Topoisomerase I-Mediated Integration of Hepadnavirus DNA In Vitro

HAI-PING WANG AND CHARLES E. ROGLER*

Marion Bessin Liver Research Center, Department of Medicine, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

Received 26 November 1990/Accepted 11 February 1991

Hepadnaviruses integrate in cellular DNA via an illegitimate recombination mechanism, and clonally propagated integrations are present in most hepatocellular carcinomas which arise in hepadnavirus carriers. Although integration is not specific for any viral or cellular sequence, highly preferred integration sites have been identified near the DR1 and DR2 sequences and in the cohesive overlap region of virion DNA. We have mapped a set of preferred topoisomerase I (Topo I) cleavage sites in the region of DR1 on plus-strand DNA and in the cohesive overlap near DR2 and have tested whether Topo I is capable of mediating illegitimate recombination of woodchuck hepatitis virus (WHV) DNA with cellular DNA by developing an in vitro assay for Topo I-mediated linking. Four in vitro-generated virus-cell hybrid molecules have been cloned, and sequence analysis demonstrated that Topo I can mediate both linkage of WHV DNA to 5'OH acceptor ends of heterologous DNA fragments and linkage of WHV DNA into internal sites of a linear double-stranded cellular DNA. The in vitro integrations occurred at preferred Topo I cleavage sites in WHV DNA adjacent to the DR1 and were nearly identical to a subset of integrations cloned from hepatocellular carcinomas. The end specificity and polarity of viral sequences in the integrations allows us to propose a prototype integration mechanism for both ends of a linearized hepadnavirus DNA molecule.

Eucaryotic cells contain two classes of DNA topoisomerases, types I and II (Topo I and Topo II). These enzymes change the superhelical state of DNA through nicking-closing reactions. Topo II makes a transient, staggered, double-stranded break and passes the unbroken DNA strand through the double-strand break (56). Topo I introduces a transient, single-strand break which allows the nicked strand to unwind (11, 56). The enzyme remains covalently bound to the 3' phosphate at the nick site during swivelling, thus remaining in position to catalyze the closing reaction (12-14). However, if Topo I cleaves a DNA molecule in a single-stranded region, the nicked strands separate and the closing reaction is interrupted. Under these circumstances, 3' covalently bound Topo I molecules participate in inter-strand linking reactions with heterologous DNA molecules containing 5'OH ends (24, 31).

Although Topo I is a universal enzyme existing in all living organisms, its function has only recently begun to be understood. In *Saccharomyces cerevisiae*, mutants deleted for Topo I are still viable, and Topo II compensates for the loss of Topo I activity (51). The phenotype of yeast Topo I plus Topo II double mutants suggests that Topo I is involved in both DNA replication and transcription (19, 23). Support for a role in DNA replication has been obtained from studies demonstrating preferential association of Topo I with replicating simian virus 40 (SV40) molecules (15, 36) and other in vitro studies (61). Regarding transcription, Topo I has been found at sites of actively transcribing genes in diverse systems (5, 22, 47). Its role in DNA replication and transcription may be to relieve superhelical tension generated by movement of the replication or transcription complexes along the DNA (17, 60, 61).

A role for eucaryotic Topo I in illegitimate recombination is suggested by a growing body of evidence (16). Mapping of

preferred Topo I cleavage sites on SV40 DNA has shown that Topo I cleaves SV40 DNA in vitro in the same positions that viral excision and rearrangement occur in vivo (6, 7). Surveys of nonhomologous recombination sites in mammalian DNA have shown that trinucleotides that are preferentially cleaved by rat liver Topo I in vitro are present in the immediate vicinity of 92% of the crossover sites (27). The most compelling evidence that Topo I can promote illegitimate recombination in vivo is the recent study showing that Topo I from vaccinia virus causes lambda prophage excision by an illegitimate recombination mechanism (46).

Hepadnaviruses infect primarily the liver and cause persistent infections in approximately 5 to 10% of infected individuals (21, 30). Long-term persistent infection is associated with a high risk of hepatocellular carcinoma (HCC) (2). Early studies of HCCs from hepatitis B virus (HBV) carriers revealed that most of the HCCs contained clonal HBV DNA integrations (3, 4, 10, 28, 35, 43). Initial cloning and sequencing data from integrations revealed that they were the result of illegitimate recombination between viral DNA and the cellular genome (18, 44) and that virus-cell junctions occurred at many sites in both the viral and cellular DNA (32). In addition, rearrangements of integrated viral DNA, including deletions and direct and inverted duplications, are common (35, 40, 44). Cellular sequences flanking integrations also contain micro- and macrodeletions (32, 39) and inverted duplications (52, 59) and have served as sites for HBV-associated chromosome translocations in several instances (26, 52).

Since a large number of integrations have been cloned and sequenced, including integrations from both precancerous liver and tumors at various stages of malignant progression, it has become clear that two classes of integrations exist in vivo (32, 58). These include simple linear integrations (genome length or near genome length) and rearranged integrations containing either very short or very long (greater-than-genome-length) segments of viral DNA. Nucleotide

* Corresponding author.

FIG. 1. Illustration of the circular structure of WHV Dane particle DNA (referred to as WHV DNA in the text), including the nucleotide sequences in the vicinity of the 5' ends of the minus (–) and plus (+) strands. The positions of the envelope (S gene), core (C gene), terminal protein-reverse transcriptase (P gene), and X genes are noted by arrows around the circular map. Roman numerals I, II, III, and IV with associated arrows refer to the prototype integration types I to IV described by Shih et al. (45). Directly repeated sequences DR1 and DR2 are denoted as boxed areas of the genome. Topo I cleavage sites mapped by primer extension are illustrated by vertical arrows above the WHV (+) strand. Topo I cleavage motifs in the region adjacent to the DR1 sequence are denoted by solid lines below the WHV (+) and (–) strands. The solid double line denotes a break in the WHV sequence.

sequence analysis of many virus-cell junctions, particularly those from the simple linear integrations, has revealed several regions of the viral genome which are highly preferred integration sites (18, 26, 32, 45, 58, 59). The most highly preferred sites reside near two 11-bp directly repeated sequences, DR1 and DR2, which are the initiation points for minus- and plus-strand DNA synthesis, respectively (21). The region of the genome between these two points, designated the cohesive overlap, is also a preferred region for integration (Fig. 1). The plus and minus strands of hepadnavirus DNA are nicked at their initiation sites as a consequence of the viral DNA replication mechanism, which includes a reverse transcription step (48).

Preferred integration in the immediate vicinity of the DR1 and DR2 sites has led to the suggestion that linear molecules are the primary substrates for hepadnavirus integration (25, 32, 52) and that hepadnaviruses integrate by a strand invasion mechanism (45). One earlier report has pointed out the existence of preferred Topo I cleavage motifs in HBV and woodchuck hepatitis virus (WHV) DNA near DR1 and suggested a possible role of Topo I in viral integration (25). Illegitimate recombination of linear DNAs transfected into mammalian cells is also thought to be mediated by Topo I (27), and linearized hepadnavirus DNA resembles transfected DNAs in some ways. However, until this report, a specific enzymatic mechanism for hepadnavirus integration has been lacking.

Integration is not a part of the hepadnavirus replication cycle (29, 48), and hepadnaviruses do not contain an integrase gene (37). Therefore, cellular enzymes must mediate the viral integration pathway. The studies in this report demonstrate that Topo I cleaves WHV DNA at a cluster of bases near the DR1 sequence in plus-strand WHV DNA and in several regions within the cohesive overlap of WHV DNA

near DR2. Using an in vitro assay for linking, we have demonstrated that Topo I can also link viral DNA molecules into internal sites in a linear double-stranded cellular target DNA. The linked molecules closely resemble viral DNA integrations which occur in vivo, and a prototype integration mechanism involving Topo I has been suggested.

MATERIALS AND METHODS

Plasmids, enzymes, primers, PCR, and DNA sequencing. (i) **Plasmids.** Plasmid pW8 (20) contains the complete 3,308-bp WHV genome cloned in the unique *EcoRI* site of WHV and was used as the source of WHV DNA for subcloning and 32 P-hybridization probes and as template for sequencing reactions requiring WHV standards. Plasmids pW805 and pW803 were subgenomic clones of WHV DNA in plasmid pBR322 which contained the WHV *BglII-PstI* fragment, nucleotides 2534 to 3047, and the WHV *BamHI-HindIII* fragment, nucleotides 1529 to 2190, respectively. Plasmid pGEM-3Z (Promega) linearized at the *HincII* site was used to clone rearranged and integrated WHV DNA molecules from polymerase chain reaction (PCR)-amplified Topo I reaction mixes.

(ii) **Enzymes.** Calf thymus Topo I was obtained from Bethesda Research Laboratories, and restriction endonucleases, T4 DNA ligase, and T4 polynucleotide kinase were obtained from New England BioLabs.

(iii) **Primers.** The oligonucleotide primers used were as follows: primer a, WHV minus-strand DNA comprising nucleotides 5'-2030 to 2011-3'; primer b, WHV minus-strand DNA comprising nucleotides 5'-2052 to 2033-3'; primer c, WHV plus-strand DNA comprising nucleotides 5'-1730 to 1749-3'; primer d, C3 human DNA containing the nucleotide sequence 5'-CATATACTCAGATTTTCTT-3'; primer e, C3 human DNA containing the nucleotide sequence 5'-ATCTGCTTAGAAATCTTCAG-3'; primer f, this primer was used for the WHV sequencing ladder in the primer extension protocol and consisted of WHV minus-strand DNA from nucleotides 5'-2020 to 2001-3'. The WHV nucleotide numbering system was based on that of Galibert et al. (20).

(iv) **PCR.** A GeneAmp DNA amplification reagent kit with AmpliTaq DNA polymerase was purchased from Perkin-Elmer-Cetus and the PCRs were performed according to the protocols described for each experiment. The PCR reaction products were analyzed by agarose gel electrophoresis through a 2.5% Nusieve agarose gel.

(v) **DNA sequencing.** A Sequenase version 2.0 DNA sequencing kit was purchased from United States Biochemical Corporation, and [35 S]dATP (>1,000 Ci/mmol) was purchased from Amersham. Dideoxy chain termination reactions (42) were performed following the manufacturer's instructions and were analyzed by electrophoresis through an 8% polyacrylamide–8 M urea gel followed by drying of the gel under vacuum on a BioRad model 483 slab gel dryer. The gels were autoradiographed with Kodak XAR-5 film at –70°C overnight without intensifying screens.

Isolation of WHV DNA from Dane particles. Dane particles were isolated from the serum of a WHV carrier woodchuck (CW 624) as previously described (50), except that the virus was pelleted twice through a 15 to 30% sucrose step gradient and between the first and second sucrose gradients the pellet was treated with 1 μ g of DNase I per ml and 20 μ g of RNase A per ml in 10 mM MgSO_4 –10 mM Tris HCl (pH 7.4) at 37°C for 30 min. Endogenous labeling of WHV Dane particle DNA was carried out in a 100- μ l reaction mix containing 7 mM MgSO_4 , 50 mM NaCl, 1 mM dithiothreitol, 50 mM Tris

HCl, 0.1% Triton X-100, 100 μ M (each) dATP, dGTP, dTTP, and 100 μ Ci of [32 P]dCTP (3,000 Ci/mmol; Amersham Research Products) at 37°C for 10 min. This reaction was followed by the addition of unlabeled dCTP to 100 μ M and further incubation for an additional 2 h to allow completion of the plus strand. Endogenously labeled WHV Dane particle DNA was purified by adding EDTA to 10 mM, and sodium dodecyl sulfate (SDS) to 0.1% followed by proteinase K digestion (100 μ g/ml) at 37°C for 60 min, phenol extraction, and ethanol precipitation. WHV Dane particle DNA in which the incomplete plus strand was filled in by the endogenous polymerase activity is referred to as WHV DNA in this report. Whenever cloned WHV DNA was used, it is specifically mentioned in the text.

Topo I cleavage of WHV DNA. Purified WHV DNA was cleaved with calf thymus Topo I (Bethesda Research Laboratories) in a Topo I cleavage buffer containing 40 mM Tris HCl (pH 7.5), 120 mM KCl, 10 mM MgCl₂, 0.1 mM EDTA, 0.1 mM dithiothreitol, and 30 μ g of bovine serum albumin per ml. Various amounts of WHV DNA were incubated with 100 U of Topo I (approximately 7 μ g of protein) at 37°C for 10 min. The Topo I reaction was arrested by adding SDS to 1% followed by the addition of EDTA and NaCl to 10 mM and 100 mM, respectively. WHV DNA was then digested with 100 μ g of proteinase K per ml at 37°C for 30 min followed by phenol extraction, addition of 25 μ g of wheat germ RNA carrier, and ethanol precipitation. (The reactions shown in Fig. 2 were treated according to the experimental protocol described above). After being pelleted, WHV DNA was washed with 70% ethanol, air dried, and redissolved in water. As a control, an equal amount of WHV DNA was carried through the above procedure without the addition of Topo I. Topo I-treated and control WHV DNAs were further digested with restriction endonucleases, and the reaction products were analyzed by agarose gel electrophoresis and Southern blotting for the experiments shown in Fig. 3.

Primer extension analysis of Topo I-cleaved WHV DNA. Oligonucleotide primer f was purified through a 20% polyacrylamide-7 M urea gel and stored frozen in distilled water. Topo I-treated and control WHV DNAs without carrier RNA were denatured in 0.2 N NaOH at room temperature for 5 min, neutralized by the addition of 0.4 volumes of 5 M ammonium acetate (pH 7.5), and then precipitated with 4 volumes of ethanol (-70°C, 5 min). After being washed with 80% ethanol, the pelleted DNA was redissolved in 10 μ l of solution containing 40 mM Tris HCl (pH 7.5), 20 mM MgCl₂, 50 mM NaCl, and 100 ng of oligonucleotide primer f. Annealing was accomplished by heating the sample at 85°C for 2 min and then slowly cooling it to 35°C. A mixture containing 1.5 mM (each) dGTP, dCTP, dTTP, 20 mM dithiothreitol plus 1 μ l of [35 S]dATP (1,000 Ci/mmol; Amersham Research Products), and 6 U of Sequenase was added to the primer-annealed WHV DNA solution, and the primer extension reaction was carried out for 5 min at room temperature followed by the addition of 10 μ l of chase buffer containing 80 μ M (each) dGTP, dATP, dCTP, dTTP and 50 mM NaCl, followed by further incubation for 5 min at 37°C. The primer extension products were precipitated with ethanol, redissolved in 3 μ l of sequencing stop buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF), and then separated by electrophoresis at 50 W for 1.5 h through an 8% polyacrylamide-8 M urea sequencing gel.

A 661-bp WHV subgenomic fragment from plasmid pW803 (*Hind*III-*Bam*HI insert) was subcloned into phage M13mp18, and single-stranded DNA from this phage was

used as template for the sequence ladder in the primer extension experiment.

Protocols for generation and detection of Topo I-mediated linking and integration of WHV DNA. The basic reaction mixtures for these experiments contained 200 ng of WHV DNA and 3.5 μ g of Topo I in 20 μ l of Topo I cleavage buffer, as described above. The reactions were carried out for 1 h at 37°C and were stopped by the addition of SDS to 1%, followed by purification of the reaction products as described above. The final Topo I-treated WHV DNA pellets were taken up in 10 to 20 μ l of water, and aliquots were used for PCR amplification of rare reaction products.

(i) PCR amplification and cloning of WHV DNA linked to C3 human DNA. A 1.6-kb fragment of human C3 DNA (25) was used as a template for PCR reactions to amplify a 139-bp internal fragment of C3 DNA. The 50- μ l mixture contained 2 μ g of 1.6-kb C3 DNA and 1 μ M C3 DNA primers d and e and was subjected to 30 cycles of PCR consisting of 1 min of denaturation (94°C), 1 min of annealing (50°C), and 1 min of extension (72°C). The PCR products were purified by phenol extraction and ethanol precipitation followed by agarose gel analysis for homogeneity of product. The WHV linking reaction was carried out in a 10- μ l reaction mixture containing 20 ng of WHV DNA and 1.4 μ g of Topo I. The reaction was incubated at 37°C for 10 min followed by the addition of 200 ng of previously amplified 139-bp C3 DNA (in 1 μ l) and further incubation for 30 min. The same reaction scheme including WHV DNA and C3 DNA without added Topo I was used as control. The reaction was stopped with 1% SDS, and the DNA was purified as previously described. Hybrid C3-WHV molecules were selectively amplified by PCR using C3 primer d and WHV primer c. The amplified DNA was analyzed by Nusieve agarose gel electrophoresis, and specific amplified fragments were isolated from the agarose gel with GeneClean (Bio 101). The isolated fragments were phosphorylated with T4 polynucleotide kinase prior to cloning into the *Hinc*II site of pGEM-3Z. Transformed colonies of *Escherichia coli* JM109 were identified as white colonies on agar containing isopropyl- β -D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside and ampicillin (100 μ g/ml). White colonies containing hybrid C3-WHV pGEM-3Z plasmids were identified as follows. A micropipet tip was touched to the white colony and then rinsed in 20 μ l of PCR reaction mixture containing the two primers (in this case primers c and d) and next in LB plus ampicillin (100 μ g/ml) to establish a stock culture. The 20- μ l PCR reaction mixture was heated to 100°C for 10 min to denature the DNA and kill endogenous protease and then was cooled on ice. Amplitaq polymerase was added, and the PCR amplification was carried out by using the protocol that was used to initially amplify the Topo I-treated WHV DNA for each separate experiment. Agarose gel analysis of the amplified DNA identified those colonies which contained cloned hybrid C3-WHV DNA by the presence of an amplified band of the predicted size. Minipreparations of plasmid DNA were used for sequence analysis which was carried out from both ends of the WHV DNA insert, using both T7 and Sp6 primers or WHV primers in some cases. The complete sequence of the cloned insert was determined.

(ii) PCR amplification of integrated WHV DNA. The target DNA for the WHV integration reaction was the 1.6-kb C3 human DNA fragment. The Topo I integration reaction was carried out in a 20- μ l reaction mixture containing 50 ng of WHV DNA, 200 ng of C3 1.6-kb DNA, and 3.5 μ g of Topo I. The control reaction contained the same components minus Topo I. After incubation at 37°C for 1 h, the reaction

was stopped with SDS and reaction products were purified as described above.

Integrated WHV molecules were selectively amplified by PCR, using two combinations of viral and cellular primers. Integrated molecules corresponding to type I integrations were amplified using WHV primer a and cellular primer e, and type II integrations were amplified using WHV primer c and cellular primer e. Approximately 25% of the purified Topo I reaction products were used in each PCR amplification. A sample of each reaction product was analyzed by agarose gel electrophoresis, and the remaining products were cloned in the *HincII* site of pGEM-3Z as described above. Recombinant plasmid DNAs containing integrated WHV DNA were identified by PCR amplification using the two combinations of primers and the method described above. Minipreparations of plasmid DNA were used for DNA sequence analysis, using both T7 and Sp6 promoter primers to sequence the entire recombinant molecules in both directions.

RESULTS

Covalent attachment and limited cleavage of WHV DNA by Topo I. Because of the unique structure of WHV virion DNA (subsequently referred to as WHV DNA), we expected that Topo I cleavage of the plus and/or minus strands opposite the DNA replication initiation sites near DR1 and DR2, respectively, would result in linearization of the virion molecule (Fig. 1). This should allow us to detect accumulation of nicked reaction intermediates due to interruption of the closing reaction. These sites contain Topo I motifs plus an 8-bp terminally redundant tail (third DNA strand) near the DR1 site and a partially hybridized RNA primer at DR2 (Fig. 1). The effect of these structures on the efficiency of Topo I cleavage at those sites was unknown.

We first characterized the reaction products produced by Topo I treatment of WHV DNA. The single-stranded gap in virion DNAs was filled in by using the endogenous polymerase reaction during virus purification, and the 5' terminal protein was removed from purified WHV DNA with proteinase K. These treatments maintain the nicks in WHV DNA and the unique nucleic acid structures at the DR1 and DR2 sites. WHV DNA isolated in this manner was treated with calf thymus Topo I, and the reactions were stopped by the addition of SDS (12, 13). Agarose gel electrophoresis (in the absence of SDS) and Southern blotting of the reaction products showed that Topo I-treated WHV DNA formed high-molecular-weight complexes (Fig. 2, lane 4). Agarose gel electrophoresis in the presence of SDS resulted in disaggregation of the complexes; however, a significant portion of viral DNA continued to exhibit retarded migration compared with that of control WHV DNA (Fig. 2, lanes 5 and 6). This suggested that some WHV molecules remain covalently bound to Topo I after SDS treatment. Proteinase K treatment caused WHV DNA to migrate as a narrower band spanning the positions expected for double-stranded open circular and linear WHV molecules without covalently bound proteins (Fig. 2, lane 2). Heating of the Topo I plus proteinase K-treated samples caused the WHV DNA to migrate at the position expected for intact single-stranded viral molecules (Fig. 2, lane 1). These results demonstrated that individual strands of WHV DNA were largely intact after Topo I treatment. When proteinase K digestion was omitted but the sample was heated, a significant fraction (at least 20%) of WHV DNA exhibited retarded mobility (Fig. 2, lane 3) because of the presence of covalently bound Topo I.

FIG. 2. Southern blot analysis of WHV DNA treated with Topo I, illustrating limited cleavage of individual strands and covalent attachment of Topo I to WHV DNA. Topo I-treated WHV DNA was subjected to various combinations of proteinase K and/or heat treatment (100°C, 2 min) as noted above the lanes. The reaction products were separated in 0.8% agarose gels in the presence (SDS +) or absence (SDS -) of SDS, and the Southern blot was hybridized with a complete 3.3-kb ³²P-WHV DNA probe. The left-hand lambda standard provides a reference for lanes 1 to 4, and the right-hand lambda standard provides a reference for lanes 5 and 6. Sizes of the *HindIII*-digested lambda DNA fragments are in kilobases, from bottom to top.

Restriction endonuclease mapping of Topo I cleavage sites in WHV DNA. The presence of genome-length and some near-genome-length (only 250 bp shorter; Fig. 2, lane 3) WHV DNA in the previous experiment indicated that Topo I cleavage probably occurred near the nick sites in viral DNA and in the cohesive overlap region since cleavage of open circular WHV DNA at other positions would generate small subgenomic viral DNA fragments upon denaturation, and such molecules were not present. In order to map the Topo I cleavage sites, WHV DNA was digested with single-cut restriction endonucleases after Topo I treatment and the reaction products were analyzed by Southern blotting. Using a *BglII-PstI* subgenomic WHV probe (pW805), we detected a series of fragments unique to Topo I digestion (Fig. 3A, asterisks) as well as to a set of fragments that were enriched by Topo I digestion (Fig. 3A, circles). This was particularly evident for samples treated with *HindIII*, *PstI*, *EcoRI*, and *BamHI*. The WHV regions present in each of the unique and enriched fragments are diagrammed as arrows around the circular WHV map to the right of each Southern blot. The Topo I-enriched fragments (circles) were present in both the Topo I-treated and control DNAs. The sizes of these fragments corresponded to those which would be expected if WHV molecules became linearized either by melting and filling in of the cohesive overlap region during the endogenous polymerase filling reaction or by Topo I cleavage of the plus strand opposite the nick in the minus strand. From these results, we estimated that molecules linearized by melting and filling in of the cohesive overlap region constituted approximately 5% of the native WHV DNA molecules in our

FIG. 3. Restriction endonuclease mapping of Topo I cleavage sites in WHV DNA by Southern blot hybridization with various ^{32}P -WHV DNA probes. The interpretation of the genomic locations of the various WHV fragments produced by restriction endonuclease cleavage of Topo I-treated WHV DNA is shown to the right of each blot. (A, left) Southern blot of Topo I-treated and control WHV DNA hybridized with WHV subgenomic probe pW805. Restriction fragments unique to Topo I cleavage are marked with asterisks and those enriched by Topo I cleavage are marked with circles. Samples were untreated (Un) or treated with *PvuI* (Pv), *HindIII* (H), *PstI* (Pst), *EcoRI* (E), or *BamHI* (B). Lane λ, size markers (in kilobases). (A, right) Partial restriction map of circular WHV DNA, with outside arrows denoting the WHV regions present in each of the unique and enriched fragments generated by Topo I plus restriction endonuclease. Asterisks and circles adjacent to the arrows correspond to the WHV fragments marked in the Southern blot. Arrows marked with an asterisk represent WHV DNA fragments which are continuous from the origin of the arrow at the appropriate restriction site, and those arrows marked with circles represent fragments which end at the internal arrowhead. (B, left) Southern blot analysis of WHV DNA treated with various combinations of Topo I and/or *HindIII* and hybridized with WHV subgenomic probe pW803. Treatments are denoted above the lanes, and fragments unique or enriched by Topo I are designated 1, 2, or 3. (B, right) Proposed WHV map positions of fragments 1, 2, and 3, illustrated by arrows around the WHV genome map. The leftmost lane of the Southern blot contains lambda *HindIII* standards (in kilobases). The rightmost lane contains ϕX174 *HaeIII* fragments (in kilobases).

preparation. The lengths of fragments unique to Topo I treatment were those expected for WHV molecules which are cleaved in the minus strand in the vicinity of the DR2 sequence (Fig. 3A, asterisks).

Since the *BglII-PstI* subgenomic probe would not detect WHV fragments generated by Topo I cleavage near both the DR1 and DR2 sites, a separate preparation of WHV DNA was cleaved by Topo I followed by *HindIII* digestion and a second Southern blot was hybridized with a subgenomic probe spanning the complete cohesive overlap region (pW803). This allowed us to identify fragments produced by Topo I cleavage near the DR1 site (Fig. 3B, bands 1), near the DR2 site (Fig. 3B, band 2), or including both the DR1 and DR2 sites (Fig. 3B, bands 3). Again, WHV genomic regions present in fragments 1, 2, and 3 are diagrammed by arrows around the WHV map to the right of the Southern blot. Fragment 1 is present in WHV DNA treated with *HindIII* only, because of the presence of linear molecules in the

WHV DNA preparation. Fragment 3 was produced by digestion at both the DR1 and DR2 sites since it was also present in the Topo I only lane.

Primer extension analysis of Topo I cleavage sites in WHV plus-stranded DNA. Since the sequences in the vicinity of DR1 are the most highly preferred integration sites *in vivo*, we chose to precisely map the Topo I cleavage sites in the vicinity of DR1 and to focus on this area for our further studies. WHV DNA was digested with Topo I, and the plus-strand cleavage products in the vicinity of DR1 were analyzed by primer extension, using a minus-strand oligonucleotide primer spanning nucleotides 5'-2020 to 2001-3'. The positions of the cleavage sites were determined by comparison with a WHV sequencing ladder generated with cloned WHV DNA and using the same primer (Fig. 4, lanes 1 to 4, minus-strand sequence shown). This analysis would only detect those molecules in which the Topo I closing reaction was interrupted. According to this analysis, most of the

FIG. 4. Mapping of Topo I cleavage sites in WHV DNA plus strands by primer extension. WHV minus-strand primer f was used for primer extension of Topo I-treated WHV DNA. Lanes 1 to 4, DNA sequence ladder of WHV minus strands obtained using primer f and cloned WHV DNA. Lane 5, primer extension products of WHV DNA cleaved with Topo I. Note the unique set of bands between nucleotides 1931 and 1938. Lane 6, primer extension products of control WHV DNA. The nucleotide sequence of plus-strand WHV DNA is illustrated to the right of the sequencing gel, and Topo I cleavage sites are denoted by arrows adjacent to the DNA sequence. The length of the arrows corresponds to the relative intensity of the Topo I-generated band on the primer extension gel. The uppermost arrow to the right of lane 6 is the fully extended WHV DNA product.

WHV DNA in both the control and Topo I reactions remained intact as judged by the common high-molecular-weight primer extension product (Fig. 4, lanes 5 and 6, see topmost arrow adjacent to lanes). However, Topo I cleavage products were identified as a cluster of six Topo I-specific

FIG. 5. (A) Illustration of the experimental protocol and proposed mechanism of Topo I-mediated linking of WHV DNA to cellular DNA. (Upper right) WHV DNA. (Upper left) C3 cellular DNA. (Left middle) Internal fragment of C3 DNA amplified by PCR using cellular primers d and e to generate a 139-bp DNA fragment with 5'OH ends. (Middle) Proposed linking mechanism (see text). (Bottom) Selective PCR amplification of virus-cell DNA reaction products using cellular primer d and WHV primer c. The narrow arrow inside the circular WHV molecule indicates a Topo I nicking site in WHV DNA. The Topo I molecule is denoted as a hatched circle. (B) Nucleotide sequence of a clone obtained by PCR amplification of a Topo I linking reaction containing WHV DNA and C3 DNA, according to the protocol in panel A. The hybrid molecule was amplified using primers c and d.

bands which corresponded to Topo I cleavage sites on the 3' side of nucleotides 1931, 1934, 1935, 1936, 1937, and 1938 (Fig. 4, lane 5). Control WHV DNA was not cleaved at these sites (Fig. 4, lane 6). These Topo I cleavage sites occurred in WHV plus-strand DNA opposite the nick in the minus strand and were partially within the terminal redundancy in virion minus-strand DNA (vertical arrows in Fig. 1 denote positions of Topo I cleavage sites).

Topo I-mediated linking of WHV DNA to 5'OH ends of cellular DNA. As a first step in studying Topo I-mediated integration, we tested whether Topo I could mediate linking of WHV DNA to free ends of cellular DNA. Previous studies had demonstrated that DNAs containing 3' covalently bound Topo I molecules could function as donors in linking reactions with heterologous DNAs containing 5'OH acceptor ends (24). Therefore, we performed an experiment to determine whether WHV molecules treated with Topo I could function as donor molecules in a linking reaction with a human cellular DNA containing 5'OH acceptor ends. We first produced a 139-bp cellular DNA fragment containing 5'OH ends by PCR (see Materials and Methods) and used it as an acceptor molecule in our linking experiment (Fig. 5). A reaction was carried out in which WHV DNA was first incubated with Topo I followed by the addition of the 139-bp fragment to the reaction mix. The reaction products were amplified by PCR using one plus-strand WHV oligonucleotide spanning nucleotides 5'-1730 to 1749-3' (primer c) and one human cellular DNA oligonucleotide (primer d) which

permitted only virus-cell hybrid DNA molecules to be amplified (Fig. 5A). One of these molecules was cloned into pGEM-3Z, and nucleotide sequence analysis showed that it was a virus-cell hybrid molecule in which WHV plus-strand DNA at nucleotide 1820 was linked to the 5'OH end of cellular DNA primer e (Fig. 5B). This finding was expected from our experimental design since primers c and d had been used to amplify hybrid molecules in the PCR reaction. This result demonstrated that a Topo I molecule mediated linking of WHV plus-strand DNA to the free 5'OH of the cellular DNA fragment. A highly preferred Topo I cleavage motif, 5'-CTT-3', was present in WHV DNA between nucleotides 1819 and 1821 (27).

Topo I-mediated WHV DNA integration into internal positions of double-stranded cellular DNA. Having established Topo I-mediated linking of ends of WHV and cellular DNA fragments, we next focused on testing whether Topo I could link WHV DNA at internal positions in double-stranded cellular DNA. For these experiments, our target DNA was a 1.6-kb human DNA fragment, C3, which had been previously cloned and partially sequenced. The C3 fragment was the normal allele from an HBV-associated HCC in which the other allele had served as an HBV integration site. Knowing the sequence across the HBV integration site in the normal allele allowed us to design an experiment to determine whether in vitro linking could occur at the same site as integration occurred in vivo. Interestingly, the in vivo integration of HBV had caused a microdeletion of C3 DNA at the integration site, and HBV and C3 cellular DNA shared a 5-bp homology at one junction site (25).

An in vitro linking reaction was performed by mixing WHV DNA and C3 DNA in the presence or absence of Topo I. We assayed for rare linked molecules by using a PCR reaction which contained one cellular and one viral primer. In this way, we selectively amplified virus-cell hybrid molecules produced by Topo I-mediated reactions. In separate PCR reactions we used either a WHV primer in the core gene region, primer a, or, in the cohesive overlap region, primer e. In this way, we were able to assay for linked WHV molecules with different polarities. Using the minus-strand core gene primer we could amplify DNA molecules containing core gene sequences, and using the plus-strand cohesive overlap primer (located immediately adjacent to DR2) we were able to amplify DNA molecules containing sequences in the entire cohesive overlap region. The C3 oligonucleotide we used (primer e) was approximately 70 bp away from the known HBV integration site in C3 DNA. This would allow us to amplify WHV integrations within 70 bp 3' of the HBV integration site and, several hundred base pairs upstream, 5' to the integration site.

Agarose gel analysis of the products of the PCR reaction which utilized the WHV core gene primer (primer a) and C3 primer e revealed a specific DNA fragment of approximately 150 bp unique to the Topo I-treated DNAs (Fig. 6, lanes 4 and 5). Similar analysis of the PCR reaction products using the WHV cohesive overlap primer c and C3 primer e revealed a major band of approximately 250 bp and two minor bands of approximately 300 and 400 bp which were unique to the Topo I-treated DNAs (Fig. 6, lanes 2 and 3). Since these fragments were likely to be virus-cell hybrid molecules, we cloned and sequenced the major 150- and 250-bp fragments.

Nucleotide sequence analysis of the smaller fragment revealed that it was a virus-cell hybrid in which WHV DNA linked to C3 DNA between WHV nucleotides 1931 and 1935 (Fig. 7, lines 1). DNA sequences proceeding toward the

FIG. 6. Agarose gel analysis of PCR amplification products from WHV integration reactions. Lanes: 1, ϕ X174 RF *Hae*III-digested DNA standards in base pairs (top to bottom); 2 and 3, PCR reaction products from control (lane 2) and Topo I-treated (lane 3) reactions utilizing WHV primer c and cellular primer e to assay for linking in the cohesive overlap region; 4 and 5, PCR reaction products from control (lane 4) and Topo I-treated (lane 5) reactions utilizing WHV primer a and cellular primer e to assay for linkage in the core gene.

WHV core gene were present in the fragment, as expected for type I integrations (45). The exact position of the WHV linkage site could not be determined unambiguously because the viral DNA shared a 5-bp homology with C3 DNA at the linkage site (Fig. 7, lines 3). In addition, WHV DNA integrated in C3 DNA near the site in which HBV had integrated in vivo (9 to 13 bp away) and the nucleotides between the two sites (in vitro versus in vivo) were the same ones which had been deleted when HBV integration occurred in this region in vivo (25).

Nucleotide sequence analysis of the larger (~250-bp) fragment from the second PCR reaction revealed that it was also a virus-cell hybrid. In this case, WHV linkage occurred between nucleotides 1938 and 1941, with WHV sequences proceeding into the cohesive overlap toward the DR2 sequence as would be expected for type II integrations (Fig. 7, lines 2). In this case, a 3-bp homology was observed between viral and cellular DNA at the linkage site (Fig. 7, lines 3). This linkage reaction occurred 10 to 13 bp (ambiguity due to the region of homology) downstream from the first in vitro linkage site (Fig. 7, lines 3) and in the opposite strand of C3 DNA. These results require that Topo I cleaved opposite strands of C3 DNA at staggered positions, 10 to 14 bp apart, in separate reactions. If such cleavage could also occur in vivo, it would facilitate breakage of the DNA and generation of two cellular DNA fragments containing cohesive ends. Such a mechanism has been previously proposed (6).

Previous reports have shown that nucleotide sequences composed of five or more A's and T's are preferred sites for nonhomologous recombination in mammalian cells (27). In agreement with this, the five-nucleotide sequence, 5'AT TTA-3' was found to be present at each of the integrations which occurred in the 31-bp region of C3 DNA (sequence denoted by bars above the C3 sequence in Fig. 7, lines 3). This included the in vivo HBV integration and the two in vitro WHV linkage sites. Therefore, the presence of three A-T stretches in addition to preferred Topo I motifs at each linkage site appears to make this region a highly preferred target.

Virus-cell hybrid molecules are not produced by PCR amplification of WHV and C3 DNAs treated with Topo I separately. Since WHV DNA shared 3 to 5 bp of homology with C3 DNA at both in vitro linkage sites, the possibility existed that WHV DNA cleaved by Topo I could have

FIG. 7. (Lines 1) Nucleotide sequence analysis of WHV-C3 DNA hybrid molecule with a virus-cell junction between WHV nucleotides 1931 and 1935. The bracketed region with an asterisk denotes a region of 5-bp homology. WHV primer a and C3 primer e were used to amplify the molecule. (Lines 2) Nucleotide sequence of a WHV C3 DNA hybrid molecule with a virus-cell junction between WHV nucleotides 1938 and 1941. The bracketed region with an asterisk denotes a region of 3-bp homology between viral and cellular DNA at the linkage site. WHV primer c and C3 primer e were used to amplify the molecule. (Lines 3) Comparison of the WHV linkage sites in C3 DNA. The locations of WHV linkage sites are illustrated in relation to the intact C3 DNA sequence. The WHV DNA sequences of intact plus and minus strands across the integration sites are illustrated on the top and bottom lines labeled (+) WHV and (-) WHV, respectively. The 5'-ATTTA-3' triplicate repeats are denoted by a solid bar above the intact C3 DNA sequence. The sequences across the integration sites of the two integrations are noted in the lines labeled 1931 integration and 1938 integration. The middle line is the intact C3 sequence. HBV integration occurred in vivo within the left-hand 5'-ATTTA-3' repeat.

served as a primer for *Taq* polymerase extension into C3 DNA during in the PCR reaction, generating virus-cell hybrid molecules. Although it was unlikely that such short homologous regions would be stable at 72°C and able to prime the *Taq* polymerase extension reaction, we conducted additional control experiments to test this and other unforeseen possibilities. In a series of reactions, WHV DNA and C3 DNA were treated separately with Topo I and then the DNAs (purified to remove Topo I) were mixed and used in PCR reactions. We added a new WHV minus-strand primer, primer b, spanning WHV nucleotides 5'-2033 to 2052 3' to the PCR reactions so that we could distinguish any new hybrid molecule from those which had been previously produced. The same C3 primer, primer e, was used as before, so that we could test for additional linkage products in the same cellular DNA region. If nicked WHV DNA could function as its own primer by hybridizing to C3 DNA, amplified fragments should be produced in the PCR reaction mixtures. PCR reactions with WHV and C3 DNAs alone were also carried out.

Agarose gel analysis of the PCR reactions containing either WHV DNA, C3 DNA, or the mixture of WHV plus C3 DNA showed that no fragments were amplified (Fig. 8, lanes 7, 8, and 9, respectively). Additional control reactions using WHV, C3, and pGEM-3Z DNAs which were not treated with Topo I also did not produce any amplified fragments (Fig. 8, lanes 3, 4, and 5, respectively). PCR reactions with the same WHV and C3 DNA mixtures, except for different primers (WHV primer c in the cohesive overlap plus C3 primer e) also did not yield any amplified fragments (Fig. 8, lanes 13, 14, and 15). These control reactions were repeated several times, each time with the same results, leading us to conclude that self priming did not account for the generation of hybrid molecules in our experiments.

Topo I-mediated integration occurs preferentially at short regions of homology. At the same time that we conducted the above negative controls, we repeated the Topo I integration

experiment we had performed earlier, except that we utilized the new WHV minus-strand primer, primer b, 20 bp 5' of the first core gene primer. In two separate PCR reactions, we detected single amplified fragments unique to the Topo I-treated DNAs. These fragments were approximately 20 bp larger than the previous integration we had produced in the core gene region (Fig. 8, lanes 1 and 10). Control mixtures utilizing the same DNAs and reaction components but not Topo I treated yielded no amplified fragments (Fig. 8, lanes 2 and 11). One of the amplified DNA fragments was cloned and sequenced, and the analysis revealed that WHV integration had occurred at the exact same position in both viral and cellular DNA (between WHV nucleotides 1931 and 1935) as had occurred in the first core gene integration. The core gene sequences in the cloned integration extended 22 bp farther into the core gene (to nucleotide 2052) than those in the first integration, proving that the molecule was produced by a separate linking reaction (Fig. 9, lower sequence). These results suggest that Topo I-mediated illegitimate recombination is enhanced by short regions of homology between viral and cellular DNA, especially if the regions of homology are coincident with preferred Topo I cleavage motifs.

DISCUSSION

The observation that Topo I cleavage sites are associated with runs of purines (or pyrimidines) and A/T-rich regions has led to the suggestion that unusual DNA structures preferentially enhance Topo I-mediated illegitimate recombination (8). In support of this, Topo I has been shown to be preferentially associated with replicating SV40 molecules (15) and camptothecin, which blocks the Topo I closing reaction, induces DNA breakage at replication forks (1). Only 11% of the potential Topo I cleavage sites were cleaved in replicating SV40 molecules, and those that were cleaved were clustered near the terminus for DNA replication (36). This cleavage also occurs asymmetrically, such that Topo I

FIG. 8. Agarose gel analysis of PCR reaction products from experiments in which WHV, C3, and pGEM-3Z DNAs were subjected to various combinations of Topo I treatment followed by PCR amplification with WHV and C3 oligonucleotide primers. Lanes: 1 and 10, Topo I treatment of WHV plus C3 DNA mixtures followed by PCR using C3 primer e and WHV minus-strand primer b; 2 and 11, control reactions for lanes 1 and 10 in which Topo I treatment of C3 and WHV DNAs was omitted; 3, 4, and 5, PCR amplification of WHV, C3, and pGEM-3Z DNAs without Topo I treatment, respectively, using the same primers as in lane 1; 7 and 8, PCR amplification of WHV (lane 7) and C3 (lane 8) DNAs treated separately with Topo I and amplified separately using the same primers as in lane 1; 9, PCR amplification of a WHV and C3 DNA mixture in which Topo I-treated DNAs from lanes 7 and 8 were mixed (after separate Topo I treatment and purification) and subjected to PCR amplification using the same primers as in lane 1; 13 and 14, PCR amplification of WHV and C3 DNAs, respectively, treated as in lanes 7 and 8 except that PCR amplification was carried out by using WHV plus-strand primer c and C3 primer e; 15, PCR amplification of a WHV and C3 DNA mixture treated as in lane 9 except that WHV plus-strand primer c and C3 primer e were used for PCR amplification.

cleavage sites are almost exclusively located on the DNA strand that is the template for discontinuous DNA synthesis (36, 53).

A close analogy can be drawn between Topo I cleavage of specific regions of SV40 DNA and Topo I cleavage of WHV DNA in the vicinity of the DR1 sequence. As shown in Fig. 1, the DR1 region is analogous to a stationary replication fork in which the WHV plus strand serves as the template for replication of the minus strand. This, of course, is not the natural mode of hepadnavirus replication. However, the presence of a forked structure makes these molecules prime targets for Topo I cleavage. This predisposition is further enhanced in WHV DNA by the presence of both preferred Topo I trinucleotide cleavage sites (underlined nucleotides in Fig. 1) and an A/T-rich region in plus-strand DNA. Topo I cleavage of WHV DNA in the plus strand opposite the nick in the minus strand would be expected to linearize the molecule and either prevent or slow down the Topo I closing reaction. Linearization was probably responsible for our ability to detect Topo I cleavage intermediates and to map the cleavage sites in WHV DNA near DR1.

It has previously been proposed that Topo I mediates illegitimate recombination by cleaving opposite strands of

FIG. 9. Complete DNA sequence analysis of two in vitro-generated WHV-C3 DNA molecules. WHV linking occurred between nucleotides 1931 and 1935, as noted by the bracket labeled viral-cell junction. WHV primers a (upper) and b (lower) plus C3 primer e were used to amplify the virus-cell hybrids.

cellular DNA such that the number of base pairs between the two break sites is insufficient to hold the two ends of the DNA molecule together and that these ends then function as either donor or acceptor ends in intermolecular reactions with linearized SV40 DNA molecules which are similarly cleaved by Topo I (6). Our in vitro-generated WHV integrations occurred in opposite strands of C3 DNA at sites which were separated by 10 to 13 bp (Fig. 7, line 3, and Fig. 10, upper right). Therefore, our data provide experimental support for the previously proposed model. In addition, the results suggest that Topo I cleavage can produce cellular DNA molecules with short cohesive ends. Since these single-strand ends are sensitive to cellular nucleases, their deletion or truncation could explain why microdeletions occur at many hepadnavirus integration sites.

We have constructed molecular models to explain the mechanism of the Topo I-mediated linking reactions (Fig. 10). The DNA sequence of the first integration is consistent with a mechanism in which WHV DNA was cleaved by Topo I at nucleotide 1931, resulting in linearization of the WHV molecule (Fig. 10A). The free 5'OH end of the plus strand was then available to link to cellular DNA which contained a 3' bound Topo I. The region of 5-bp homology between WHV and cellular DNA suggests that it stabilized the formation of the hybrid molecule during the linking reaction. In the second case (Fig. 10B), we propose that WHV plus-strand DNA was cleaved at nucleotide 1941 and that the 3' end of the plus-strand DNA containing a 3' bound Topo I molecule became linked to a 5'OH acceptor end in cellular DNA. In this case, a short 3-bp region of homology between viral and cellular DNA was evident. According to these models, WHV DNA served as a 5'OH acceptor end in one case (Fig. 10A) and in the other case (Fig. 10B) it provided a 3' Topo I-containing donor end. Our experiments only tested linking of a single strand of WHV DNA at each site. In vivo, it would be necessary to complete the integration of the opposite strand either by a second Topo I linking reaction or via other cellular DNA polymerases and ligases.

The end specificity and polarity of the separate virus-cell DNA hybrid molecules lead us to propose a prototype model for Topo I-mediated integration of both ends of a linearized WHV molecule (Fig. 11). According to the model, Topo I linearizes circular WHV DNA and the Topo I-containing end functions as a donor during WHV integration. The opposite end, containing a 5'OH group, functions as an acceptor site for other heterologous cellular DNAs containing 3' bound Topo I molecules. In order for a linearized viral molecule to

FIG. 10. (A) Molecular model illustrating the proposed mechanisms for Topo I-mediated linking at WHV nucleotide 1931. WHV DNA (circular molecule) and C3 DNA (linear double-stranded molecule) are nicked by Topo I. (Middle) The WHV molecule has been cleaved by Topo I at nucleotide 1931. The free 3' end is the naturally occurring 3' end of the WHV minus strand. C3 DNA (right) is nicked in opposite strands, yielding two fragments, one of which becomes linked to WHV DNA. (Bottom) Proposed linking reaction. (B) Molecular model illustrating the proposed mechanism for Topo I-mediated linking at WHV nucleotide 1941. (Top line) Topo I cleaves WHV DNA (right) and C3 DNA (left). (Middle left) Topo I nicks C3 DNA at positions 12 bp apart in opposite strands. (Middle right) A WHV molecule is nicked at nucleotide 1941 in the plus strand and contains a 3' bound Topo I (circle with hash marks). (Bottom) Proposed linking reaction.

integrate, it must insert into the reaction site in cellular DNA at the moment of Topo I nicking and provide an alternate acceptor or donor end. Such ends might occur preferentially at DNA replication forks or within actively transcribed genes, because Topo I is present at these locations. The availability of donor and acceptor ends is probably the rate-limiting step for viral DNA integration *in vivo*, and the complexity of such reactions may be an important factor which accounts for the low frequency of integration observed *in vivo*.

The hepadnavirus integration mechanism presented has some features in common with lambda phage integration (for reviews, see references 9 and 54). The lambda integrase protein (Int) is a topoisomerase which generates staggered nicks in lambda DNA. The integrase protein normally mediates integration of lambda DNA into the *E. coli* genome in a site-specific manner in association with accessory proteins.

FIG. 11. Proposed prototype mechanism for Topo I-mediated hepadnavirus integration. (Top line) WHV virion DNA left, cellular DNA right. (Middle line) Topo I linearizes WHV DNA by nicking the plus strand opposite the terminal redundancy in the minus strand. Topo I also cleaves cellular DNA at staggered positions, causing them to separate. (Bottom line) Topo I remains covalently bound to the 3' phosphate of WHV plus strands and mediates linking (left virus-cell junction). The opposite end of the plus strand contains a free 5'OH group which becomes linked to a cellular DNA via a 3' bound Topo I molecule (right virus-cell junction). The DR1 and DR2 sequences in WHV virus DNA are noted by boxes labeled 1 and 2, respectively. Circles with hash marks are Topo I molecules. Arrows immediately adjacent to DNA molecules are Topo I nicking sites. Thick squiggle line opposite DR2 is the plus-strand RNA primer. Dots are regions of the molecules which must be filled in by cellular DNA polymerases.

However, the Int protein can be induced to execute single-strand DNA transfer in a nonreciprocal manner when altered integration targets are used (33). If the integration targets are altered such that the integration intermediate, cleaved at staggered positions, is not held together by Watson-Crick base pairing, the strands separate after Int cleavage and participate in illegitimate recombination (33, 34). Thus, other viral integration systems appear to have important common features with our model, while also exhibiting their own unique characteristics.

Our integration model also proposes that primary integrations will usually contain linear viral DNA sequences which may have deleted part of the WHV genome but are not extensively rearranged. Most of the examples of *in vivo* integrations with this structure have been cloned from pre-cancerous liver tissue or small developing HCCs (40, 58). In contrast, most integrations cloned from larger highly malignant HCCs contain highly rearranged viral sequences (32, 35). In a series of experiments to be published separately, we have generated inversions in WHV DNA *in vitro*, utilizing Topo I, that closely resemble some of the inversions observed in this second group of rearranged integrations. Whether rearrangements of viral DNA occur before or after integration is unknown.

Two pathways have been demonstrated as sources of open circular molecules which are converted to covalently closed circular (CCC) molecules in the nucleus of persistently infected hepatocytes. These include molecules from infect-

ing viral Dane particles and replicative intermediates in core particles which are cycled into the nucleus instead of being excreted (54, 55, 57). The exact mechanism and site of conversion of open circular molecules to CCC DNA are unknown. However, it is quite likely that one step in the pathway involves uncoating and release of open circular virion DNA molecules into the nucleus followed by cleavage of the 5' terminal protein. These open circular molecules would be identical in structure to those which we utilized for our in vitro integration experiments and would be expected to be present in the nucleus and available for cleavage by Topo I.

The recent demonstration that recycling of viral DNA into the nucleus is suppressed to a very low level during productive infection (49) helps to explain why CCC DNA copy number is stable in the nucleus during persistent infection and may also explain why integrations are rare under normal conditions. The further demonstration that a blockage in viral envelope production prevents core particle excretion and causes a significant accumulation of open circular and CCC DNAs in the nucleus (49) supports the notion that conditions can be established during persistent infection which would promote the occurrence of viral integration. This idea is consistent with studies showing that viral integrations accumulate in the livers of long-term persistent HBV carriers which have developed a blockage in the viral replication cycle (38).

ACKNOWLEDGMENTS

We thank Okio Hino for providing a plasmid clone of the 1.6-kb C3 human DNA and the nucleotide sequence across the HBV integration site and Walter Ogston for WHV subgenomic clones pW803 and pW805. The authors also thank David Shafritz, Robert Burk, Raju Kucherlapati, Stewart Shuman, and Peter Bullock for critical evaluation of the manuscript.

This work was supported by Public Health Service grant CA-37232 from the National Cancer Institute, grant DK-17702 from the Digestive Disease Center grant program, and grant FRA-316 from the American Cancer Society.

REFERENCES

- Aveman, K., R. Knippers, T. Koller, and J. M. Sogo. 1988. Camptothecin, a specific inhibitor of type I DNA topoisomerase, induces DNA breakage at replication forks. *Mol. Cell. Biol.* **8**:3026-3034.
- Beasley, R. P., C. C. Kiu, L. Y. Hwang, et al. 1981. Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22,707 men in Taiwan. *Lancet* **ii**:1129-1133.
- Brechot, C. M., M. Hadchouel, J. Scotto, M. Fonck, F. Potet, G. N. Vyas, and P. Tiollais. 1981. State of hepatitis B virus DNA in hepatocytes of patients with hepatitis B surface antigen-positive and -negative liver diseases. *Proc. Natl. Acad. Sci. USA* **78**:3906-3910.
- Brechot, C., C. Pourcel, A. Louise, B. Rain, and P. Tiollais. 1980. Presence of integrated hepatitis B virus DNA sequences in cellular DNA in human hepatocellular carcinoma. *Nature (London)* **286**:533-535.
- Brill, S. J., S. DiNardo, K. Voelkel-Meiman, and R. Sternglanz. 1987. Need for DNA topoisomerase I activity as a swivel for DNA replication and for transcription of ribosomal RNA. *Nature (London)* **326**:414-416.
- Bullock, P., J. J. Champoux, and M. Botchan. 1985. Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. *Science* **230**:954-958.
- 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.
- Camilloni, G., E. Di Martino, M. Caserta, and E. di Mauro. 1988. Eukaryotic DNA topoisomerase I reaction is topology dependent. *Nucleic Acids Res.* **16**:7071-7085.
- Campbell, A. 1984. Types of recombination: common problems and common strategies. *Cold Spring Harbor Symp. Quant. Biol.* **49**:839-844.
- Chakraborty, P. R., N. Ruiz-Opazo, D. Shouval, and D. A. Shafritz. 1980. Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular carcinoma cell line. *Nature (London)* **286**:531-533.
- Champoux, J. J. 1976. Evidence for an intermediate with a single-strand break in the reaction catalyzed by the DNA untwisting enzyme. *Proc. Natl. Acad. Sci. USA* **73**:3488-3491.
- Champoux, J. J. 1977. Strand breakage by the DNA untwisting enzyme results in the covalent attachment of the enzyme to DNA. *Proc. Natl. Acad. Sci. USA* **74**:3800-3804.
- Champoux, J. J. 1978. Mechanism of the reaction catalyzed by the DNA untwisting enzyme: attachment of the enzyme to the 3'-terminus of the nicked DNA. *J. Mol. Biol.* **118**:441-446.
- Champoux, J. J. 1981. DNA is linked to the rat liver DNA nicking-closing enzyme by a phosphodiester bond to tyrosine. *J. Biol. Chem.* **256**:4805-4809.
- Champoux, J. J. 1988. Topoisomerase I is preferentially associated with isolated replicating simian virus 40 molecules after treatment of infected cells with camptothecin. *J. Virol.* **62**:3675-3683.
- Champoux, J. J., and P. A. Bullock. 1988. A possible role for the eucaryotic type I topoisomerase in illegitimate recombination, p. 655-666. In R. Kucherlapati and G. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.
- Champoux, J. J., and R. Dulbecco. 1972. An activity from mammalian cells that untwists superhelical DNA: a possible swivel for DNA replication. *Proc. Natl. Acad. Sci. USA* **69**:143-146.
- 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.
- DiNardo, S., K. Voelkel, and R. Sternglanz. 1984. DNA topoisomerase II mutant of *Saccharomyces cerevisiae*: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proc. Natl. Acad. Sci. USA* **81**:2616-2620.
- Galibert, F., T. N. Chen, and E. Mandart. 1982. Nucleotide sequence of a cloned woodchuck hepatitis virus genome: comparison with the hepatitis B virus sequence. *J. Virol.* **41**:51-65.
- Ganem, D., and H. E. Varmus. 1987. The molecular biology of the hepatitis B virus. *Annu. Rev. Biochem.* **56**:652-693.
- Gilmour, D. S., and S. C. R. Elgin. 1987. Localization of specific topoisomerase I interactions within the transcribed region of active heat shock genes by using the inhibitor camptothecin. *Mol. Cell. Biol.* **7**:141-148.
- Goto, T., and J. C. Wang. 1985. Cloning of yeast TOP 1, the gene encoding DNA topoisomerase I, and construction of mutants defective in both DNA topoisomerase I and DNA topoisomerase II. *Proc. Natl. Acad. Sci. USA* **82**:7178-7182.
- Halligan, B. D., J. L. Davis, K. A. Edwards, and L. F. Liu. 1982. Intra- and intermolecular strand transfer by HeLa DNA topoisomerase I. *J. Biol. Chem.* **257**:3995-4000.
- Hino, O., K. Ohtake, and C. E. Rogler. 1989. Features of two hepatitis B virus (HBV) DNA integrations suggest mechanisms of HBV integration. *J. Virol.* **63**:2638-2643.
- Hino, O., T. B. Shows, and C. E. Rogler. 1986. Hepatitis B virus integration site in hepatocellular carcinoma at chromosome 17:18 translocation. *Proc. Natl. Acad. Sci. USA* **83**:8338-8342.
- Konopka, A. K. 1988. Compilation of DNA strand exchange sites for nonhomologous recombination in somatic cells. *Nucleic Acids Res.* **16**:1739-1758.
- 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.
- Mason, W. S., M. S. Halpern, J. Newbold, C. E. Rogler, K. L.

- Molnar-Kimber, and J. Summers. 1984. Molecular biology of the replication of hepatitis B viruses, p. 23-41. In P. W. J. Rigby and N. M. Wilkie (ed.), *Viruses and cancer*. Cambridge University Press, Cambridge.
30. Mason, W. S., and J. M. Taylor. 1989. Experimental systems for the study of hepadnavirus and hepatitis delta virus infections. *Hepatology* 9:635-645.
31. McCoubrey, W. K., and J. J. Champoux. 1986. The role of single-strand breaks in the catenation reaction catalyzed by the rat type I topoisomerase. *J. Biol. Chem.* 261:5130-5137.
32. Nagaya, T., T. Nakamura, T. Tokino, T. Tsurimoto, M. Imai, T. Mayumi, K. Kamino, K. Yamamura, and K. Matsubara. 1987. The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. *Genes Dev.* 1:773-782.
33. Nash, H. A., and C. A. Robertson. 1989. Heteroduplex substrates for bacteriophage lambda site-specific recombination: cleavage and strand transfer products. *EMBO J.* 8:3523-3533.
34. Nunes-Duby, S. E., L. Matsumoto, and A. Landy. 1987. Site specific recombination intermediates trapped in suicide substrates. *Cell* 50:779-788.
35. Ogston, 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. Porter, S. E., and J. J. Champoux. 1988. Mapping in vivo topoisomerase I sites on simian virus 40 DNA: asymmetric distribution of sites on replicating molecules. *Mol. Cell. Biol.* 9:541-550.
37. Radziwill, G., W. Tucker, and H. Schaller. 1990. Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. *J. Virol.* 64:613-620.
38. Raimondo, G., R. D. Burk, H. M. Lieberman, J. Muschel, S. J. Hadziyannis, H. Will, M. C. Kew, J. M. Dusheiko, and D. A. Shafritz. 1988. Interrupted replication of hepatitis B virus in liver tissue of HBsAg carriers with hepatocellular carcinoma. *Virology* 166:103-112.
39. Rogler, C. E., M. Sherman, C. Y. Su, and D. A. Shafritz. 1985. Deletion in chromosome 11p associated with a hepatitis B integration site in hepatocellular carcinoma. *Science* 230:319-322.
40. Rogler, C. E., and J. Summers. 1984. Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from a chronically infected liver. *J. Virol.* 50:832-837.
41. Sadowski, P. 1986. Site-specific recombinases: changing partners and doing the twist. *J. Bacteriol.* 165:341-347.
42. Sanger, F., S. Niklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
43. Shafritz, D. A., D. Shouval, H. I. Sherman, S. J. Hadziyannis, and M. C. Kew. 1981. Integration of hepatitis B virus DNA into the genome of liver cells in chronic liver disease and hepatocellular carcinoma. *N. Engl. J. Med.* 305:1067-1073.
44. Shaul, Y., M. Ziemer, P. D. Garcia, R. Crawford, H. Hsu, P. Valenzuela, and W. J. Rutter. 1984. Cloning and analysis of integrated hepatitis virus sequences from a human hepatoma cell line. *J. Virol.* 51:776-787.
45. Shih, C., K. Burke, M.-J. Chow, J. B. Zeldis, C.-S. Yang, C.-S. Lee, K. J. Isselbacher, J. R. Wands, and H. M. Goodman. 1987. Tight clustering of human hepatitis B virus integration sites in hepatomas near a triple-stranded region. *J. Virol.* 61:3491-3498.
46. Shuman, S. 1989. Vaccinia DNA topoisomerase I promotes illegitimate recombination in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 86:3489-3493.
47. Stewart, A. F., and G. Schutz. 1987. Camptothecin-induced "in vivo" topoisomerase I cleavage in the transcriptionally active tyrosine aminotransferase gene. *Cell* 50:1109-1117.
48. 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.
49. Summers, J., P. M. Smith, and A. L. Horwich. 1990. Hepadnavirus envelope proteins regulate covalently closed circular DNA amplification. *J. Virol.* 64:2819-2824.
50. Summers, J., J. Smolec, and R. Snyder. 1978. A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proc. Natl. Acad. Sci. USA* 75:4533-4537.
51. Thrash, C., K. Voelkel, S. DiNardo, and R. Sternglanz. 1984. Identification of *Saccharomyces cerevisiae* mutants deficient in DNA topoisomerase I activity. *J. Biol. Chem.* 259:1375-1377.
52. Tokino, T., S. Fukushige, T. Nakamura, T. Nagaya, T. Murotsu, K. Shiga, N. Aoki, and K. Matsubara. 1987. Chromosomal translocation and inverted duplication associated with integrated hepatitis B virus in hepatocellular carcinoma. *J. Virol.* 61:3848-3854.
53. Tsui, S., M. E. Anderson, and P. Tegtmeyer. 1989. Topoisomerase I sites cluster asymmetrically at the ends of the simian virus 40 core origin of replication. *J. Virol.* 63:5175-5183.
54. 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.
55. Tuttleman, J. S., J. C. Pugh, and J. W. Summers. 1986. In vitro experimental infection of primary duck hepatocyte cultures with duck hepatitis B virus. *J. Virol.* 58:17-25.
56. Wang, J. C. 1985. DNA topoisomerases. *Annu. Rev. Biochem.* 54:665-697.
57. Wu, T.-T., L. Coates, C. E. Aldvhi, 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.
58. Yaginuma, K., H. Kobayashi, M. Kobayashi, T. Morishima, K. Matsuyama, and K. Koike. 1987. Multiple integration site of hepatitis B virus DNA in hepatocellular carcinoma and chronic active hepatitis tissues from children. *J. Virol.* 61:1808-1813.
59. Yaginuma, K., M. Kobayashi, E. Yoshida, and K. Koike. 1985. Hepatitis B virus integration hepatocellular carcinoma DNA: duplication of cellular flanking sequences at the integration site. *Proc. Natl. Acad. Sci. USA* 82:4458-4462.
60. Yang, L., M. S. Wold, J. J. Li, T. J. Kelly, and L. F. Liu. 1987. Roles of DNA topoisomerases I in the transcription of human ribosomal RNA genes. *Cell* 85:1060-1064.
61. Yang, L., M. S. Wold, J. J. Li, T. J. Kelly, and L. F. Liu. 1987. Roles of DNA topoisomerase in simian virus 40 DNA replication in vitro. *Proc. Natl. Acad. Sci. USA* 84:950-954.