Tight Clustering of Human Hepatitis B Virus Integration Sites in Hepatomas Near a Triple-Stranded Region

CHIAHO SHIH,^{1*} KATHY BURKE,² MIN-JI CHOU,³ JEROME B. ZELDIS,⁴ CZAU-SIUNG YANG,³ CHUE-SHUE LEE,³ KURT J. ISSELBACHER,⁵ JACK R. WANDS,⁵ AND HOWARD M. GOODMAN²

Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059¹; Department of Molecular Biology² and Gastrointestinal Unit,⁵ Massachusetts General Hospital, and Departments of Genetics² and Medicine,⁴ Harvard Medical School, Boston, Massachusetts 02114; and College of Medicine, National Taiwan University, Taipei, Taiwan³

Received 26 March 1987/Accepted 17 July 1987

The open circular genome of human hepatitis B virus (HBV) is known to contain a partially double-stranded DNA with a single-stranded gap region of variable length. This circular structure of the genome is maintained by base-pairing of the 5' ends of the two DNA stands, the long or L(−) strand and the short or S(+) strand. By cloning, mapping, and sequencing studies, we have localized three recombinational junctions of the integrated HBV in two hepatoma samples, HT14 and FOCUS. Breakpoints of recombination derived from these results and those of others appear to be clustered and coincidental with the identified 5' or the deduced 3' end of the long-strand DNA, respectively. Statistical analysis of these results supports the hypothesis that integration preferentially occurs in an extremely narrow region on the HBV genome. This site-specific recombinational mechanism appears to be conserved among different HBV subtypes. No extensive sequence homology was found between each pair of the recombining parental molecules; however, at the site of crossover, 2- to 3-base-pair junctional homology was consistently observed. Examination of the patterns of the integrated HBV DNAs allowed us to categorize these various patterns into four different groups according to their end specificity and strand polarity. The molecular form of relaxed circle is proposed to be one major substrate for HBV integration. The effect of free strand in the integration of HBV is emphasized in this model. Unlike any other known DNA animal viruses, the site specificity of HBV integration appears to be similar to that of the retroviruses.

Human hepatitis B virus (HBV) has been known to be associated with a wide variety of diseases, such as acute hepatitis, chronic hepatitis, and the development of hepatocellular carcinoma (HCC) (1). The molecular basis of the link between HBV infection and hepatoma development has thus far remained elusive. Since HBV DNA is integrated in most hepatomas associated with chronic HBV infection (53), it has been speculated that viral integration and consequent insertional mutagenesis may contribute to the oncogenic process in the liver (35). This speculation has gained some support. (i) The nonacute type of oncogenic retroviruses can integrate into the vicinity of an endogenous cellular protooncogene (e.g., avian leukosis virus [ALV]/bursal lymphoma [33, 40], ALV/erythroblastosis [13], and mouse mammary tumor virus/mammary carcinoma [34]), (ii) HBV appears to be related to retroviruses. It has been proposed by Summers and Mason that duck hepatitis B virus replicates its DNA genome via an RNA intermediate (50). By sequence comparison, a common evolutionary origin of HBV and retroviruses has been reported recently (29, 54). (iii) The long latency period between HBV infection and the development of hepatoma is consistent with a tumor induction pattern characteristic of the nonacute-type oncogenic retroviruses.

There are several animal model systems for the study of HBV, including woodchuck hepatitis virus, ground squirrel hepatitis virus, duck hepatitis B virus (DHBV) (reviewed in reference 49), and tree squirrel hepatitis virus (10). These viruses have common structural features and genetic organization and thus belong to the same group, collectively

called hepadnaviruses. Viral integration in HCC has been demonstrated in human (reviewed in reference 53), woodchuck (35), ground squirrel (24), and duck systems (62). In both human and woodchuck studies, extensive rearrangement of oligomeric viral DNA has been reported (35, 63).

The structure of HBV genomic DNA is a small circular molecule (37, 42). This circular structure appears to contain a partially double-stranded DNA with a single-stranded gap region (51), which is variable in size from particle to particle (22). The circular conformation is open and held together via a cohesive overlap (approximately 225 base pairs [bp]) between the 5' ends of the long or L(−) strand and the short or S(+) strand (45, 51). An 11-bp direct repeat (DR1 and DR2) has been found to be located near the extremities of the cohesive ends (6). The copy near the end of the long strand is designated DR1 and that near the 5' end of the short strand, DR2. These direct repeats have been shown to be involved in the initiation step of DNA synthesis of L(−) or S(+) strands (23, 31, 46, 59).

Despite the fact that the mode of DNA replication of HBV is similar to that of retroviruses (50), integration of HBV does not appear to be an essential step in its life cycle (reviewed in reference 53 and in W. S. Mason, J. M. Taylor, and R. Hull, *Adv. Virus Res.*, in press), although it can and usually does occur in HCC. It is therefore interesting to ask whether such integrations of HBV in HCC show any features of site specificity or structural mutagenesis. Several reports on the DNA sequence of the HBV integration site have been published (5, 6, 20, 21, 30, 61, 63). As far as the target of HBV insertion on the host chromosome is concerned, there appears so far to be no preferred site. In two occasions when the preintegration sites of the cellular DNA

* Corresponding author.

were analyzed and compared with the allele with integrated HBV, only minor or no rearrangement was detected (5, 61). However, there are examples as well where extensive rearrangement has occurred at the integration site (30, 44). As far as the integration site of the viral DNA is concerned, both a recombinational mechanism via the single-stranded gap region (21) and preference for integration within the double-stranded region (63) of the viral genome have been proposed following characterization of one or more of the many integrations in the same hepatoma cell line (PLC/PRF/5). A similar analysis has been performed with DNA from primary tumors. Observation of specific breakpoints of recombination within an 11-bp direct repeat (DR1 or DR2) on the HBV genome in two independent hepatomas has led to the hypothesis of a site-specific integration mechanism (6). In both cases the integrated HBV DNA appears to have one fixed end at the virus-cell junction within the DRs, while the other end appears to be at variable positions (6). By analogy with retrovirus integration, it was speculated that a 2-bp deletion (TT) from within the DRs (TTCACCTCTGC) may occur upon integration (6). Additional supporting data for this hypothesis of sequence-specific integration within the DRs has not yet been reported. In fact, this sequence-specific mechanism of integration via the DRs was not supported by studies of human hepatoma-derived cell lines huSP (30) and huH2-2 (61). In the former case, the integration site of HBV is localized in the middle of the cohesive-end region (see Fig. 2) (30). In the latter case, one virus-cell junction is about 26 nucleotides away from DR1 (61). The data from huH2-2 and huSP appear to be more consistent with the hypothesis that the whole cohesive-end region is predisposed to integration activity.

We have identified two hepatoma samples with a single site of integrated HBV. Here we report the characterization by cloning and sequencing of three recombinational junctions of HBV from two different hepatomas. Comparison of our results with those of others (5, 6, 20, 30, 36, 58, 61) has led us to the conclusion that the preferred site of HBV recombination may be localized in an extremely narrow region (ca. 9 bp) within the cohesive overlap (ca. 200 bp). In our model, site-specific recombination may occur not only within the direct repeats as suggested previously (6), but it may often occur outside in the vicinity of the direct repeats. This phenomenon is discussed with respect to various interpretations, including one dealing with a putative triple-strand structure created on the circular viral genome as a result of a short terminal redundancy of the L(–) strand (46, 59; Mason et al., in press) (see Fig. 2). A model hypothesizing the relaxed circular DNA as a major preintegration substrate is discussed in detail.

(A preliminary account of these results was presented at the Cold Spring Harbor meetings on HBV held 30 May 1985 and 30 August 1986.)

MATERIALS AND METHODS

Human hepatoma DNA. HT14 is an HCC from an Oriental patient in Taiwan. This patient was male, 37 years old, and serum positive for HBV surface antigen and alpha-fetoprotein. The FOCUS primary tumor was from a Caucasian patient at Massachusetts General Hospital. This patient was male, 63 years old, and serum negative for both surface antigen and alpha-fetoprotein, but positive for antibodies to HBV surface antigen (16). High-molecular-weight DNAs were extracted as described elsewhere (47).

Genomic library construction. Genomic DNAs from primary tumors HT14 and FOCUS with *Mbo*I partial digestion

were size fractionated by sucrose gradient centrifugation as described previously (47). Fractions with a size range between 18 and 21 kilobases (kb) were pooled. Such fragments were ligated to the phage arms of the EMBL-3 vector (11) and packaged in vitro by the procedure of Hohn (17). An estimated 2×10^6 PFU from both HT14 and FOCUS were plated onto *Escherichia coli* LE392 cells and screened with the radiolabeled 3.1-kb *Eco*RI vector-free HBV DNA fragment from the pAO1 plasmid (4) by the procedure of Benton and Davis (2).

Plasmid subcloning and DNA sequencing. Briefly, a 1.8-kb *Hind*III-*Bam*HI fragment containing the integrated HBV DNA of FOCUS hepatoma was subcloned into pUC12. This plasmid DNA was cleaved with *Xba*I and then 5' or 3' end labeled at this site, following the sequencing method of Maxam and Gilbert (26). A 0.65-kb *Pst*I-*Pst*I fragment derived from HT14 HBV and its preintegration site were subcloned into pUC12, linearized with enzymes which cut in the polylinker region, and end labeled at these sites for sequencing. A University of Wisconsin Genetics Computer Group "BESTFIT" computer program (8) was used for the comparisons of sequence homology.

RESULTS

Cloning, mapping, and sequencing. As an initial approach to understanding the relationship between HBV integration and hepatocyte transformation, we used the Southern blot procedure (48) to screen a number of human hepatomas for HBV DNA. The goal was to identify tumors containing only a single site of HBV integration. Two of 30 hepatomas, namely HT14 and FOCUS (16), appeared to fulfill the single-site criterion. The remaining tumors contained between zero and seven independent copies of integrated HBV DNA (data not shown). Digestion of high-molecular-weight DNA from both HT14 and FOCUS with various restriction enzymes known to cut infrequently within the HBV DNA and hybridization to an HBV DNA probe showed a single hybridizing band (data not shown). Neither integrated nor unintegrated DNA was detectable in the uninvolved region of the liver of the FOCUS patient, but covalently closed and relaxed circular forms of HBV DNA were seen in the normal liver of other hepatoma patients (data not shown). Uninvolved liver was not available from the HT14 patient.

Genomic libraries of HT14 and FOCUS DNA were constructed in the bacteriophage vector EMBL-3 (11). Phage isolates with positive hybridization signals to HBV DNA were plaque purified and their restriction maps were determined (data not shown).

The relevant portions of the DNA fragments spanning the integration or preintegration site from both HT14 and FOCUS tumors were subcloned into pUC12 (57) and sequenced by the Maxam and Gilbert procedure (26). FOCUS tumor DNA and the phage insert DNA had identical restriction maps. In contrast, differences could be detected between the HT14 phage DNA and genomic DNA by Southern blot analysis. Both sides of the flanking cellular DNA at the HT14 HBV integration site contained concatemers of centromere-specific repeats (data not shown). We therefore propose that deletions occurred during cloning due to the recombination between tandem concatemers of centromere repeats, as reported previously (60).

HT14 viral-cellular junction. The virus-cell junction of HT14 was either at position 1825/1828 or 1830/1831 (Fig. 1a) when the junctional sequences of HT14 were compared with the nucleotide sequence of ayw-HBV (14). The same results

FIG. 1. Mapping the recombinational junctions of HBV. (a) Localization of the HT14 HBV virus-cell junction by comparison of the nucleotide sequences of HBV, HT14, and HT14NCP (normal counterpart). The direction from left to right corresponds to the polarity from 5' to 3'. There are two possibilities for the alignment. In the first possibility, there is no homology at the site of crossing over. One extraneous nucleotide C (underlined) was added to the ends. The mapped junction of HT14 HBV is about 5 bp downstream from that of IA22 (6). In the second possibility, a 2-bp homology (CA) at the site of crossover occurs. One gap with a 2-bp deletion (TC) at 1830-1831 was introduced to maximize the extent of homology. In this case, the mapped junction of HT14 HBV coincides with that of tumor IA22 (6). (b) Localization of an inversional junction of FOCUS HBV by a sequence comparison computer program (8). Note that the L(-) strand DNA has fused into the S(+) strand as a consequence of an inversional recombination. At position 1827, nucleotide A (underlined) has been converted into a C. (c) Localization of a deletional junction of an adr variant (36). (d) Nucleotide sequences of HT14 virus-virus inversional junction. The mutated residues are underlined. To achieve the maximal junctional homology, the nucleotide sequences of HBV subtype adr (36) was used.

were obtained when the nucleotide sequences from other HBV subtypes were used for alignment (12, 19, 36, 39, 56).

When the HT14 sequence was compared with that of the normal counterpart of the unintegrated allele (HT14NCP), the homology appeared to end at a CA sequence (positions 166 and 167), which is 1 bp away from the junction defined by comparison between sequences of HBV and HT14 (Fig. 1a, top). One can interpret these data as if breakage took place between HBV sequence position 1831 and position 168 of normal cellular DNA (Fig. 1a, top). In this model, one extraneous nucleotide, C (underlined), of unknown origin must then be added to the ends. The breakage position at 1830/1831 is about 5 bp away from the breakpoint of IA22, a hepatoma characterized previously (6). An alternative explanation, perhaps as likely, could be that breakage of HBV DNA occurred between positions 1825 and 1826, preceded

or followed by the deletion of a CT (1829-1830 or 1831-1832) or TC (1830-1831) dinucleotide (Fig. 1a, bottom). In this case, a 2-bp (CA) homology at the site of crossing-over was evident (Fig. 1a); limited junctional homology was also seen on other occasions (Fig. 1b and c). This homology therefore generated a junctional ambiguity with three potential breakpoints at 1825/1826, 1826/1827, and 1827/1828. If it is 1825/1826, HT14 will coincide exactly with tumor IA22 (6) at the recombinational breakpoints.

FOCUS junction. A viral-viral inversional junction of FOCUS HBV was located very close to the HT14 junction mentioned above. As shown in Fig. 3b, this inversional joint was formed by fusing an AA dinucleotide at 1818-1819 of the S(+) strand into the core antigen gene region at nucleotide positions 2329 and 2328 of the L(-) strand. A previous study (36) reported a deletional endpoint of an adr variant of HBV

FIG. 2. Putative recombinational hotspot (1818 to 1827) on the HBV genome as shown from the sequence organization of the integrated HBV DNAs in hepatomas (5, 6, 30, 61). The wavy lines indicate flanking cellular sequences; broken lines indicate inversional junctions of HBV DNAs, shown here as solid boxes. Asterisks indicate the locations of DR1 (to the right) and DR2 (to the left). The approximate positions of the virus-cell junctions of FOCUS and HT14 HBV are derived from our unpublished results of DNA sequencing, restriction mapping, and electron microscopy (performed by H. Potter). The question marks indicate that the exact sequences at the junctions are still not available (6). In the case of tumor huSP (30), the inverted duplication of the flanking cellular DNA is reported to be longer than 2 kb. (b) Tight clustering of the ends of recombination, replication, and transcription near the extremities of the HBV genome. Different subtypes of HBV are indicated within parentheses. Data for DT1 and IA22 are from Dejean et al. (6). The HBV deletion variant is from Ono et al. (36). The breakpoints of PLC/PRF/5 are from clones A10.5 and A10.7 (20). The data for huH2-2, HC1707-1, and HC1707-7 are from the discussion section of Yaginuma et al. (61). HBV +6-bp insertion variant is from Will et al. (58). Ends of L(-) and S(+) strands are from Will et al. (59). Open circle at the 5' end of the L(-) strand represents a protein which is covalently bound to this end (15). The wavy line at the 5' end of the S(+) strand represents a capped oligoribonucleotide primer (23). The S(+) strand nucleotide sequence is shown here (14). The unique *EcoRI* site is counted as position 0, clockwise in the direction of the S(+) strand. The dotted, upward arrow of HT14 and DT1 indicates a second potential breakpoint.

that happened to coincide exactly with this FOCUS breakpoint (Fig. 1c). More coincidentally, an AA dinucleotide homology was observed at the site of mutation in both events: the FOCUS inversion and the adr variant 27-bp deletion (Fig. 1b and c). Although the breakpoint of FOCUS was near the HT14 and IA22 breakpoints, it was located outside the direct repeats instead of being within them (Fig. 2b).

HT14 Junction. Another junction of HT14 HBV, located at position 2131-2133/2864-2862, resulted from an inversional recombination, whereby nucleotides 2864 to 2862 of the L(-) strand from the pre-S region fused into nucleotides 2131 to 2133 of the S(+) strand from the core antigen region (Fig. 1d). This junction was far from the other junctions that we have described so far. A 3-bp TGT homology at the junction was observed (Fig. 1d). All three recombinational events whose junctions we have characterized appeared to show base changes near the junctions (data not shown).

Conserved mechanism of recombination. We compared 350 bp (positions 1826 to 2176) of HT14 HBV with the DNA sequences from various HBV subtypes. HT14 HBV was most similar to the adr subtype, at least within this 350-bp region. The pairwise comparisons showed that the heterogeneity between HT14 and adr (36) was only about 4.3%, while HT14 compared with adw (36) or ayw (14) demonstrated an 11% sequence divergence (data not shown), a number approximately the same as that reported previously for different HBV subtypes (36). The subtype of FOCUS, ayw, was also determined by sequence comparisons (data not shown). As shown in Fig. 2b, the ayw HBV showed breakpoints at positions 1591/1592 (tumor DT1), 1817/1820 (FOCUS), and 1825/1826 (tumor IA22), while the adr subtype showed breakpoints at 1817/1820 (deletion variant) or 1825/1828 (HT14). In other words, different subtypes can exhibit the same breakpoint. For example, HT14 (adr) from

Taiwan coincided with IA22 (ayw) from Africa (6), while FOCUS (ayw) from Greece coincided with a deletion variant (adr) isolated in Japan (36). This strongly suggests that this site-specific recombination is a conserved mechanism among HBV serotypes prevalent in different areas of the world.

DISCUSSION

Statistical significance of site-specific recombination. A compilation of our results and others (Fig. 2a and b) shows that there are eight breakpoints of recombination near the DR1 region and two breakpoints near the DR2 region. These results together have made a very strong statistical argument for the existence of a site-specific mechanism of recombination in human HBV.

In support of this hypothesis are results from Hino and Rogler (personal communication) mapping recombinational breakpoints of HBV in additional tumors to the hotspot discussed above.

Replication and recombination. Studies of the synthesis of DHBV DNA have led to the proposal of a reverse transcription model of DHBV replication (50). Recent studies of RNA transcription with DHBV and the other animal models have identified a 3.5-kb pregenomic mRNA (3, 9, 32) that is larger than the DNA genome size and is believed to be a template for reverse transcription. Indeed, the 5' end of hepadnavirus pregenomic RNA has been implicated as corresponding to the 3' end of the L(-) DNA strand (31, 46, 59; Mason et al., in press). The 5' end of the HBV pregenomic RNA has been mapped to position 1818. If the result from the nonhuman hepadna viruses can be extended to the HBV system, then the 3' end of the HBV L(-) strand should be near position 1818 (59). This deduced 3' end of the L(-) strand appears to coincide with the breakpoint of FOCUS HBV and one adr deletion variant, while the 5' end of the L(-) strand also coincides with the HT14 breakpoint.

It has been reported that integrated HBV DNA in hepatoma appears to have one fixed end of virus-cell junction near the DRs, while the other end of the virus-cell junction appears to occur at variable positions (6). As will be discussed further, our results are consistent with this characteristic feature of integrated HBV DNA (Fig. 2a). The variation of the position near the other end can be explained if replication intermediates with a variable size can serve as preintegration substrates, rather than assuming that a secondary event of random rearrangement often occurred at the other end of the full-length integrated HBV DNA. Since little or no rearrangement of cellular DNA at the HBV integration site has been observed in some cases (5, 61), this has provided an additional strong support for the link between preintegration substrate and replication intermediate. As a side note, it is perhaps worth mentioning that integration of retroviruses has been proposed to be mediated via circle junction on the two long terminal repeat circular genome (38). This hypothesis predicts that most of the integrated retroviruses should be structurally intact with a two-long terminal repeat linear configuration. However, deleted forms of provirus have been found in many cases (13, 33, 40, 41). It would be interesting to see whether these "deletions" occur after integration or some of them actually derive from replication intermediates, as for the HBV proposed here.

Existing models of HBV site-specific integration. There are several models being proposed for the site specificity of HBV recombination (6, 21, 61). One model proposed by Dejean et al. involves the assumption of a sequence-dependent integration event from the covalently closed

circles (6). Although this model can explain the end specificity of the integrated HBV DNA quite well, it does not offer a satisfactory explanation of why the other end of the integrated HBV DNA is often at variable positions.

Yaginuma et al. proposed another model of HBV DNA integration via a single-stranded L(-) DNA intermediate (61). This model is based on the observation that the cohesive-end region and gene C are often separated in the integrated HBV DNA. Consistent with this model is an earlier finding of the accumulation of free minus-strand DNA as a replication intermediate in all systems (25; Mason et al., in press). Since our results with HT14 and FOCUS suggest a strong preference for integration to take place near both 5' and 3' ends of the L(-) strand, they therefore seem to be consistent with that model (61) as well. However, this model also predicts a "polarity" of the integrated HBV DNA, i.e., regions more upstream in the L(-) strand, such as gene X, will be preferentially retained in the integrated HBV DNA compared with the downstream regions of core gene. As shown in Fig. 2a, our results with HT14 and FOCUS and the results from DT1 (5, 6) do not appear to lend support to that model, because the X gene region does not appear to be more frequently retained than gene C in these cases. Since the cohesive overlap region is not directly linked to gene C in the integrated HBV DNA of IA22 (6), HT14, and FOCUS, an S(+) linear-strand model does not offer a satisfactory explanation either. In fact, single-stranded plus-strand DNA has never yet been observed anywhere in any hepadna virus system. Although single-stranded minus-strand DNA has been found to accumulate in the virion during DNA replication (25), it has not been detected in the nucleus (28, 55). Further examination of all the integration patterns of HBV DNA currently available in the literature (5, 6, 20, 21, 30, 35, 43, 61) has allowed us to divide these patterns into four different groups according to their end specificity and strand polarity (Table 1). From this table, it is apparent that a new model is required to explain the integration pattern with a strand polarity away from the gene X region, such as HT14 and FOCUS.

Triple-stranded structure and relaxed circle roll-in model. At this stage, we entertain the speculation that relaxed circular DNA may be able to "roll-in" to the host chromosomal DNA via the triple-stranded region near DR1. This hypothesis is consistent with the finding of the relaxed circular form in the nucleus (28).

On the relaxed circle, a putative triple-strand region with a short stretch of terminal redundancy at the extremities of the L(-) strand of HBV DNA can be generated during the process of DNA synthesis (46, 59; Mason et al., in press). This putative structure has been hypothesized to be used for "jumping" by the nascent plus-strand DNA from the 5' to the 3' end of the minus-strand template, allowing the continuation and completion of HBV plus-strand synthesis (46, 59; Mason et al., in press). This kind of triple-strand conformation is reminiscent of the recombinational intermediate formed during the process of strand invasion and displacement (Fig. 3b). If the cellular enzymes involved in the formation of Holliday-like structures can also be involved here by recognizing a similar configuration around the fork point with a triple-strand discontinuity region (Fig. 3a), such a unique structure may increase the probability of a strand transfer or invasion event (18, 27).

It may be important to make a distinction here between the recombinogenic effects of free end from that of free strand. Although free end in general has been considered influential to recombinational activity, it is not clear that the

TABLE 1. Four different integration patterns of HBV DNAs in hepatocytes

		STRAND POLARITY	
		core → surface → X	X → surface → core
END	DR 1	HT14 FOCUS IA22 (ref. 5) HW99-0 (ref. 36)	huSP (ref. 31) huH2-2 (ref. 62) HC1707-7 (ref. 62) HC1707-1 (ref. 62) CW309 (ref. 36)
		I	II
SPECIFICITY	DR 2	DT1 (ref. 6) HW197-2 (ref. 44)	A10.5 (ref. 20)
		III	IV

recombinational activity of HBV integration can be totally explained by free ends. The 5' end of the L(−) strand of the rolling circle is bound by a protein (15), while the 5' end of the S(+) strand is a capped ribonucleotide. The 3' end of the S(+) strand is believed to be associated with an endogenous polymerase and therefore is likely to have no free end either.

The only available free end is the 3' terminus of the L(−) strand. We chose to interpret that this free end may be a more potent DNA donor than usual because it is sitting on a free strand or near a triple-stranded structure (Fig. 3).

We envision that this 3' free end of the L(−) strand may lead the invasion event into a recessed 5' end of cellular DNA, and a transient base-pairing between the viral and the cellular DNAs near the ends in contact will be formed. This predicts a short stretch of junctional homology (Fig. 1). Such transient base-pairing can become stabilized as soon as the copy process of the complementary strand of the L(−) DNA takes place. This process of displacement synthesis will allow the L(−) strand to be peeled off from the S(+) strand, and the growing point will eventually be blocked when it reaches the 3' end of the S(+) strand. At such a position, presumably both strands of viral DNA are tightly clamped by the virally encoded endogenous polymerase. The repair system of the host will have to resolve this problem first and eventually give rise to patterns like HT14, FOCUS, IA22 (6), and HW99-O (35), with end specificity near the 3' terminus of the L(−) strand, but with a strand polarity not containing the cohesive overlap (type I pattern, Table 1).

The type II integration pattern (Table 1) can be explained by the model of Yaginuma et al. (61) as well as by a model with strand invasion from the cellular DNA into the single-stranded gap region of the relaxed circle, followed by displacement synthesis throughout the cohesive overlap region. We do not yet have a satisfactory model for the third integration pattern of HBV DNA, such as DT1 (6) or HW197-2 (43). The type IV integration pattern has rarely been found (Table 1). Such a low frequency of occurrence is consistent with the relaxed-circle model, since the majority of the virion DNA contains a single-stranded gap region from positions 900 to 1600 near DR2 (7). Among these four different integration patterns, end specificity near DR1 is a more frequent event than the pattern with end specificity near DR2. In addition, microheterogeneity of the break-

FIG. 3. (a) Hypothetical equilibrium at the putative triple-strand discontinuity region of HBV (1818 to 1826), underlined here for clarity. It should be noted that the L(−) strand shown here is not lariatlike, i.e., there is no covalent bonding between nucleotide G of the L(−) strand at position 1817 and the nucleotide T of the L(−) strand at its hypothetical 3' terminus. (b) Cartoon illustrating the molecular events of recombination and the similarity between the recombinational intermediate (steps 3 and 4) and the putative triple-strand discontinuity region of HBV. Panels: 1, two different DNA molecules before recombination; 2, single-strand (or double-strand) break; 3, strand invasion and D-loop formation; 4, formation of Holliday junction (18).

points of HBV integration appears to be asymmetrically distributed with regard to the triple-stranded region (Fig. 2b). In other words, we have not observed any breakpoint of virus-cell junction mapped to the right of the 5' end of L(-) strand (i.e., to the right of the cohesive overlap region). Distribution of the breakpoints of recombination to the left of the 3' end of L(-) strand (i.e., to the left of the triple-stranded region within the cohesive overlap; see Fig. 2b) appears to be associated only with those patterns with a strand polarity away from the core gene, such as huSP (position 1740) or huH2 (position 1798).

Finally, it should also be pointed out that our model does not imply that the relaxed circular form is either necessary or sufficient for integration. It is not entirely inconceivable that a less permissive environment for viral replication in the infected host cell will favor the pathway of integration rather than replication.

Conclusions. We have confirmed and refined the site specificity of HBV integration down to an extremely narrow region on the HBV genome. This region appears to coincide with a triple-stranded structure (46, 59; Mason et al., in press) which is partially overlapped with DR1. Although the integration patterns of HBV DNA in hepatomas seem to be highly complicated in general, we have attempted to rationalize this phenomenon by categorizing these patterns into four different groups. The integration patterns of HT14 and FOCUS (type I, Table 1) have led us to a new model of integration which assumes that the identity of the proximal precursor to the integrated HBV DNA may be a relaxed circular form. The potential influence on recombination from the unpaired, single-stranded, or triple-stranded region, in addition to the 3' free end of the L(-) strand of the HBV genome, is discussed. It should be noted that this model does not necessarily exclude other models. Clearly, more integration sites will need to be characterized and compared before any definite conclusion can be reached. Recently it was speculated that integration of a tandem dimer or concatemer may be related to chronic infection (52). In addition, the potential of an oncogenic role of HBV integration in the early stages of hepatoma development remains elusive. It may therefore be interesting to further our understanding of the mechanism of HBV integration in the future. This site-specific mechanism, whether it is mediated via DNA intermediates of single-stranded linear (61), double-stranded closed circular (6), or triple-stranded relaxed circular DNA can be elucidated in molecular detail if a functional assay of recombination can be established.

ACKNOWLEDGMENTS

We thank Jesse Summers for HBV plasmid pAO1, Bruce Burnette for the EMBL-3 phage stock, Mark Conklin for advice on DNA sequencing techniques, and William Mason and John Taylor for stimulating discussions and careful readings of this manuscript. We also thank C. Rogler, R. Miller, W. Robinson, and J. Ou for communication of unpublished results.

This work was supported by grants from Hoechst AG, Research Fund No. 3-71072 of the University of Pennsylvania; by Public Health Service grants CA-16520, CA-43835-01, and CA-35711 from the National Institutes of Health; by BRS grant S07-RR-05415-25; by American Cancer Society Institutional Research grant IN-135; and by a W. W. Smith award to C.S. through the University of Pennsylvania Cancer Center.

LITERATURE CITED

1. Beasley, R. P., C. C. Lin, L.-Y. Hwang, 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. Benton, W. D., and R. W. Davis. 1977. Screening gt recombinant clones by hybridization to a single plaque *in situ*. *Science* 176:180-182.
3. Buscher, M., W. Reiser, H. Will, and H. Schaller. 1985. Transcripts and the putative RNA pregenome of duck hepatitis B virus: implications for reverse transcription. *Cell* 40:717-724.
4. Cummins, I. W., J. K. Browne, W. A. Salser, G. V. Tyler, R. L. Snyder, and J. Summers. 1980. Isolation, characterization, and comparison of recombinant DNAs derived from genomes of human hepatitis B virus and woodchuck hepatitis virus. *Proc. Natl. Acad. Sci. USA* 77:1842-1846.
5. Dejean, A., L. Bougueleret, 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.
6. 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.
7. Delius, H., N. M. Gough, C. H. Cameron, and K. Murray. 1983. Structure of the hepatitis B virus genome. *J. Virol.* 47:337-343.
8. Devereux, J., P. Haeblerli, and P. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-395.
9. Enders, G. H., D. Ganem, and H. Varmus. 1985. Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. *Cell* 42:297-308.
10. Feitelson, M. A., I. Millman, T. Halbherr, H. Simmons, and B. S. Blumberg. 1986. A newly identified hepatitis B type virus in tree shrews. *Proc. Natl. Acad. Sci. USA* 83:2233-2237.
11. Frishauf, A. M., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170:827-842.
12. Fujiyama, A., A. Miyanohara, C. Nozaki, T. Yonayama, N. Ohtoma, and K. Matsubara. 1983. Cloning and structural analysis of hepatitis B virus DNAs, subtype adr. *Nucleic Acids Res.* 11:4601-4610.
13. Fung, Y. K., A. M. Fadly, L. B. Crittenden, and H. J. Kung. 1981. On the mechanism of retrovirus-induced avian lymphoid leukosis: deletion and integration of the proviruses. *Proc. Natl. Acad. Sci. USA* 78:3418-3422.
14. 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.
15. Gerlich, W. H., and W. S. Robinson. 1980. Hepatitis B virus contains protein attached to the 5' terminus of its complete DNA strand. *Cell* 21:801-809.
16. He, L., K. J. Isselbacher, J. R. Wands, H. M. Goodman, C. Shih, and A. Quaroni. 1984. Establishment and characterization of a new human hepatocellular carcinoma cell line. *In Vitro* 20:493-504.
17. Hohn, B. 1979. In vitro packaging of cosmid DNA. *Methods Enzymol.* 68:299-309.
18. Holliday, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* 5:282-304.
19. Kobayashi, M., and K. Koike. 1984. Complete nucleotide sequence of hepatitis B virus DNA of subtype adr and its conserved gene organization. *Gene* 30:227-232.
20. Koch, S., A. F. Von Loringhoven, P. H. Hofschneider, and R. Kochy. 1984. Amplification and rearrangement in hepatoma cell DNA associated with integrated hepatitis B virus DNA. *EMBO J.* 3:2185-2189.
21. Koshy, R., S. Koch, A. F. Von Loringhoven, R. Kahmann, K. Murray, and P. H. Hoffschneider. 1983. Integration of hepatitis B virus DNA: evidence for integration in the single-stranded gap. *Cell* 34:215-233.
22. Landers, T. A., H. B. Greenberg, and W. S. Robinson. 1977. Structure of hepatitis B Dane particle DNA and nature of the endogenous DNA polymerase reaction. *J. Virol.* 23:368-376.
23. Lien, J. M., C. E. Aldrich, and W. S. Mason. 1986. Evidence that a capped oligonucleotide is the primer for duck hepatitis B

- virus plus-strand DNA synthesis. *J. Virol.* 57:229-239.
24. Marion, P. L., M. J. V. Davelaar, S. S. Knight, F. H. Salazar, G. Garcia, H. Popper, and W. S. Robinson. 1986. Hepatocellular carcinoma in ground squirrels persistently infected with ground squirrel hepatitis virus. *Proc. Natl. Acad. Sci. USA* 83:4543-4546.
25. Mason, W. S., C. E. Aldrich, J. Summers, and J. M. Taylor. 1982. Asymmetric replication of duck hepatitis B virus DNA in liver cells: free minus strand DNA. *Proc. Natl. Acad. Sci. USA* 79:3997-4001.
26. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
27. Meselson, M. S., and C. M. Radding. 1975. A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA* 72:358-361.
28. Miller, R. H., and W. S. Robinson. 1984. Hepatitis B virus DNA forms in nuclear and cytoplasmic fractions of infected human liver. *Virology* 137:390-399.
29. Miller, R. H., and W. S. Robinson. 1986. Common evolutionary origin of hepatitis B virus and retroviruses. *Proc. Natl. Acad. Sci. USA* 83:2531-2535.
30. 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.
31. Molnar-Kimber, K., J. W. Summers, and W. S. Mason. 1984. Mapping of the cohesive overlap of duck hepatitis B virus DNA and of the site of initiation of reverse transcription. *J. Virol.* 51:181-191.
32. Moroy, T., J. Etienne, C. Trepo, P. Tiollais, and M. A. Buendia. 1985. Transcription of woodchuck hepatitis virus in the chronically infected liver. *EMBO J.* 4:1507-1514.
33. Neel, B. G., W. S. Hayward, H. L. Robinson, J. Fang, and S. M. Astrin. 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. *Cell* 23:323-334.
34. Nusse, R., and H. E. Varmus. 1982. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* 31:99-109.
35. Ogston, C. W., G. J. Jonak, C. E. Rogler, S. M. Astrin, and J. Summers. 1982. Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from hepatocellular carcinomas of woodchucks. *Cell* 29:385-394.
36. Ono, Y., H. Onda, R. Sasada, K. Igarashi, Y. Sugino, and K. Nishioka. 1983. The complete nucleotide sequences of the cloned hepatitis B virus DNA: subtype adr and adw. *Nucleic Acids Res.* 11:1747-1757.
37. Overby, L. R., P. P. Hung, J. C.-H. Mao, C. M. Ling, and T. Kakefuda. 1975. Rolling circular DNA associated with Dane particles in hepatitis B virus. *Nature (London)* 255:84-85.
38. Paganiban, A. 1985. Retroviral DNA integration. *Cell* 42:5-6.
39. Pasek, M., T. Goto, W. Gilbert, B. Zink, H. Schaller, P. MacKay, G. Leadbetter, and K. Murray. 1979. Hepatitis B virus genes and their expression in *E. coli*. *Nature (London)* 282:575-579.
40. Payne, G. S., S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, and H. E. Varmus. 1981. Analysis of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for maintenance of the tumor state. *Cell* 23:311-322.
41. Robinson, H. L., and G. C. Gagnon. 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. *J. Virol.* 57:28-36.
42. Robinson, W. S., D. A. Clayton, and R. L. Greenman. 1974. DNA of a human hepatitis B virus candidate. *J. Virol.* 14:384-391.
43. Rogler, C., and J. Summers. 1984. Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from a chronically infected liver. *J. Virol.* 50:832-837.
44. 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.
45. Sattler, F., and W. S. Robinson. 1979. Hepatitis B viral DNA molecules have cohesive ends. *J. Virol.* 32:226-233.
46. Seeger, C., D. Ganem, and H. E. Varmus. 1986. Biochemical and genetic evidence for the hepatitis B virus replication strategy. *Science* 232:477-484.
47. Shih, C., and R. A. Weinberg. 1982. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* 29:161-169.
48. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
49. Summers, J. 1981. Three recently described animal virus models for human hepatitis B virus. *Hepatology* 1:179-183.
50. Summers, J., and W. S. Mason. 1982. Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* 29:403-415.
51. Summers, J., A. O'Connell, and I. Millman. 1975. Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. *Proc. Natl. Acad. Sci. USA* 72:4597-4601.
52. Sureau, C., J. L. Romet-Lemonne, J. I. Mullins, and M. Essex. 1986. Production of hepatitis B virus by a differentiated human hepatoma cell line after transfection with cloned circular HBV DNA. *Cell* 47:37-47.
53. Tiollais, P., C. Pourcel, and A. Dejean. 1985. The hepatitis B virus. *Nature (London)* 317:489-495.
54. Toh, H., H. Hayashida, and M. Takashi. 1983. Sequence homology between retroviral reverse transcriptase and putative polymerase of hepatitis B virus and cauliflower mosaic virus. *Nature (London)* 305:829-831.
55. Tuttleman, J. S., C. Pourcel, and J. Summers. 1986. Formation of the pool of covalently closed circular viral DNA in hepadna virus-infected cells. *Cell* 47:451-460.
56. Valenzuela, P., M. Quiroga, J. Saldivar, P. Gray, and W. J. Rutter. 1981. 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.
57. Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19:259-268.
58. Will, H., C. Kuhn, R. Cattaneo, and H. Schaller. 1982. Structure and function of the hepatitis B virus genome, p. 237-247. In M. Miwa et al. (ed.), *Primary and tertiary structure of nucleic acids and cancer research*. Japan Science Society Press, Tokyo.
59. Will, H., W. Reiser, T. Weimer, E. Pfaff, M. Buscher, R. Sprengel, R. Cattaneo, and H. Schaller. 1987. Replication strategy of human hepatitis B virus. *J. Virol.* 61:904-911.
60. Wolfe, J., S. M. Darling, R. P. Ericksow, I. W. Craig, V. J. Buckle, P. W. Rigby, H. F. Willard, and P. N. Goodfellow. 1985. Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *J. Mol. Biol.* 182:477-485.
61. 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.
62. Yokosuka, O., M. Omata, Y. Z. Zhou, F. Imazeki, and K. Okuda. 1985. Duck hepatitis B virus DNA in liver and serum of Chinese ducks: integration of viral DNA in a hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA* 82:5180-5184.
63. Ziemer, M., P. 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.