Illegitimate Replication of Linear Hepadnavirus DNA through Nonhomologous Recombination

WENGANG YANG AND JESSE SUMMERS*

Department of Cell Biology, The University of New Mexico School of Medicine, Albuquerque, New Mexico 87131

Received 17 February 1995/Accepted 21 March 1995

Linear hepadnavirus DNA in primary hepatocyte cultures efficiently participates in intra- and intermolecular nonhomologous recombination at its ends. The products of this recombination are (i) monomeric covalently closed circular DNAs (cccDNAs) with deletions and insertions around the site of joining and (ii) oligomeric forms in which monomers are joined near the ends in random orientation. A fraction of monomeric cccDNAs can serve as intermediates in further DNA replication through at least five generations of nonhomologous recombination in a process we call illegitimate replication. We suggest that the monomeric and oligomeric linear DNAs produced by illegitimate replication may be precursors of the integrated and other high-molecular-weight hepadnaviral DNA forms seen in chronic infection.

Initiation of hepadnavirus DNA replication begins with the conversion of genomic relaxed circular DNA of the infecting virus to covalently closed circular DNA (cccDNA) in the nucleus of the infected cell (18, 27). cccDNA is transcribed to produce genomic RNA molecules (pregenomes), which are transported to the cytoplasm and translated to produce the two replication proteins, capsid protein and P protein (3, 7). Pregenomes are also packaged into nucleocapsid particles, along with the viral reverse transcriptase (P protein), where they are reverse transcribed to produce new relaxed circular DNA molecules (24). Relaxed circular DNA, synthesized *de novo*, is efficiently converted to new cccDNAs in the nucleus of the infected cell (25, 27, 29) or packaged into enveloped particles and delivered into the nuclei of uninfected cells, where new cccDNAs can be formed, initiating a new cycle of viral DNA replication. In this form of replication, the relaxed circular viral genomes produced are genetically identical to that of the infecting virus. In this paper, we refer to this form of replication as legitimate DNA replication.

The formation of relaxed circular DNA depends on a specific step in the initiation of second (or plus)-strand synthesis, shown in Fig. 1. The RNA primer for plus-strand synthesis is derived from the terminal 18 ribonucleotides of the 5' end of the pregenome through cleavage by the RNase H activity of the viral reverse transcriptase (12, 13, 15, 23). Because of the precise cleavage of the primer, it contains a copy of the 12-nucleotide sequence dr1 located at its 3' end. The presence of this sequence allows the primer to be translocated to a site near the 5' end of the first strand (minus strand), where it duplexes with the minus strand through a complementary 12-nucleotide sequence, dr2. The translocated primer initiates plus-strand synthesis, copying the minus strand to near its 5' end. Further elongation of the plus strand depends on a template switch to the single-stranded 3' end of the minus strand, vacated by the translocated primer. The plus-strand template switch is mediated by a nine-nucleotide terminal repeat (r) in the minus strand and results in circularization of the DNA, reproducing a faithful copy of the infecting viral DNA.

Mutations that lie just 3' of dr1, reduce the homology be-

tween dr1 and dr2, or change the spacing between dr1 and the 5' cap of the pregenome prevent the primer translocation step (15, 23). However, such mutations do not necessarily prevent plus-strand initiation and elongation, which then occur at the original site of primer binding at the 3' end of the minus strand. This process has been termed *in situ* priming (23). Spontaneous *in situ* priming occurs during replication of wild-type hepadnaviruses at a frequency of about 5%. The resulting product is a linear double-stranded molecule with the plus-strand RNA primer presumably present in a 18-nucleotide RNA-DNA duplex at the 3' end of the minus strand, represented at the bottom of Fig. 1. Linear DNAs contain the nine-nucleotide repeat (r) at either end of the minus strand. Linear double-stranded molecules have been found as a minor species in both the cytoplasm and nucleus of infected cells, but their biological activity has not been extensively investigated.

In this study, we report evidence that linear double-stranded DNAs produced by *in situ* priming of mutant or wild-type genomes efficiently undergo nonhomologous recombination in the nucleus of infected cells at sites near the ends of the linear molecules to produce a variety of monomeric and oligomeric products. Our evidence indicates that some of the resulting monomeric cccDNA molecules can produce pregenomes that can be reverse transcribed to generate more linear DNAs. Multiple rounds of synthesis of linear DNAs occur in this way to produce progeny cccDNA molecules with nucleotide sequences at the region of joining that are different from those of the parental cccDNA and linear DNA molecules. We have termed this process illegitimate DNA replication. Illegitimate replication is slow compared with legitimate replication, but may be self-sustaining in individual cells. Moreover, the efficiency of intra- and intermolecular nonhomologous recombination of linear DNAs suggests that these species may be the precursors of integrated hepadnavirus DNA.

MATERIALS AND METHODS

Plasmids. Infectious plasmids containing duck hepatitis B virus type 3 (DHBV3) DNA (22) with mutations resulting in defects in plus-strand primer translocation were a generous gift of Dan Loeb. The mutations in these plasmids are shown in Fig. 2. The pregenome-encoding sequences from these plasmids were subcloned into an expression vector, pUC119CMVDHBV. A mutation ablating production of the large envelope protein (1165A) was introduced (26) by subcloning a subgenomic *Bg/III-KpnI* fragment (nucleotides 391 to 1294) from the DHBV16 strain (16). Dimer genome-containing plasmids were constructed

* Corresponding author. Phone: (505) 277-7979. Fax: (505) 277-9494.

FIG. 1. Generation of relaxed circular and linear double-stranded DNA during hepatitis B virus replication. See text for details. Thin lines represent DNA strands, and thick lines represent RNA. The positions of *dr1*, *dr2*, and *r* are indicated.

by substituting the *AflII*-*NsiI* (nucleotides 2526 to 2849) fragment, containing joint regions altered in vivo by nonhomologous recombination between the ends of linear DNA into an *EcoRI* monomer of DHBV16 cloned into vector pSP65, excising the resulting monomer, and recloning as a dimer into the same vector, as previously described (25). The helper plasmid pSPADR1, in which DNA synthesis was prevented by a precise deletion of *dr1*, has been described previously (6). The mutant genome *NheI* (see Fig. 5) was constructed by cloning a synthetic linker, TTAACAAAGCTAGCTTTTG, into the 3' *AflII* site in pUC119CMVDHBV. The resulting pregenome carried the linker at its 3' end. The linker does not interfere with relaxed circular and cccDNA synthesis, but transcription of the resulting cccDNA produced a pregenome with the linker at both ends, and reverse transcription of this pregenome produced only linear double-stranded DNA (data not shown).

Transfections. Transfection of the chicken hepatoma cell line LMH (4, 9), isolation and assay of replicative intermediates by Southern blot hybridization (26), and concentration and assay of enveloped virus by the pronase-DNase I method were performed as previously described (11).

Primary hepatocytes. Duck primary hepatocyte cultures were prepared and maintained as previously described (20). The virus used to infect 2- to 3-day-old cultures was concentrated from the supernatants of transfected LMH cells by precipitation with 10% polyethylene glycol 8000 (26), adjusted to 10 mM magnesium acetate, and treated for 30 min with 100 μ g of DNase I (type II; Sigma) per ml at 37°C. Concentrated virus was added to the culture supernatants and allowed to infect the cells overnight before the medium was changed. The selective isolation of viral replicative intermediates and cccDNA by the protein-detergent precipitation method was carried out as previously described (25). Total nuclear DNA for analysis of high-molecular-weight viral DNA (see Fig. 8) was isolated by lysis of the infected cells with 1% Nonidet P-40. Nuclei and insoluble cell debris were pelleted by microcentrifuge centrifugation, and the pellet was dissolved in 1% sodium dodecyl sulfate (SDS) and digested with 500 μ g of pronase per ml. Nucleic acids were recovered by phenol extraction and ethanol precipitation.

Gel purification of cccDNA and amplification of altered and wild-type joints. DNA from the cccDNA fraction extracted from infected hepatocytes was mixed

with 1 μ g of plasmid pSP65, which comigrated with DHBV cccDNA, and subjected to electrophoresis through a 1% low-melting-point agarose gel (Sea-Plaque). The ethidium bromide-staining band was excised and extracted from the gel slice by melting at 68°C, phenol extraction, and ethanol precipitation. A portion of the DNA was amplified through 25 cycles of PCR with the Invitrogen PCR kit according to the manufacturer's instructions and with primers whose sequence corresponded to DHBV sequences 70 to 45 and 2217 to 2240. The reaction product was diluted 10-fold in fresh PCR reaction mix and amplified through two more cycles to eliminate heteroduplexes that might form in the final unproductive cycles of amplification. The reaction products were ligated to the TA cloning vector pCRII (Stratagene) and transformed into *Escherichia coli* DH5 α . Colonies were selected for kanamycin resistance and screened for inserts. For construction of libraries, a portion of the transformed cell preparation was grown directly in liquid culture under kanamycin selection. Plasmid DNAs were extracted and purified by standard methods.

Sequencing. All sequencing was carried out on plasmid DNA with the Sequenase version 2.0 kit (US Biochemicals) according to the manufacturer's instructions.

RESULTS

Replication of wild-type and primer translocation mutant viruses in primary hepatocytes. We infected primary duck hepatocyte cultures with a series of mutant DHBV genomes that were defective to various degrees in plus-strand primer translocation. The properties of these mutants, including production of various amounts of double-stranded linear DNA at the expense of relaxed circular DNA, were described previously (15, 23). The mutations producing these defects, shown in Table 1, are located within or near *dr1*. These mutant genomes were further altered to eliminate production of the large envelope protein p36 (1165A mutation) in order to stimulate the accumulation of nuclear viral DNA (26). Infection-

Deletions	(46 joints)
66 parent	gaacttAAGAATTACacgcggttact
2-11	gaacttAAGAATTAC ggcggttact
66 parent	gaacttAAGAATTACacgcggttact
2-13	gaacttAAGAATTAC t
66 parent	gaacttAAGAATTACacgcggttact
2-18	gaac ttact
Insertions	(12 joints)
66 parent	gaacttAAGAATTAC acgcggttact
1-17	gaacttAAGAATTACATacgcggttact
wt parent	gaacttAAGAATTAC acccctctact
wt-5	gaacttAAGAATTACGAATTACaccctctact
66 parent	gaacttAAGAA TTAC acgcggttact
1-18	gaacttAAGAAGAAATTACTATacgcggttact
66 parent	gaacttAAGAATTAC acgcggttact
1-21	gaacttAAGAATTACTTACGAAGAATTACTTACTACTATacgcggttact
Deletions and insertions	(8 joints)
66 parent	gaacttAAGAA TTACacgcggttact
1-13	gaacttAAGAAGAAATTAC ggcggttact
66 parent	gaacttAAGAA TTACacgcggttact
1-4	gaacttAAGAAGAAATTACAAGAAGAAATTAC ccggttact
12 parent	gaacttAA GAATTACaccctctact
12-9	gaa GAGAAGAAATT Caccctctact

FIG. 2. Sequence alignments at 66 recombination joints. Sequences are expressed as the plus strand. Uppercase letters indicate sequences derived from *r*. Since the parental 66 plasmid produced pregenomes with mutations only at the 5' end, we show the sequence at the initiation site for minus-strand synthesis as the wild-type TTAC instead of the mutant TTAT.

TABLE 1. Mutations affecting plus-strand primer translocation

Genome	Plus-strand sequence ^a	Location in pregenome
Wild type	<u>TACACCCCTCTCCTTCGG</u>	5' and 3' ends
DR1-13	-----G-----	5' and 3' ends
DR1-12	-----A-----	5' and 3' ends
DR1-66	--T--G--GT--A-----	5' end only
DR1-Pvu	-----C	5' and 3' ends

^a The dr1 sequence is underlined.

competent viral particles were produced by cotransfection of the permissive chicken hepatoma cell line LMH with expression plasmids for these mutant genomes and a second plasmid, ΔDR1, that supplied the missing envelope proteins. Equivalent amounts of virus particles (by total DNA content) were used to infect primary hepatocyte cultures.

As shown in the Southern blot in Fig. 3, the accumulation of DNA replicative intermediates and cccDNA in infected hepatocyte cultures was inhibited in accordance with the strength of the primer translocation defect. In addition, as previously reported (15, 23), the pattern of the replicative intermediates reflected the strength of the primer translocation defect, with the weakest defect, in mutant 13, resulting in an approximately 1:1 ratio of linear to circular DNA, and the strongest defect, Pvu, producing a corresponding ratio of >20:1 (not visible in the autoradiogram).

Comparison of cccDNA formed by mutant and wild-type viruses. The relative rates of replication of the primer translocation mutants suggested that replication might be entirely dependent on the synthesis of relaxed circular DNA. Since replication in primary hepatocytes depends on cccDNA amplification (11, 26, 27), it seemed likely that cccDNA formed by the mutants might be derived exclusively from the reduced amounts of relaxed circular DNA that they produced, thus explaining the relationship between their relative rates of replication and the amounts of relaxed circular DNA made. However, it was also possible that linear DNA could also be used for cccDNA synthesis but that cccDNA produced from linear DNA was not functional for replication.

FIG. 3. Replication of primer translocation mutants in primary hepatocytes. Primary hepatocyte cultures were infected with the indicated mutant viruses, and at 5, 8, 11, and 17 days postinfection, replicative intermediates and cccDNA were extracted. DNAs from one-fourth of a 60-mm plate were assayed by electrophoresis through a 1.5% agarose gel, blotting, and hybridization with a riboprobe specific for detection of the minus strand. Hybridization molecular weight standards (lanes M) consisted of 0.5 ng of pSPDHBV5.1 (2X) digested with *Bam*HI to produce 4.6-, 3.0-, and 1.4-kbp fragments. The standard shown was 0.5 ng of plasmid DNA. The migration positions of relaxed circular (rc), linear (l), single-stranded (ss), and covalently closed circular (ccc) DNAs are indicated.

FIG. 4. Sequencing DNA from a library of mutant and wild-type (wt) recombination joints. Sequence ladders were extended in the direction of dr1 from either upstream (primer 1, nucleotides 2399 to 2419) or downstream (primer 2, 2662 to 2641) of dr1. Sequence divergence around dr1 resulted in the generation of unreadable overlapping sequence ladders.

In order to determine if linear DNA molecules were converted to cccDNAs, we sequenced the regions of these cccDNA molecules that would correspond to the ends of linear double-stranded DNA. We speculated that ligation of linear DNA molecules to produce cccDNAs might result in ligation joints different from those produced from relaxed circular DNA. The region of the putative joints was amplified by PCR from gel-purified cccDNA. Control reactions showed that relaxed circular DNA molecules were poorly amplified, if at all, by the conditions that we used, as previously reported (10), presumably because of the lack of continuity of either strand through the region bracketed by the primers. PCR products were diluted and reamplified through two cycles to eliminate heteroduplexes that might be formed by the annealing of two related but nonidentical strands. These products were cloned and sequenced.

Initially we sequenced a pooled library of plasmids derived from approximately 200 clones of PCR-amplified joints from cccDNA purified from cells infected by wild-type virus and from more than 400 clones derived from the cccDNA of mutant 1165/66-infected cells. Two sequencing primers that bracketed the putative joints were used and generated sequence ladders in the direction of the joint of either plus-strand (primer 1) or minus-strand (primer 2) polarity. As shown in Fig. 4, while sequence ladders could be read through the regions upstream and downstream of dr1, sequence divergence in the library was detected around the position of dr1 with both primers. The result indicated that the library of putative joint regions contained diverse sequences between and around nucleotide 2537, which corresponds to the 5' end of the minus strand of linear DNA, and nucleotide 2547, which corresponds to the 5' end of the plus strand of in situ-primed linear DNA.

To examine the structures of individual joints, we sequenced clones derived from the cccDNA of wild-type- and mutant-infected cells. Sequencing of 71 clones of DNA amplified from the cccDNA purified from cells infected with the 1165/66 mutant revealed that 69 putative joint regions contained sequence alterations, with a total of 54 different sequences. Similarly, 12 of 15 joints amplified from cccDNA of the 1165/12 mutant contained sequence alterations, with 11 different sequences. In contrast, only 1 of 48 clones sequenced from the corresponding DNA of wild-type-infected cells showed a sequence alteration

FIG. 5. Efficiency of cccDNA formation during initiation of infection by virions containing relaxed circular (rc) or linear (l) DNAs. Primary hepatocytes were infected with virus particles concentrated from the culture fluids of transfected LMH cells. Hepatocytes were harvested at 2 and 4 days postinfection and assayed for cccDNA by selective extraction and Southern blot hybridization. DNA in enveloped particles used for the infection is shown in the upper panel, and cccDNAs are shown in the lower two panels. The upper panel was exposed for 2 h, and the lower panels were exposed for 44 h.

(wt-5, Fig. 2), with 47 having perfect alignment of DNA sequences with the original plasmid DNA, indicating that these cccDNAs were derived from relaxed circular DNA.

Of a total of 66 different altered joints, 46 suffered deletions that included one or both regions corresponding to the expected ends of in situ-primed linear DNA. Twelve joints contained an insertion of short sequences or complex repeats of sequences derived from the repeat sequence AAGAATTAC, located at either end of linear DNA. Eight joints suffered deletions accompanied by such insertions. Examples of the variety of joints that we observed are shown in Fig. 2.

Efficiency of cccDNA formation from linear DNA. Since altered joints were a characteristic of cccDNA found in mutants synthesizing predominantly linear double-stranded DNA, it seemed reasonable to conclude that the cccDNAs were generated by nonhomologous recombination near the ends of linear DNA precursors. The high frequency of altered joints suggested that formation of cccDNA from linear DNA occurred readily in the infected cells. In order to directly test the relative efficiencies of cccDNA formation from linear and relaxed circular DNAs, we infected primary duck hepatocytes with equivalent amounts of viral particles with various ratios of linear and relaxed circular double-stranded viral DNAs. De novo synthesis of cccDNA was prevented in the primary hepatocytes by the presence of 100 μ g of phosphonoformic acid per ml in the culture medium (19), so that any cccDNA produced after infection was derived from the infecting virus. At 2 and 4 days postinfection, cccDNA was isolated from the infected hepatocytes and assayed by Southern blot (Fig. 5).

Although the ratios of circular to linear double-stranded viral DNAs in the inocula varied from about 20:1 to about 1:20, the amounts of cccDNA formed after infection of the primary hepatocytes were equivalent for all the inocula at 2 and 4 days postinfection. This result indicated that during the initiation of infection, linear DNA and relaxed circular DNA could be converted to cccDNA with similar efficiencies. We concluded that a defect in the formation of cccDNA from linear double-stranded DNA was probably not responsible for the low replication rate of the primer translocation mutants. On the other hand, synthesis of cccDNA functional for further replication might occur less frequently from linear DNA than from relaxed circular DNA because of deletions and insertions occurring as a consequence of nonhomologous recombination.

Infection of primary hepatocytes can be initiated by linear-DNA-containing virus. The preceding experiments suggested that linear DNA might serve as an alternative to relaxed circular DNA as a substrate for cccDNA amplification, provided that some molecules of cccDNA formed by nonhomologous recombination were functional for further DNA synthesis. We confirmed this hypothesis by demonstrating that linear-DNA-containing virus, derived from cccDNAs with altered joints, could initiate an infection, resulting in the further production of progeny linear DNAs.

We subcloned two altered joint regions derived from either wild-type- or mutant-infected hepatocytes (wt-5 and 66-1-18, Fig. 2) into dimer viral DNA-containing plasmids and transfected these plasmids into LMH cells. The plasmids containing these joints produced nearly wild-type levels of DNA-containing virus particles in the transfections (not shown). The virus particles, however, contained linear DNA rather than relaxed circular DNA. We infected primary duck hepatocytes with the linear-DNA-containing viruses produced by transfection with these two plasmids and assayed for replication in the hepatocytes at 6, 10, and 14 days postinfection. As shown in Fig. 6, both viruses were able to initiate an infection in primary hepatocytes that resulted primarily in the production of linear DNA replicative intermediates and enveloped linear DNA-containing viruses. In addition, as illustrated by the presence of about 20% relaxed circular DNA in hepatocytes infected by the wt-5 virus particles, nonhomologous recombination occurring during cccDNA formation could restore wild-type or nearly wild-type joints, giving rise to the synthesis of relaxed circular DNA. The results directly demonstrated two cycles of synthesis of functional cccDNA by nonhomologous recombination, i.e., that which generated the wt-5 and 66-1-18 joints, and that which resulted in productive infection by linear-DNA-containing virus in the primary hepatocytes.

Evidence for multiple cycles of linear DNA replication. Evidence for multiple cycles of linear DNA replication within individual cells could be seen in the sequences found in 19 joints that could not have been generated by single non-homologous recombinations between the ends of linear DNA. The occurrence of multiple copies of sequences derived from the r region, AAGAATTAC, within these joints indicated se-

FIG. 6. Infection of primary hepatocytes by virus particles containing mutant or wild-type linear DNAs. Primary hepatocytes were infected with virus particles concentrated from the culture fluid of LMH cells transfected with the dimer-containing plasmids with the normal joint regions replaced by the indicated altered joints. At 6, 10, and 14 days, cccDNA and replicative intermediates were extracted and analyzed by Southern blot hybridization (cells). DNA extracted from enveloped virus particles (virus) present in the culture fluids at 14 days postinfection was assayed as previously described (11). The upper portion of the blot of replicative intermediates was removed to eliminate the interfering hybridization signal from a contaminating plasmid DNA that can be seen in the virus lanes. rc, relaxed circular DNA; l, linear DNA; ss, single-stranded DNA.

Generation 1 (deletion)

linear ...gaacttAAGAATTAC
 AAGAATTATacgcgttact...
 ccc ...gaacttAAGAATTACcgcgttact...

Generation 2 (partial duplication)

linear ...gaacttAAGAATTAC
 AAGAATTACcgcgttact...
 ccc ...gaacttAAGAAGAATTACcgcgttact...

Generation 3 (reduplication)

linear ...gaacttAAGAAGAATTAC
 AAGAAGAATTACcgcgttact...
 ccc ...gaacttAAGAAGAATTACAAGAAGAATTACcgcgttact...

FIG. 7. Derivation of the sequence of the 66-1-4 joint. Successive nonhomologous recombinations of the ends of linear DNAs are shown, expressed as the plus strand. The end containing the 5' terminus of the minus strand is shown on the top, and the end containing the 3' terminus of the minus strand is shown on the bottom. Arrows indicate the sites of joining. Uppercase letters designate nucleotides derived from the original r sequence AAGAATTAC.

quential replication and recombination events. An example of the evolution of one of these complex joint sequences by three cycles of replication and nonhomologous recombination is shown in Fig. 7. All 19 complex joints that we observed could be explained by this process, with as many as five generations of replication being required to produce the most complex joints (e.g., 66-1-21, Fig. 2). The data imply, therefore, that linear DNA replication was sustained through at least five generations in some infected cells during the 17 days in culture.

Intermolecular nonhomologous recombination of linear DNAs. The production of linear viral DNAs in the mutant-infected hepatocyte cultures also resulted in the formation of oligomeric viral DNA through intermolecular nonhomologous recombination. Inspection of the replicative intermediate fractions (which contain the bulk of cellular DNA) in Fig. 3 revealed the presence of higher-molecular-weight viral DNA species in cells producing high levels of linear viral DNA. These species, present as a ladder of discrete sizes, are shown in greater detail in the total nuclear DNA extracted at 17 days postinfection from 1165/12-infected cells in Fig. 8. Total nuclear DNA extracted from wild-type-infected hepatocytes did not contain detectable high-molecular-weight viral DNA.

In order to investigate the nature of these molecules, we digested the total nuclear DNAs with a variety of restriction enzymes. Enzymes that did not cut DHBV DNA but did cleave cellular or plasmid DNAs (*Pst*I, *Nde*I, and *Sal*I) did not affect the mobility of the high-molecular-weight viral DNAs, suggesting that they consisted predominantly of viral sequences. Digestions with enzymes that cut viral DNA (*Bam*HI plus *Nco*I, *Eco*RI, and *Nco*I) generated novel fragments (Fig. 8, arrowheads) in the DNA extracted from 1165/12-infected cells but not in the DNA from wild-type virus-infected cells. The sizes of these fragments were consistent with those expected if the high-molecular-weight DNAs contained joints derived from joining the ends of linear DNAs in head-to-head and tail-to-tail orientations, as shown at the bottom of Fig. 8.

In order to determine whether the high-molecular-weight

forms also contained head-to-tail joints, we removed the monomeric viral DNA forms (relaxed circular, linear, and cccDNAs) by gel purification and digested the purified high-molecular-weight forms. Cleavage with *Bgl*II, *Eco*RI, *Nco*I, or *Bam*HI plus *Nco*I produced two sizes of hybridizing DNA fragments characteristic of head-to-tail and either head-to-head (*Eco*RI, 5.1 kbp; *Bgl*II, 4.3 kbp) or tail-to-tail (*Nco*I, 5.7 kbp; *Bam*HI plus *Nco*I, 3.9 kbp) joints. Comparison of the hybridization signals attributable to head-to-head, tail-to-tail, and the two possible types of head-to-tail joints indicated that all orientations occurred with similar frequencies in the high-molecular-weight DNA.

In a separate experiment (not shown), we confirmed that two of the fragments diagnostic of head-to-head (*Eco*RI, 5.1 kbp) and tail-to-tail (*Nco*I, 5.7 kbp) arrangements consisted of inverted duplications around the proposed sites of joining. We showed that the gel-purified fragments produced snap-back molecules of the expected sizes (2.6 and 2.8 kbp) after denaturation. The snap-back products were shown to be double stranded by resistance to digestion with mung bean nuclease and sensitivity to cleavage with *Bgl*II. We concluded that the high-molecular-weight viral DNA forms were produced by recombinational joining near the ends of linear viral DNAs, regardless of the polarity of the recombining species. Such joining independent of the polarity of the viral sequences is consistent with the nonhomologous recombination that we inferred from the previous experiments.

DISCUSSION

Illegitimate replication of linear DNA. We have presented evidence that linear DHBV DNA entering the nucleus of hepatocytes frequently participates in both inter- and intramolecular recombination. Intramolecular recombination resulted in the production of cccDNA molecules, some of which participated in further DNA replication through the formation of

FIG. 8. Analysis of high-molecular-weight DNAs produced by 1165/12 virus in primary hepatocytes. Total nuclear DNA was isolated from hepatocytes infected with wild-type or 1165/12 viruses produced in LMH cells. (Top) Undigested total DNA (first two lanes), total DNA (middle lanes), and gel-purified DNA (right lanes) digested with the indicated restriction enzymes were analyzed by electrophoresis through a 1.5% agarose gel and Southern blotting. The hybridization marker lane contained 0.5 ng of digested plasmid DNA, as in Fig. 3. (Bottom) The expected diagnostic fragments for head-to-head and tail-to-tail joining are indicated and are marked by arrowheads in the Southern blot. Restriction sites: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; N, *Nco*I.

FIG. 9. Illegitimate replication of linear DNA. Linear DNA introduced by a virus or produced in an infected cell undergoes nonhomologous recombination at the ends (a) to produce a cccDNA molecule with an altered joint (b) containing a partial repeat of the original r sequences. Transcription of the cccDNA and reverse transcription of the pregenome produces new linear DNAs with altered r sequences (c). Nonhomologous recombination at the ends of these linear DNAs (d) produces new cccDNAs with a variety of joints, some of which are functional (e) and some of which are nonfunctional (X) for further DNA synthesis (f). Functional cccDNAs are transcribed to produce pregenomes which are reverse transcribed to produce new linear DNAs with further altered r sequences (g). In wild-type DNAs, alterations in the cccDNA joints restore the ability to produce relaxed circular DNA and legitimate replication (h). Brackets indicate hypothetical nonhomologous recombination.

double-stranded linear DNA intermediates that gave rise to progeny cccDNA molecules through nonhomologous recombination. From sequence analysis of the recombination joints of cccDNA molecules produced from linear double-stranded DNA, we infer that as many as five generations of cccDNAs could be formed by nonhomologous recombination at the ends of linear DNAs. Additional generations due to homologous recombination at the ends of linear DNAs would not have been detected by sequence analysis. Since each cycle of replication mediated through linear DNA intermediates would result in the formation of novel progeny DNA molecules that are not identical to the parental DNA, we have termed this process illegitimate replication. A schematic representation of illegitimate replication is shown in Fig. 9.

We would like to point out that nonhomologous recombination at the ends of wild-type linear DNA favors the formation of cccDNA molecules that produce linear rather than relaxed circular DNA progeny. Deletion or insertion of sequences at the recombination joint will cause a corresponding

change in the 5' end of the pregenome. When such a change involves a deletion of part of dr1 or alters the distance between the 5' end of the pregenome and dr1, a defect in primer translocation (15) will be produced. Therefore, hepatocytes infected by viruses containing wild-type linear DNAs are likely to initiate cycles of illegitimate replication.

We do not know the mechanism of nonhomologous recombination that joins the ends of linear DNA. Recombination appears to depend on a cellular function, since we found that, unlike primary hepatocytes, LMH cells did not efficiently convert linear viral DNA to cccDNA (data not presented). Thirty-eight different sites for end-to-end joining in a single cycle of recombination are summarized in Fig. 10. About 60% of the recombination joints (27 of 40) resulted in the deletion of sequences that were required for (i) transcription of the pregenome (8, 14) or (ii) minus-strand synthesis (1, 2, 5, 28), and would not be expected to contribute further to viral DNA synthesis. About 10% of the recombination joints (four in the shaded squares) occurred within the repeat sequence r at both ends of the linear DNA, such that no deletion of r sequences occurred (see below).

Generation and replication of r sequence sets by illegitimate replication. Illegitimate replication differs from conventional modes of DNA replication in two related aspects: (i) the progeny DNAs are not necessarily identical to the parental DNAs, and (ii) a majority of the progeny molecules cannot function further in DNA replication. These properties do not, however, preclude that illegitimate replication can sustain the existence of the hepadnavirus genome over many generations. Examination of the matrix of possible recombination joints in Fig. 10 reveals that 55 of the possible joints (shaded squares) formed by recombination within the nine-nucleotide r region, including 10 joints that regenerate the original r sequence, will result in duplications within the r region of zero to 9 nucleotides, generating r sequences of 9 to 18 nucleotides in the second-generation linear DNA. Each of these 55 possible linear DNAs would be capable of regenerating the original r sequence by

FIG. 10. Different recombination joints resulting from single recombination events at the ends of linear DNA. Sequences are expressed as the plus strand. The sequence at the top of the matrix corresponds to the end of an in situ-primed linear DNA containing the 3' end of the minus strand. The sequence at the left corresponds to the end containing the 5' end of the minus strand. A marked square in the matrix indicates a recombination joint in which the nucleotide in the column was joined to the nucleotide in the row. Three recombination joints outside the matrix were omitted.

recombination in the second generation. Therefore, each member of the set of these 55 possible linear DNAs is capable of generating all members of the set in one or two subsequent generations. In addition, a similar relationship exists between any member of this set of 55 joints and the corresponding subset of its progeny resulting from a second generation of cccDNA synthesis from linear DNA.

Many possible recombinations generate cccDNA molecules that cannot produce reversion to the original *r* sequence (outside the shaded boxes, Fig. 10). Such recombination events do not affect the lineage of the parental *r* sequence, since DNA replication by hepadnaviruses is conservative and new joints are created without loss of the parental DNA. In addition, many cccDNA molecules with joints that cannot give rise to the original parental *r* sequence can nevertheless participate in further DNA replication, and these joints will produce lineages of sets and subsets of *r* sequences that are revertible to the parental sequence of the set.

Thus, the illegitimate and legitimate modes of replication differ significantly in their mechanisms for maintaining an *r* sequence over time. Illegitimate replication replicates sets of interconvertible sequences through multiple generations, while legitimate replication results in the precise replication of an individual *r* sequence in each generation.

Variation in rates of virus replication by legitimate and illegitimate replication. A second significant difference between legitimate and illegitimate replication is the level of virus replication resulting from each process. Legitimate replication reproduces an *r* sequence that is optimal for high rates of DNA replication (through relaxed circular DNA), while illegitimate replication reproduces *r* sequence sets that contain cccDNA joints with either low rates of DNA replication (exclusively through linear DNA) or high rates of DNA replication (through complete or partial reversion to legitimate replication). Thus, cells infected by viruses containing relaxed circular DNA will produce homogeneously high levels of virus because all of their cccDNA molecules participate in DNA synthesis. In contrast, cells infected with viruses containing linear DNA will be heterogeneous with respect to levels of virus production, according to the fraction of the cccDNA pool that participates in DNA synthesis. Consistent with this prediction, we have observed a characteristic broad range of capsid antigen immunostaining in primary hepatocytes infected with linear-DNA-containing viruses (not shown).

In the liver, the low replication potential of illegitimate replication may offer some advantage to infected cells if hepatocytes with low levels of virus expression are better able to survive an antigen-specific cellular immune response than hepatocytes with high levels of expression. Chronic hepatitis in humans and woodchucks is known to be associated with the progressive appearance and accumulation of hepatocytes that display low or undetectable levels of virus antigen expression or abnormal patterns of expression. We suggest that pleomorphic virus expression in infected livers may be correlated with the presence of hepatocytes that express defective cccDNA molecules formed by nonhomologous recombination. Whether cells that carry out predominantly illegitimate replication occur *in vivo*, however, remains to be determined.

Novel forms and integrated viral DNA. We previously reported the cloning of multimeric viral DNA sequences from DNA extracted from liver chronically infected with the woodchuck hepatitis virus, a member of the hepadnavirus family that is closely related to the human hepatitis B virus (21). These sequences were remarkable in the occurrence of multiple direct and inverted copies of partial and complete viral DNA segments, with no evidence for the presence of cellular

DNA sequences. Similar high-molecular-weight forms of viral DNA were cloned from total DNA extracted from livers chronically infected with the ground squirrel hepatitis virus (17). The presence of these multimeric forms of mammalian hepadnavirus DNAs suggests that high levels of linear DNA synthesis occur in some cells during chronic infections by mammalian hepadnaviruses. Randomly oriented multimeric viral DNA in DHBV-infected hepatocytes was a hallmark of import of large amounts of linear DNA into the nucleus in the experiments reported here. Analogous multimeric molecules in mammalian hepadnavirus-infected liver seem likely precursors of the high-molecular-weight viral DNA "novel forms" cloned from chronically infected woodchuck or ground squirrel livers.

The recombination properties of linear hepadnavirus DNA demonstrated in this study suggest to us that monomeric or oligomeric linear DNA molecules may occasionally participate in nonhomologous recombination with cellular DNA. Such rare events could be responsible for producing the various forms of integrated hepadnaviral DNA that have been found regularly in hepatomas associated with mammalian hepadnavirus infections.

ACKNOWLEDGMENTS

We are grateful to Dan Loeb for generously providing the primer translocation mutants, to Dan Loeb and William Mason for helpful discussions, and to Timothy Powell for excellent technical assistance.

This work was supported by Public Health Service grant CA-42542 from the National Cancer Institute.

REFERENCES

1. Bartenschlager, R., and H. Schaller. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* **11**:3413-3420.
2. Calvert, J., and J. Summers. 1994. Two regions of an avian hepadnavirus RNA pregenome are required in *cis* for encapsidation. *J. Virol.* **68**:2084-2090.
3. Chang, L.-J., D. Ganem, and H. Varmus. 1990. Mechanism of translation of the hepadnaviral polymerase (P) gene. *Proc. Natl. Acad. Sci. USA* **87**:5158-5162.
4. Condreay, L., C. Aldrich, L. Coates, W. Mason, and T.-T. Wu. 1990. Efficient duck hepatitis B virus production by an avian tumor cell line. *J. Virol.* **64**:3249-3258.
5. Hirsch, R. C., D. D. Loeb, J. R. Pollack, and D. Ganem. 1991. *cis*-Acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. *J. Virol.* **65**:3309-3316.
6. Horwich, A. L., K. Furtak, J. C. Pugh, and J. Summers. 1990. Synthesis of hepadnavirus particles containing replication-defective duck hepatitis B virus genomes in cultured Huh-7 cells. *J. Virol.* **64**:642-650.
7. Huang, M., and J. Summers. 1991. Infection initiated by the RNA pregenome of a DNA virus. *J. Virol.* **65**:5435-5439.
8. Huang, M., and J. Summers. 1994. *pet*, a small sequence distal to the pregenome cap site, is required for transcription of the duck hepatitis B virus pregenome. *J. Virol.* **68**:1564-1572.
9. Kawaguchi, T., K. Nomura, Y. Hirayama, and T. Kitagawa. 1987. Establishment and characterization of a chicken hepatocellular carcinoma cell line LMH. *Cancer Res.* **47**:4460-4464.
10. Koeck, J., and H.-J. Schlicht. 1993. Analysis of the earliest steps of hepadnaviral replication: genome repair after infectious entry into hepatocytes does not depend on viral polymerase activity. *J. Virol.* **67**:4867-4874.
11. Lenhoff, R., and J. Summers. 1994. Construction of avian hepadnavirus variants with enhanced replication and cytopathicity in primary hepatocytes. *J. Virol.* **68**:5706-5713.
12. Lien, J. M., C. 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**:229-236.
13. Lien, J. M., D. Petcu, C. Aldrich, and W. S. Mason. 1987. Initiation and termination of duck hepatitis B virus DNA synthesis during virus maturation. *J. Virol.* **61**:3832-3840.
14. Liu, C., W. S. Mason, and J. B. Burch. 1994. Identification of factor-binding sites in the duck hepatitis B virus enhancer and *in vivo*: effects of enhancer mutations. *J. Virol.* **68**:2286-2296.
15. Loeb, D., R. Hirsch, and D. Ganem. 1991. Sequence-independent RNA cleavages generate the primers for plus strand DNA synthesis in hepatitis B viruses: implications for other reverse transcribing elements. *EMBO J.* **10**:3533-3540.
16. Mandart, E., A. Kay, and F. Galibert. 1984. Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison with woodchuck and human

- hepatitis B virus sequences. *J. Virol.* **49**:782-792.
17. **Marion, P. L., W. S. Robinson, C. E. Rogler, and J. Summers.** 1982. High molecular weight GSHV-specific DNA in chronically infected ground squirrel liver. *J. Cell. Biochem. Suppl.* **6**:203.
 18. **Mason, W., M. Halpern, J. England, G. Seal, J. England, L. Coates, C. Aldrich, and J. Summers.** 1983. Experimental transmission of duck hepatitis B virus. *Virology* **131**:375-384.
 19. **Mason, W., J.-M. Lien, D. Petcu, L. Coates, W. T. London, A. O'Connell, C. Aldrich, and R. Custer.** 1987. *In vivo* and *in vitro* studies on duck hepatitis B virus replication, p. 3-16. *In* W. Robinson, K. Koike, and H. Will (ed.), *Hepadnaviruses*. Alan R. Liss, Inc., New York.
 20. **Pugh, J., and J. Summers.** 1989. Infection and uptake of duck hepatitis B virus by duck hepatocytes maintained in the presence of dimethyl sulfoxide. *Virology* **172**:564-572.
 21. **Rogler, C. E., and J. Summers.** 1982. Novel forms of woodchuck hepatitis virus DNA isolated from chronically infected woodchuck liver nuclei. *J. Virol.* **44**:852-863.
 22. **Sprenkel, R., C. Kuhn, H. Will, and H. Schaller.** 1985. Comparative sequence analysis of duck and human hepatitis B virus genomes. *J. Med. Virol.* **15**:323-333.
 23. **Staprans, S., D. Loeb, and D. Ganem.** 1991. Mutations affecting hepadnavirus plus-strand synthesis dissociate primer cleavage from translocation and reveal the origin of linear viral DNA. *J. Virol.* **65**:1255-1262.
 24. **Summers, J., and W. Mason.** 1982. Replication of the genome of hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**:403-415.
 25. **Summers, J., P. Smith, and A. Horwich.** 1990. Hepadnaviral envelope proteins regulate amplification of covalently closed circular DNA. *J. Virol.* **64**:2819-2824.
 26. **Summers, J., P. Smith, M. Huang, and M.-S. Yu.** 1991. Morphogenetic and regulatory effects of mutations in the envelope proteins of an avian hepadnavirus. *J. Virol.* **65**:1310-1317.
 27. **Tuttleman, J., C. Pourcel, and J. Summers.** 1986. Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells. *Cell* **47**:451-460.
 28. **Wang, G., and C. Seeger.** 1993. Novel mechanism for reverse transcription in hepatitis B viruses. *J. Virol.* **67**:6507-6512.
 29. **Wu, T.-T., L. Coates, C. E. Aldrich, J. Summers, and W. S. Mason.** 1990. In hepatocytes infected with duck hepatitis B virus, the template for viral RNA synthesis is amplified by an intracellular pathway. *Virology* **175**:255-261.