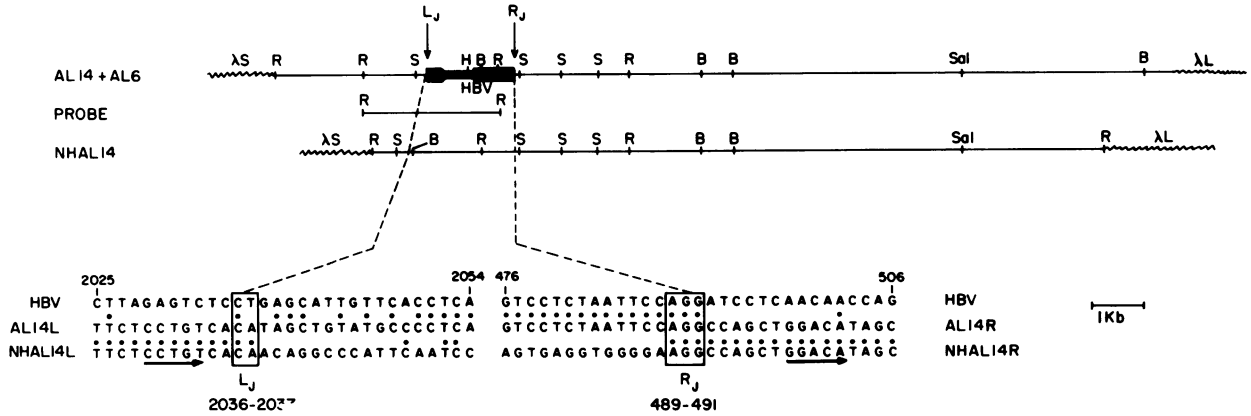

FIG. 2. Restriction maps and partial sequences of AL-14 and NHAL14. The combined restriction map of AL-14 and its homologous clone AL-6 are taken from a previous report (20); wavy lines show the λ arms (S, short; L, long), and solid lines designate the integrated HBV sequences. The DNA probe that was used to isolate the unoccupied site NHAL14 is shown. At the bottom of the figure the sequences of HBV (27) and AL-14L (left host-virus junction), AL-14R (right host-virus junction), and the related sequences in NHAL14 are compared. The sequences at the left- (L_j) and right (R_j)-end junctions are boxed. The following restriction enzymes were used: *Eco*RI (R), *Sma*I (S), *Hin*dIII (H), *Bam*HI (B), and *Sal*I (Sal).

explained by the integration of 2.8 kb of HBV sequences in AL-26 inside the 0.4-kb *Sma*I fragment. This close similarity between the restriction map of the two DNAs indicates that no major deletion or other cellular DNA rearrangements occurred at the integration site.

To determine the sequence of the virus-host junctions, the appropriate DNA fragments of AL-26 and NHAL26 were sequenced and compared with prototype HBV DNA (Fig. 3). This comparison mapped the left end of the virus-host junction at nucleotide 2536 on the HBV genome because downstream from this nucleotide the sequence of AL-26 matched that of the HBV DNA, while the sequence upstream from this point was not homologous to that of HBV but rather to that of the unoccupied DNA (NHAL26). This left-end junction site was about 3 bases away from the one reported previously (Winocour et al., in press). Significantly, the right-end junction (Fig. 5) mapped at nucleotide 1797 of the HBV genome, about 182 base pairs (bp) away from nucleotide 1615, which was reported previously to be the integration site (8, 26, 27). It is also evident that in contrast to AL-14, here only a small DNA fragment of 17 bp was deleted at the integration site. Interestingly, inspection of the adjacent host sequences revealed a stretch of 12 bases that was homologous to HBV DNA (positions 1891 to 1902). This dodecanucleotide sequence was positioned 100 bases downstream from the right end of the integration site, which is

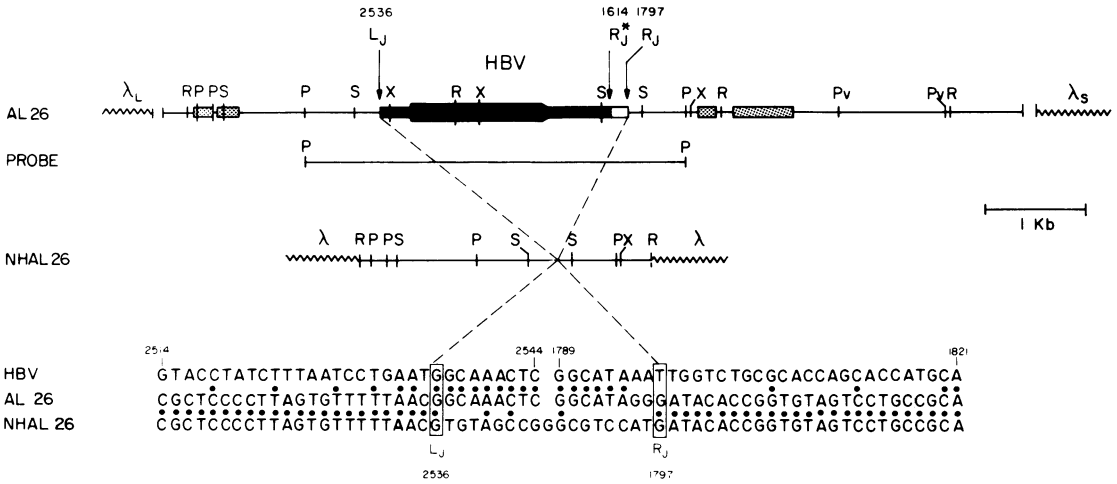

FIG. 3. Restriction maps and partial sequences of AL-26 and NHAL26. The restriction map of AL-26 is taken from a previous report (20). The DNA probe that was used for isolation of NHAL26 is shown. At the bottom of the figure the sequences of HBV DNA, AL-26, and NHAL26 are compared. Homologous bases are marked by dots. The black box on the restriction map of AL-26 is the integrated HBV DNA; the white box is the newly detected HBV DNA, which is positioned between the previously mapped right-end junction (R_j*) on our new map (R_j); the dotted boxes indicate the presence of *Alu* family repeats. The enzymes and the symbols used here are described in the legend to Fig. 1, except that here we also used the enzymes *Xba*I (X), *Pst*I (P), and *Pvu*II (Pv).

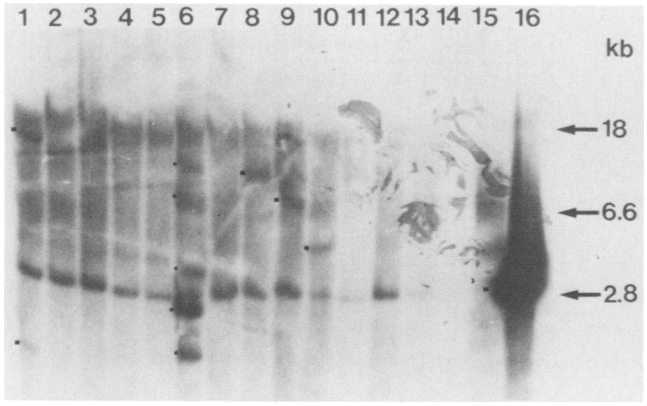

FIG. 4. Southern blot analysis of AL-26 unoccupied sites in normal and hepatoma cells. DNA samples (20 μg) were prepared from hepatoma cell lines Hep3B (lane 1), SK-Hep1 (lane 2), HepG2 (lane 3), Mahlavu (lane 4), and Alexander (lane 6); primary hepatoma cells (lanes 7 to 10); nonhepatoma human tumor cell lines GM3299 (lane 11), HeLa (lane 12), and HL60 (lane 13); normal human cells (lane 5); *Drosophila* (lane 14) and monkey cells (lane 15); and monkey cells containing AL-26 DNA (lane 16). These samples were digested with *Eco*RI, electrophoresed in agarose gel, Southern blotted, and hybridized to the AL-26 probe shown in Fig. 3. The HBV-positive bands are marked by dots.

almost similar in length to the size of the displaced HBV DNA, measured from the integration site (1797) to these 12 bases (1891) (see Fig. 7).

Detection of an unusual HBV sequence in AL-26. Because the 182 bp at the 5′ end of the integration site in AL-26,

```
HBV 1465   CTCGGGGCC GTTTGGGGCT CTACCGTCCC CTTCTTCTTC TGCCGTTCCG GCCGACCACG
           * ******* * ***** ** *** ****** ******* ** ****** ***  *********
AL26   1   CCCGGGGCC GCTTGGGACT CTATCGTCCC CTTCTTCGTC TGCCGTACCG TCCGACCACG
           **** ****     **    *     ** **  ** *  **** * *  *  **  **  * ***
WHV 1555   CCCGAGGCC TTCTGCCGGT TCTGCGGCCT CTCCCGCGTC TTCGCTTTCG GCCTCCGACG

HBV 1524   GGGCGCACCT CTCTTTACGC GGTCTCCCCG TCTGTGCCTT CTCATCTGCC GGACCGTGTG
           **** ***** ********** *** ****** ***** ***  ********** ** ******
AL26  60   GGGCCCACCT CTCTTTACGC GGTTTCCCCG TCTGTTCCTG CTCATCTGCC GGTCCGTGTA
            * *  * ** * ****  **  *  ******  ***** *   ***  *  ** *********
WHV 1614   AGTCGGATCT CCCTTTGGGC CGCCTCCCCG CCTGTTTCGC CTCGGCGTCC GGTCCGTGTT

HBV 1584   CACTTCG CT TCACCTCTGC ACGTCGCATG GAGACCACCG TGAACGCCCA CCA  AATAT
           ***** * ** ********** *** ****** **    **   **  * ***  ***  ** *
AL26 120   CACTTTG CT TCACCTCTGC ACGACGCATG GA    AC   TGG CACCCG CCATGAACA
             *** * *  ****** *** * **   **  *     **   ***   **   *  *****
WHV 1674   G CTTGGTCG TCACCTGTGC A GA   ATT GCGA  ACCA TGGATTCCAC CG TGAACTT

HBV 1641   TGCCCAAGGT CTTACATAAG AGGACTCTTG GACTCTCAGC AATGT  CAA CGA  CCGAC
                    **********  *********  *** * *   *****  ***   *    ***
AL26 172            CTTACATAAG TGGACTCTTG AACTGT ATT AATGTAGCAA TAATTATGAC
                    * * ** **   *    ****      *     ** *   ***   *
WHV 1726   TGTCTCCTGG CATGCA AAT CGTCAACTTG GCATG C  C AA G   CAA GGA  CCTT

HBV 1697   CTTGAGGCAT       ACTT CAA AGACTG TTTGTTTAAA GACTGGGAGG AGTTGGGGGA
            *** *  **         *  ** * *    * ** *   *  * ***  ** *     *
AL26 221   ATTGTGTAAT GTCGGTTATG TAACA ATCA ATAATTAATG GTTTAGGAAC AGATTAATGT
            * *  *  *        ***  *** * **** ** *****      * ***  **  *  *
WHV 1775    TGGACTCCT       TATA TAAGAGATCA ATTATTAACT AAATGGGAGG AGGGCAGCAT

HBV 1750   GGA GA TTA GGTTAAAGGT CTT TGTACT AGGAGGCTGT  AGGCATAAA TTGGTCTGCG
             * *    * ****** * *     * ***   *********  *******          *
AL26 280   TTATGTCCAA GGTTAATGAT AAACTATACA TGGAGGCTGT TAGGCATAGG GATACACGG
           * *  *** * * ***   ** *   * **    *********  *******       **  *
WHV 1828   TGA  TCCTA GATTATCAAT ATT TGTATT AGGAGGCTGT  AGGCATAAA  TG CATGCG
                                                                 **********
NHAL26 305                                                       GATACACGG
```

FIG. 5. Comparison of the sequences of the novel HBV DNA in AL-26 with prototype HBV and WHV DNAs. The sequence of a portion of AL-26 is compared with those of the prototype HBV (25) and WHV (5) DNAs, in a manner that displays maximal homology. The homologous bases are marked by asterisks.

```
1    AGCGGGCAGT GCAGCTGCTT GGACAGCAGG GGTGTTTCTT CAACCCAGGC TGCCCTCCTT
              Lj
61   CTCCTGTCAC AACAGGCCCA TTCAATTCTG AACCTGCAAG CCAACTCCAA TCCAACCCAG

121  ATCCTCAACA ATGTGCCAGA CGCCAGGTCC AGAAGCACGC ACCTACCATG GGGGCCTCTT

181  CCTCCTTTCT GGACGGGGTT CCGCTTTGGC CCCTGATGCT AACCTCGCCC CGTGTGCCGT

241  GCAGCTGGAG TCAAACGGAG GCAGACAGCG CATGAGGTGA CAAGCCACAC AGCAGGGTTC
                                                              933-c

                                                  847-ccacac ccctgtcccc

301  AGGGCAGCCA CTCCCCACTC CCCACTCCCC ACTCCCGCGG CATCTGTGCC CC  ACCCGGGT
     cgggaccect gtccccacac ccctgtcccc acaccc-897 (73%)
        904-cca caccectgtc cccacaccc tgtccccaga cccctgtccc caggacccctgt
     aggacc cct gtccccacac ccctgtcccc acaccc-797 (69%)
        804-cca caccectgtc cccacaccc tgtccccaggaccccctgtccc cac accectgt
        747-cca caccectgtc cccacaccc tgtccccaca cccctgtccc cag accectgt
        634-cca caccectgtc cccacaccc tgtccccaga cccctgtccc cac accectgt
        564-cca caccectgtc cccacaccc tgtccccaca cccctgtccc cgggaccectgt

361  GCCAAG CCAC TGTGCCATGC ATCCCCCCAC CCCCAACCTG CTGCAGGTGT GTGGCCTTTG

     ccccacaccc tgtccccagg accectgtcc ccacaccc-811 (65%)

     ccccacaccc tgtccc-733 (71%)
     ccccacaccc tgtccccggg accectgtcc ccacaccc-655 (65%)
     ccccagaccc tgtccc-564 (72%)
     ccccacaccc tgtccccggg acccc-484 (65%)

421  CCTCTTTTCC CAGGGGGAAC CAAAAACCCT ..... 1800 bp .....

2251 TTATTTATTA CATACATTCA CACAGAAACA CATATACAGA TATTACAATC CTCTGGCCTA

2311 CTTCACCCAC TACTAGTACT CCTCTCAAGC CTGATGGTCC ACATTGCAAC ACTGACAGCT

2371 GTAGCAGACA GCCTGGCAGC TCAGCAGCAC AGGGACCTGC CATGCCTTGG GGAGTACCAC

2431 AGCGCCAAGA GTCTGCATAA ACAAGGTCAC CTTGTCGCGC CACGCCAGGC CACCTCTCCT

2491 CTGCTCCAAA GGCCAGCAGC GGGGCCTGAA GATCTGGGTT TCCTGCTGCC CTGCCTCACC

2541 AGGCTGAGTG ATGGGTGCAG GACTGCCCAG AGCCCTGGAG CATGCAATGT CCAGTACCCT
                                                                    RJ
2601 CCCCTCTTGT GCATCTGCCA GTGGGGGCTC CAGGCTCCAT GTGCCAGTGA GGTGGGGAAG

2661 GCCAGCTGGA CATAGCAGGC CCGGAGTGTC CCCCTGTGAG GGACCGTGGG AACCAGATCA

2721 CGTGGGCCTC CCTCCCCCTG TACCCCATTG CTCTCCTGTT GTCACACTCA GTCCCACAAA

2781 GGCAGCCCC
```

FIG. 6. Homology of the sequence of NHAL14 to the human insulin gene polymorphic region. The sequences of the unoccupied site of AL-14 (NHAL14) corresponding to the left (L_j) and right (R_j) virus-host junctions are compared with the repetitive DNA sequences found previously at the polymorphic region of the insulin gene (1) (lowercase letters). The position of the insulin gene sequences and the presence of homology of these sequences to NHAL14 are indicated. The deleted DNA sequence in AL-14 is marked by lines.

which was previously believed to be of cellular origin, were not present in the unoccupied site, we performed Southern blot analysis of genomic DNAs by using a specific probe for these 182 bp, and we failed to detect hybridization (data not shown). This supports the conclusion that, in contrast to what was previously believed, these 182 bases are not of cellular origin. We therefore performed a computer search to test the possibility that these 182 bases might be of viral origin. We found that these 182 bp are highly homologous to sequences in HBV and woodchuck HBV (WHV) DNA (Fig. 5). Therefore, in AL-26 the integrated HBV sequences are not terminated at position 1615 on the HBV genome, as was proposed before, but there is significant homology between HBV DNA and AL-26 sequences up to position 1797 on the HBV genome, where the sequence is identical to that of the unoccupied DNA (NHAL26) (Fig. 5). This novel HBV sequence was only about 45% homologous to all the HBV subtypes published so far. A portion of this sequence, from base positions 237 to 290 on AL-26 DNA, was significantly

more homologous to WHV (32 of 53 bases; 60%) than to HBV (21 of 53 bases; 40%).

The unoccupied sites contain minisatellite-like sequences. To understand the nature of the loci at which HBV DNAs were integrated, we performed a computer search of the cellular side of the junctions versus the gene data bank. In NHAL14, the unoccupied site of AL-14, sequences about 240 bases downstream from the left-end integration site were found to be highly homologous to the polymorphic region of the human insulin gene (Fig. 6). This region contains sequences with short repeat units that are characteristic of insulin minisatellite sequences (1). In AL-26 HBV was also integrated at sequences with short repetitive units (Fig. 7) that were highly homologous to those found previously in the human ζ-globin gene (16).

DISCUSSION

Integration of HBV next to the cellular minisatellite-like sequences. It has been reported previously (8, 20) that the Alexander cell line contains at least seven integrated forms of HBV DNA. The nature of the normal homologous cellular alleles of the integration site has not yet been defined, except for clone AL-23 (20), which has been characterized to be made of satellite III sequences (19). Short repetitive elements, albeit of a different kind, were also found at the unoccupied sites of the two additional clones, AL-14 and

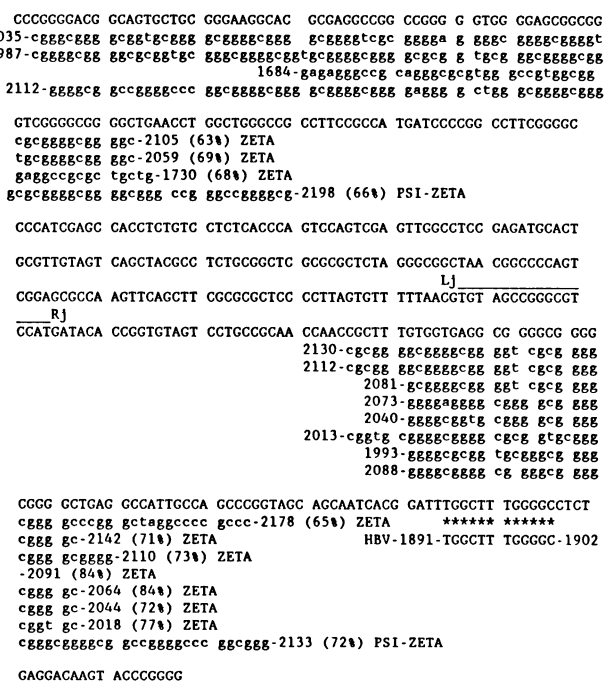

FIG. 7. Homology of the sequence of the NHAL26 to human ζ- and ψ-ζ-globin gene repetitive elements. The sequence of the unoccupied site of AL-26 (NHAL26) corresponding to the left (L_j) and right (R_j) ends of the integrated HBV DNA is compared with the repetitive DNA sequence previously found in ζ-globin and ψ-ζ-globin genes (16) (lowercase letters). The position of the related globin gene sequences and the presence of their homology to NHAL26 sequences are indicated. The asterisks show the homology found between the HBV DNA and the host sequences. The deleted DNA sequences in AL-26 are marked by lines.

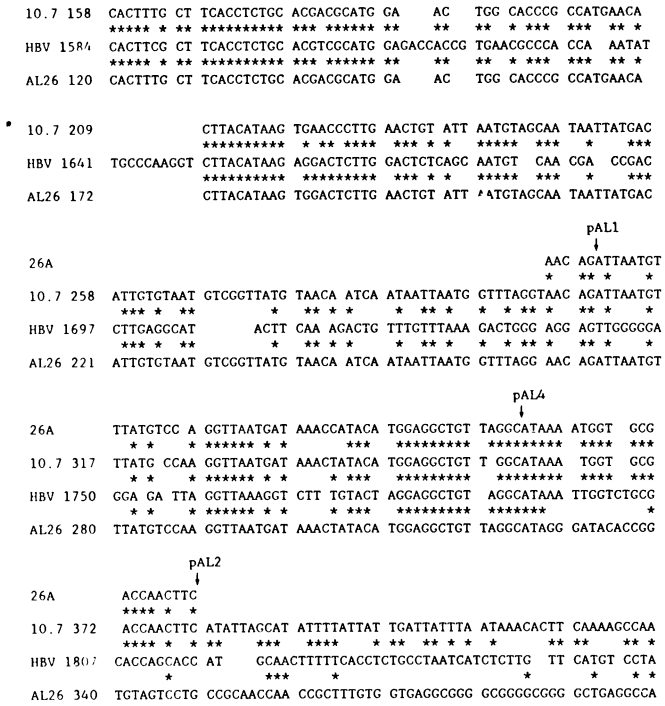

FIG. 8. Comparison of the sequences of the prototype HBV DNA with that of the novel HBV DNA in AL-26, A-10.7, and 26A. The sequences of A-10.7 (26) and 26A (15) were obtained from previous reports. The asterisks indicate the homology found to HBV DNA, and pAL1, pAL4, and pAL2 indicate the 3′ ends of HBV mRNAs found in Alexander cells (15).

AL-26, analyzed here. AL-14 contains sequences that are homologous to the repetitive elements found previously (1) to account for the DNA polymorphism of the 5′-end region of the insulin gene. It was recently proposed (7, 21) that a human genome contains many dispersed tandem repetitive elements, named minisatellite sequences, that share a 10- to 15-bp core sequence similar to the generalized recombination signal X (5′-GCTGGTGG-3′) of *E. coli*. The minisatellite sequences are highly polymorphic and exhibit a high rate of unequal exchange, which suggests that they are hot spots for meiotic recombination. These repetitive sequences display similarity to the tandem repeat of the insulin gene, and indeed, evidence for increased recombination near the human insulin gene has been reported recently (4). Therefore, it is tempting to speculate that the repetitive sequences found next to the integration site of AL-14 are also hot spots for recombination.

The unoccupied site of AL-26 also contains short repetitive sequences that are highly rich in cytosines and guanines, which is a characteristic feature of minisatellite sequences (7). These repetitive sequences are homologous in part to those found previously in the intron of the human ζ-globin gene (16).

We have recently shown (19) that one of the integrated HBV DNAs in Alexander cells is integrated at cellular satellite sequences. Thus, in all cases tested so far, the integration of HBV DNA in Alexander cells occurred at or next to cellular short repetitive units that might serve as recombination signals. The implication of these observations is that the cellular integration sites are not random. On the

basis of cellular-viral sequence homologies close to the integration sites, it was previously proposed (9) that these stretches of sequence homology may facilitate recombination. Indeed, we found that the unoccupied site of AL-26 contains a stretch of 12 bases which is fully homologous to HBV DNA. Therefore, it seems that cellular DNA sequences, which display sequence homology to viral DNA, as well as short cellular repetitive sequences, are preferred target sites for HBV integration.

Detection of a novel HBV sequence in the Alexander cell line. We have reported previously (20) the cloning and sequence analysis of AL-26. The HBV sequence of AL-26 is about 95 to 98% homologous to published prototypes of HBV DNA. A sharp transition from high to low homology is observed at nucleotide 1615 on the HBV genome next to the viral DR2 sequence; therefore, this position was considered previously to be the virus-host junction (26, 27). We have shown here, however, that the 182 bp immediately to the right of the AL-26 3′ end, which were expected to belong to the host sequence, are missing not only from the AL-26-related, unoccupied site but also from the entire cellular genome. Detailed analysis of these 182 bp revealed that they are 40% homologous to those of HBV DNA at a region corresponding to that of the viral sequence between DR2 and DR1. We therefore propose that these 182 bp are viral sequences. The possibility that this diverged HBV sequence that was detected in AL-26 is the product of extensive random mutations of HBV DNA subsequent to its insertion in the host genome is very unlikely. This is mainly because such random mutations cannot explain the high homology observed between this sequence and WHV DNA at the region where only poor homology to HBV DNA is found (Fig. 5). We propose that AL-26 HBV DNA is a novel HBV sequence. This idea is further supported by the fact that several other integrated forms of HBV DNA in Alexander cells also contain this novel HBV sequence, i.e., clones A-10.7 and A-10.5, which have been isolated by Koch et al. (9), and 26A, which was found by Ou and Rutter (15). Furthermore, we compared the sequences of these clones (except for A-10.5, for which the sequence is not yet published) and AL-26 with that of prototype HBV DNA (Fig. 8). The homology between HBV and 26A and A-10.7 sequences extends beyond nucleoide 1795, where the integration site of AL-26 was mapped (Fig. 3). This implies that the integration sites of these clones are not identical, which is in sharp contrast to the previous assumption that all these clones are the product of amplification of a single integrated HBV DNA (8). Our present results indicate that each of these integrations is derived from an independent integration of this novel HBV DNA. It is also demonstrated in Fig. 8 that the 3′ ends of HBV mRNAs produced by Alexander cells, which were mapped previously by analysis of the cDNA clones pAL1, pAL4, and pAL2 by Ou and Rutter (15), are all located at this unusual HBV sequence and not at the adjacent host sequences, as was suggested previously (15). Therefore, the idea of synthesis of virus-host read-through transcripts in Alexander cells should be reevaluated.

ACKNOWLEDGMENTS

We thank D. Shafritz and E. Winocour for critical reading of the manuscript and D. Fisher for supplying us with the blood cells.

This research was supported by grants from Israel Cancer Association 5/86, Israel Cancer Research Foundation, and Leukemia Research Foundation, Chicago, Ill.

LITERATURE CITED

1. **Bell, G. I., M. J. Selby, and W. J. Rutter.** 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature (London) **295:** 31–35.
2. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196:**180–182.
3. **Bullock, P., W. Forrester, and M. Botchan.** 1984. DNA sequence studies of simian virus 40 chromosomal excision and integration in rat cells. J. Mol. Biol. **174:**55–84.
4. **Chakravarti, A., S. C. Elbein, and M. A. Permutt.** 1986. Evidence for increased recombination near the human insulin gene: implication for disease association studies. Proc. Natl. Acad. Sci. USA **83:**1045–1049.
5. **Galibert, F., T. N. Chen, and E. Mandart.** 1982. Nucleotide sequence of cloned woodchuck hepatitis virus genome: comparison with the hepatitis B virus sequence. J. Virol. **41:**51–65.
6. **Goodbourn, S. E. Y., D. R. Higgs, J. B. Clegg, and D. J. Weatherall.** 1983. Molecular basis of length polymorphism in the human zetaglobin gene complex. Proc. Natl. Acad. Sci. USA **80:**5022–5026.
7. **Jeffreys, A. J., V. Wilson, and S. L. Thein.** 1985. Hypervariable minisatellite regions in human DNA. Nature (London) **341:** 67–73.
8. **Koch, S., A. F. Von Loringhoven, P. H. Hofschneider, and R. Koshy.** 1984. Amplification and rearrangement in hepatoma cell DNA associated with integrated hepatitis B virus DNA. EMBO J. **3:**2185–2189.
9. **Koch, S., A. F. Von Loringhoven, R. Kahamann, P. H. Hohfschneider, and R. Koshy.** 1984. The genetic organization of integrated hepatitis B virus DNA in the human hepatoma cell line PLC/PRF/5. Nucleic Acids Res. **12:**6871–6886.
10. **Koshy, R., P. Maupas, R. Muller, and P. H. Hofschneider.** 1981. Detection of hepatitis B virus specific DNA in the genomes of human hepatocellular carcinoma and liver cirrhosis tissues. J. Gen. Virol. **57:**95–102.
11. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15:** 687–701.
13. **Marvo, S. L., S. R. King, and S. R. Jaskunas.** 1983. Role of short regions of homology in intermolecular illegitimate recombination events. Proc. Natl. Acad. Sci. USA **80:**2452–2456.
14. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and DNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
15. **Ou, J. H., and W. J. Rutter.** 1985. Hybrid hepatitis B virus-host transcripts in a human hepatoma cell. Proc. Natl. Acad. Sci. USA **82:**83–87.
16. **Proudfoot, N. J., A. Gil, and T. Maniatis.** 1982. The structure of human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell **31:**553–563.
17. **Rigby, P., M. Dickmann, 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.
18. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
19. **Shaul, Y., P. D. Garcia, S. Schonberg, and W. J. Rutter.** 1986. Integration of hepatitis B virus DNA in chromosome-specific satellite sequences. J. Virol. **59:**731–734.
20. **Shaul, Y., M. Ziemer, P. D. Garcia, R. Crawford, H. Hsu, P. Valenzuela, and W. J. Rutter.** 1984. Cloning and analysis of integrated hepatitis B virus sequences from a human hepatoma cell line. J. Virol. **51:**776–787.
21. **Smith, G. R.** 1983. Chi hotspots of generalized recombination.

Cell **34**:709–710.
22. **Southern, E.** 1975. Detection of specific sequence among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503–517.
23. **Sternberg, N., D. Tiemeier, and L. Enquist.** 1977. In vitro packaging of a Dam vector containing EcoRI DNA fragments of E. coli and phage P1. Gene **1**:255–280.
24. **Tiollais, P., C. Pourcel, and A. Dejean.** 1985. The hepatitis B virus. Nature (London) **317**:489–495.
25. **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 genetic. Academic Press, Inc., New York.
26. **Von Loringhoven, 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.
27. **Ziemer, M., P. D. Garcia, Y. Shaul, and W. J. Rutter.** 1985. Sequence of hepatitis B virus DNA incorporated into the genome of a human hepatoma cell line. J. Virol. **53**:885–892.