Molecular Analysis of the Function of Direct Repeats and a Polypurine Tract for Plus-Strand DNA Priming in Woodchuck Hepatitis Virus

CHRISTOPH SEEGER* AND JOHNNA MARAGOS

Department of Microbiology, Immunology and Parasitology, New York State College of Veterinary Medicine, Cornell University, Ithaca, New York 14853-6401

Received 8 November 1988/Accepted 23 January 1989

The replication of the hepadnavirus DNA genome is initiated by reverse transcription of pregenome RNA into minus-strand DNA followed by plus-strand DNA synthesis. The priming of plus-strand DNA requires the transfer of an RNA primer from pregenome RNA to the primer-binding site on minus-strand DNA. Annealing of the primer to the primer-binding site is facilitated by short direct repeats, DR1 and DR2. To investigate the mechanism of plus-strand primer formation, we have introduced specific mutations into DR1 and DR2 and measured the effect of these mutants on initiation of plus-strand DNA synthesis. To facilitate such an analysis, we have constructed a vector for the efficient expression of woodchuck hepatitis virus in cultured cells. Our results suggest that the 3' end of the RNA primer is determined prior to its transfer to the primer-binding site and that the determination of the 3' end of the primer does not depend on a specific sequence motif at the cleavage site. In addition, we have identified an alternative initiation site for plus-strand DNA synthesis at a purine-rich sequence between DR1 and DR2. Initiation at this site occurs by a mechanism that is independent of the direct repeats and does not require the transfer of an RNA primer to the primer-binding site.

Woodchuck hepatitis virus (WHV), a member of the hepadnavirus family, has a small 3.3-kilobase (kb)-long relaxed circular DNA genome that replicates by reverse transcription of an RNA template, the pregenome (for reviews, see references 11 and 19). Virion DNA is composed of a complete minus strand covalently linked to a protein at its 5' end and of an incomplete plus strand with heterogeneous 3' ends (12). The genome is held in a circular conformation by a cohesive overlap, 215 nucleotides in length, representing the distance between the 5' ends of plus- and minus-strand DNA (23). Upon infection, the relaxed circular virion DNA is converted into a covalently closed circular (CCC) molecule present in nuclei of virus-infected cells (18, 34; Fig. 1A). CCC DNA is the template for the transcription of a 3.6-kb terminally redundant RNA initiating near the 5' end of the precore gene and terminating within the core antigen region (8; Fig. 1A and B). This transcript has heterogeneous 5' ends bracketing the first AUG of the precore gene (8). The shortest transcript, initiating within the precore region, is the template for reverse transcription and is termed pregenome RNA (27). Reverse transcription of pregenome RNA into minus-strand DNA occurs in immature virus core particles present in the cytoplasm of infected cells (32).

The primer for minus-strand DNA synthesis is most likely a protein, the structure and origin of which have not yet been determined (12, 20). The 5' end of minus-strand DNA lies within the terminally redundant region of pregenome RNA at DR1 (16, 27). The primer for plus-strand DNA synthesis is an RNA oligomer which is translocated from the DR1 region of pregenome RNA to the priming site at DR2 (15, 27). Annealing of this 18-nucleotide-long primer is facilitated by the direct repeats, DR1 and DR2, 11 nucleotides in length (Fig. 1). Since the sequence motif of the RNA primer maps to the terminally redundant portion of pregenome RNA, the

origin of the primer cannot unambiguously be determined from the sequences of cloned viral genomes (8, 27). The observation that the 5' end of the RNA primer of duck hepatitis B virus (DHBV) is modified (capped) provides strong circumstantial evidence that the primer is derived from the 5' end of pregenome RNA (15). Further support for this model is derived from biochemical studies which indicate that the position of the 5' end of the RNA primer coincides with the position of the 3' end of minus-strand DNA (16, 27). The mechanism for the formation of the exact 3' end of this primer is not known. By analogy to retroviral systems, it has been speculated that the 3' end is formed by the action of an RNase H-like activity (27, 38).

In the past, models for the replication strategy of hepadnaviruses had mainly been derived from the physical analysis of replicative intermediates isolated from livers and sera of infected animals. Later, it became feasible to infect animals with cloned viral DNA, permitting a genetic analysis of the hepadnavirus replication scheme (26, 39). However, this system allows for only the genetic analysis of altered viral genomes that remain competent through all stages of the replication cycle. Most recently, it has become possible to propagate human hepatitis B virus (HBV) and DHBV in cell culture systems, providing a basis for a detailed genetic analysis of this virus class (1, 5, 21, 33, 36, 40). To facilitate such an analysis, we constructed a cytomegalovirus (CMV)-based expression vector for the efficient production of infectious WHV in cultured cells. An important feature of this system is the possibility to genetically manipulate the terminal repeats of pregenome RNA independently of each other. Using this methodology, we have investigated how the 3' end of the RNA primer for plus-strand DNA synthesis is determined. Specifically, we addressed the question of whether the 3' end of the RNA primer is determined prior to or after its transfer to DR2. Our results provide strong evidence that the 3' end of the RNA primer is determined before its transfer to DR2 and that the genesis of the 3' end

* Corresponding author.

FIG. 1. Transcriptional and translational map of pCMW82. (A) The top half of the figure shows the structure of the WHV virion DNA. The protein covalently linked to the 5' end of minus-strand DNA is indicated with the solid oblong symbol, and the RNA primer attached to the 5' end of the plus strand is represented by the wavy line (dashes imply variability at the 3' end of plus-strand DNA). The positions of the direct repeats, DR1 and DR2, are indicated with heavy lines. The bottom half of the figure shows the covalently closed circular (CCC) DNA and the 5' and 3' ends of pregenome RNA. R, Terminally redundant portion of pregenome RNA. The positions of DR1 and DR2 are indicated as described before. (B) The CMV-IE promoter-enhancer region is represented by the hatched box, the WHV sequences are represented by the dotted box, and the four WHV gene coding regions are represented by open boxes (PC, pre-core; S1, large surface antigen; S2, middle surface antigen; SUR, major surface antigen). The solid bar above the translational map of WHV marks the length of the CCC DNA, the natural template for transcription of pregenome RNA. DR1 and DR2 are symbolized by vertical bars. The 5' and 3' ends of pregenome RNA are symbolized by arrows. The restriction sites used for the construction of pCMW82 are indicated (for details, see Materials and Methods): S, *Sst*I; Sp, *Sph*I; H, *Hind*III. bp, Base pairs.

of this primer does not depend upon a specific sequence pattern at the cleavage site. Furthermore, we demonstrate that priming of plus-strand DNA of WHV can initiate at a purine-rich site by a mechanism independent of the DR1-DR2 sequence motif.

MATERIALS AND METHODS

Molecular clones. For the construction of pCMW8, the precursor plasmid of pCMW82, an *Sst*I-to-*Hind*III fragment

of plasmid pBC12/CMV/IL-2 (a gift from Bryan Cullen [7]) was replaced with an *Sph*I-to-*Hind*III fragment from pWHV81-2 (14, 28). The correct fusion between the CMV and WHV sequences was confirmed by nucleotide sequence analysis. Subsequently, a full-length genome of WHV clone WHV2 was released from pWHV81-2 by cleavage with *Hind*III and inserted into the *Hind*III site of pCMW8, yielding pCMW82. The mutants 82DR12, 82DR1C, and 82DR1/2C were obtained by oligomer-directed site-specific mutagenesis as described by Taylor et al. (35) by using a kit from Amersham Corp. The mutant genotypes were confirmed by nucleotide sequence analysis.

Cell culture. HepG2 cells (2) were cultured in minimal essential medium (GIBCO Laboratories) supplemented with 10% bovine calf serum (HyClone Laboratories, Inc.) in the presence of penicillin (90 U/ml) and streptomycin (90 µg/ml) at 37°C and 3% CO₂. The tissue culture medium was changed in 3- and 4-day intervals. The transfection of cells with plasmid DNA was accomplished by the calcium phosphate procedure as described by Chen and Okayama (6).

Isolation of virion DNA. Supernatant from transfected cells was collected in 3- and 4-day intervals over a period of 7 to 30 days after DNA transfection. The supernatant was cleared of cellular debris by centrifugation in a Sorvall SS34 rotor at 9 krpm for 10 min. The cleared supernatant (7.5 ml) was layered on 3 ml of a 10% sucrose and on 1 ml of a 20% sucrose solution in 20 mM Tris (pH 8)–150 mM NaCl. The virus was collected by centrifugation in a Beckman SW41 rotor for 6 h at 40 krpm or for 16 h at 33 krpm at 10°C. The supernatants were decanted, and the pellets were suspended in 100 µl of a solution containing 10 mM Tris (pH 8), 100 mM NaCl, and 0.1% β-mercaptoethanol for at least 6 h at 4°C. One-tenth volume of TM (100 mM Tris [pH 8], 100 mM MgCl₂) and 5 µg of DNase I (Bethesda Research Laboratories [BRL]) were added. After an incubation of 30 min at 37°C, 50 µl of a solution containing 3% sodium dodecyl sulfate, 2 mg of proteinase K (BRL) per ml, 40 mM EDTA (pH 8), and 200 µg of yeast tRNA (BRL) was added and the suspension was incubated for 2 to 3 h at 37°C. DNA was extracted twice with phenol followed by one extraction with a solution of 70% butanol and 30% isopropanol and ethanol precipitation in the presence of LiCl (1 M final concentration) at –70°C. Pellets were washed with 70% ethanol, suspended in 20 µl of TE (10 mM Tris [pH 8], 1 mM EDTA), and stored at –20°C.

Primer extension analysis. To 5 µl of a virion DNA preparation, 1 µl (200 pg) of 5'-end-labeled [³²P]DNA oligomer primer was added, and the mixture was denatured at 95°C for 2 min. Samples were chilled immediately on ice. The primer extension reaction was performed under the following conditions. Five microliters of the virion DNA-primer mixture was supplemented with dATP, dCTP, dGTP, and TTP (100 µM final concentration), 50 mM Tris (pH 8), 6 mM MgCl₂, 40 mM KCl, 1 mM dithiothreitol (Sigma Chemical Co.), 0.1 mg of acetylated bovine serum albumin (Promega Biotec) and 2 U of avian myeloblastosis virus reverse transcriptase (Seikagaku America Inc.) per ml. The final volume of the reaction was 10 µl. The reaction was incubated at 42°C for 60 min and stopped by the addition of 6 µl of formamide. Samples were denatured at 95°C for 3 min prior to electrophoresis through a 7% polyacrylamide–8 M urea sequencing gel. Gels were dried and exposed to Kodak XAR-5 film (Eastman Kodak Co.) with an intensifying screen for 24 to 72 h at –70°C. For the digestion of the RNA primer prior to the addition of avian myeloblastosis virus reverse transcriptase, RNase T₁ (BRL) or RNase CL3

(BRL) was added to the reaction mixture prior to the addition of avian myeloblastosis virus reverse transcriptase to yield a final concentration of 140 or 0.8 U/ μ l, respectively, and then samples were incubated for 30 min at 37°C. For treatment of virion DNA with sodium hydroxide, 10 μ l of purified virion DNA was incubated for 5 min at 95°C in the presence of 200 mM sodium hydroxide. Tris hydrochloride was added to a final concentration of 333 mM, and DNA was precipitated with ethanol at -70°C. Pellets were suspended in 10 μ l of TE. DNA oligomers (10 ng) were treated at their 5' ends with polynucleotide kinase at 0.9 U/ μ l (Pharmacia) in the presence of [γ -³²P]ATP (3×10^3 Ci/mmol; 1.7 μ M), 40 mM Tris hydrochloride (pH 7.4), 10 mM MgCl₂, 100 mM KCl, 5 mM dithiothreitol (Sigma) in a total volume of 10 μ l. The reaction was incubated for 30 min at 37°C and stopped by incubation for 10 min at 65°C. The mixture was diluted 10-fold with TE. The sizes of the primer-extended products were determined with the help of a sequencing ladder derived from M13 clone MW2sh containing a *Hind*III-*Sst*II fragment of WHV2 (14). Primer extension and sequencing reactions were primed with the same DNA oligomer.

DNA oligomers. Oligonucleotides for primer extension analysis correspond to position 3248 to 3229 (CS2) and to position 26 to 3316 (DR17P) on the WHV genome.

Nomenclature. To facilitate a comparison of the presented data with previously published results, the nucleotide sequence numbering of the WHV2 genome (14) was arranged to begin with the first ATG of the precore open reading frame (position 1931 [14]) in accord with the published sequences of HBV and ground squirrel hepatitis virus (25, 37).

RESULTS

Construction of an expression vector for the production of WHV in cultured cells. The production of HBV and DHBV in selected human hepatoma cell lines transfected with cloned viral DNA has previously been described (1, 5, 33, 40). To express WHV in cultured cells, we fused WHV sequences to the CMV immediate-early (CMV-IE) promoter-enhancer region (7, 31). The recombinant plasmid pCMW82 was designed to direct the synthesis of pregenome RNA, the template for reverse transcription (Fig. 1B). This strategy was based on two considerations important for a rapid and detailed genetic analysis of the WHV genome: (i) the efficient production and secretion of virus particles for a biochemical analysis of virion DNA and (ii) the manipulation of the terminally redundant ends of pregenome RNA independently of each other. We chose the CMV-IE promoter-enhancer complex because it drives very high levels of transcription and is relatively nonspecific for cell type (3). By deleting the WHV promoter and thereby reducing the length of the terminally redundant DNA template for transcription, it was possible to minimize the chance for homologous recombination of the redundant ends upon transfection of cultured cells with cloned DNA.

To investigate whether pCMW82 could direct the production of virion DNA, HepG2 cells were transfected with this construct and virion DNA was purified from the culture supernatant as described in Materials and Methods. The ability of this system to lead to correct synthesis of virion DNA was verified by mapping the 5' end of plus-strand DNA by primer extension analysis as shown schematically in Fig. 2A. Figure 2B shows a comparison of primer-extended products obtained from virion DNA isolated from an infected woodchuck, W1451 (lanes 4 and 7), and from virion

FIG. 2. Comparison of the 5' end of plus-strand DNA from a WHV-infected animal and from WHV produced in HepG2 cells. (A) The position of oligonucleotide CS2 used for primer extension analysis of the 5' end of virion DNA is indicated (position [POS.] 3248 to 3229). The position of the 5' end of plus-strand DNA (position 3120) and the sequence of the RNA primer of WHV are shown as determined previously (27). The lengths of the primer-extended products are indicated (NT). (B) Primer extension analysis was performed as described in Materials and Methods by using DNA oligomer CS2 as a primer. The lengths of the extended products are indicated. A nucleotide-sequencing ladder derived from M13 clone MWs, primed with oligomer CS2, provided size markers (lanes 1 to 3). Lanes 4 and 7 show the extended primer from reactions with virion DNA isolated from woodchuck 1451; lanes 5 and 6 show the extended primer from reactions with virion DNA isolated from HepG2 cells transfected with pCMW82 (lane 5) or with p82DR12, a derivative of pCMW82 (lane 6) (Fig. 3B). Lane 8 contains equal amounts of the reaction products shown in lanes 4 and 5. nt, Nucleotides.

DNA isolated from the supernatant of HepG2 cells previously transfected with pCMW82 (lane 5) or p82DR12 (lane 6), a derivative of pCMW82 (see below). The DNA products extended on virion DNA derived from tissue culture cells are 146 nucleotides in length, 1 nucleotide short of the products derived from virion DNA isolated from an infected woodchuck. Mixing the extended products from those two reactions prior to gel electrophoresis confirmed this conclusion (Fig. 2B, lane 8). The occurrence of doublets in the primer extension reactions is most likely a consequence of premature disengagement of the reverse transcriptase from the DNA template. Removal of the RNA primer with RNase CL3 prior to primer extension results in extension products of the same length (see Fig. 4B, lanes 3 and 4). This indicates that the position of the 5' ends of plus-strand DNA is

identical in virus propagated in tissue culture cells and in virus obtained from infected animals. We conclude from these results that the RNA primer produced in cultured cells is 1 nucleotide shorter (17 nucleotides in length) than the RNA primer made *in vivo* (27). On the basis of the distance between the TATA box and the cap site of the original CMV-IE transcript, the 5' end of pregenome RNA transcribed from the CMV-IE promoter maps to position 5 on the WHV genome, 1 nucleotide short of the transcript made by WHV *in vivo* (8, 27, 31). This result provides direct genetic evidence for the origin of the RNA primer for plus-strand DNA synthesis from the 5' end of pregenome RNA; the truncation of pregenome RNA by 1 nucleotide results in a 1-nucleotide-shorter RNA primer for plus-strand DNA synthesis. A comparison of the 5' ends of minus-strand DNA produced by pCMW82 in cultured cells with that of WHV made *in vivo* revealed no qualitative difference, further confirming the competence of this expression system to direct the synthesis of WHV in transfected cells. Therefore, we conclude that pCMW82 is a suitable vector for the genetic analysis of the mechanism for plus-strand DNA synthesis.

Is the 3' end of the RNA primer for plus-strand DNA synthesis determined prior to its transfer to DR2? Previous studies have shown that the 5' ends of plus-strand DNA map to position 3120, the first nucleotide 3' of DR2 (15, 27). The accurate initiation of plus-strand DNA thus depends upon the precise formation of the 3' end of the RNA primer. At least two different mechanisms can be envisaged which could lead to the formation of an exact 3' end of the RNA primer; models for these mechanisms are shown in Fig. 3A. (i) If the 3' end of the primer is determined before its transfer to DR2, cleavage could occur at DR1 by a sequence-specific RNase H-like endonuclease activity, followed by the transfer of the primer to DR2. (ii) Alternatively, if the 3' end is determined after the transfer, primers with heterogeneous 3' ends could anneal to DR2 and the exact 3' ends would be determined by a nuclease activity, in the simplest model an exonuclease, that trims the overhanging 3' ends to the position where the homology between DR1 and DR2 begins.

The observation that the sequence motif dCCT(T/A), with the first dC (deoxycytidine) residue at position 21 specifying the last nucleotide of DR1, is present in the sequenced genomes of mammalian and avian hepadnaviruses favors the first of the two models (10, 14, 17, 25, 30, 37). According to this model, an RNase H-mediated cleavage occurs between the two rC residues at position 21 and 22, respectively (Fig. 3B, line a). To test this hypothesis, we have changed the nucleotide sequence flanking DR1 in pCMW82, shown in Fig. 3B (line a), and measured its effect on plus-strand DNA priming. The sequence motif 5'-dCCTT-3' was moved in the 5' direction by 1 nucleotide. To assure proper annealing of the primer to DR2, the penultimate nucleotide of DR2 (position 3118) had to be changed from dG to dC. The DR1 and DR2 sequences of the resulting mutant p82DR1/2C are shown in Fig. 3B, line b. If the rCCUU sequence pattern served as a recognition sequence for an endonucleolytic cleavage, the length of the RNA primer should be reduced by 1 nucleotide at its 3' end, thus moving the initiation site for plus-strand DNA synthesis in the 5' direction by 1 nucleotide, from position 3120 to 3119.

HepG2 cells were transfected with p82DR1/2C, and virion DNA was purified from culture supernatants and subjected to primer extension analysis. To determine the exact position of the 5' end of plus-strand DNA, the RNA primer was removed by treatment with sodium hydroxide. As shown in

FIG. 3. Mutations for the determination of the mechanism for the creation of the 3' end of the RNA primer. (A) The figure shows two possible mechanisms for the formation of the 3' end of the RNA primer, depending on whether the 3' end is determined before (i) or after (ii) its transfer to DR2. The protein covalently attached to the 5' end of minus-strand DNA is indicated with the solid oval, and the 5' end of the RNA primer is tagged with the symbol 5'. The small arrows in lane 1 indicate random RNase H cleavage sites on pregenome RNA. The large arrow in part i, lane 1, marks the specific RNase H cleavage site for the determination of the 3' end of the RNA primer. (B) The nucleotide sequence of DR1 and DR2 are shown, and their positions are indicated as described in Materials and Methods. *, Mutations generated by site-directed mutagenesis: (a) the wild-type sequence of pCMW82; (b) p82DR1/2C with base changes at positions 20, 22, 24, and 3118 (a T-to-G change at position 24 was introduced to preserve the amino acid composition of the precore gene); (c) p82DR1C, which is the same as p82DR1/2C except for position 3118, which remained unchanged; (d) p82DR12, which has a single base change at position 3120 from dA to dC. +, Positions of the 5' ends of plus-strand DNA.

Fig. 4A (lanes 6 and 7), the length of the extended product from mutant p82DR1/2C is identical to that of the wild-type pCMW82. The authenticity of the mutant clone was confirmed by treatment of virion DNA with RNase T₁ prior to primer extension analysis (Fig. 4A, lanes 1 to 3). The change of the penultimate nucleotide of DR1 from dG to dC (position 20) removes an RNase T₁ recognition site and consequently leads to an extension product 2 nucleotides longer than the products obtained with wild-type DR1 sequence motifs (Fig. 3B). In summary, the introduced base changes do not affect the position of the 5' end of the plus-strand DNA, indicating that the 3' end of the RNA primer is created by a mechanism independent of the exact sequence at the 3' end of DR1.

A possible explanation of the foregoing result is that the 3' end of the RNA primer is formed after its transfer to DR2 at the last nucleotide where DR1 and DR2 are homologous. To test this possibility, the following experiment was designed. The homology between DR1 and DR2 was extended from 11 to 12 nucleotides at DR2 by introducing a single nucleotide change at position 3120, converting a dA to a dC residue (Fig. 3B, line d). If the 3' end of the RNA primer was defined

FIG. 4. Mapping the 5' ends of plus-strand DNA of mutants p82DR1/2C and p82DR12. Primer extension analysis was done as described in the legend to Fig. 2B, except that virion DNA was treated with either RNase T₁, sodium hydroxide, or RNase CL3 prior to the primer extension. (A) Reaction products obtained after treatment with RNase T₁ (lanes 1 to 3) and sodium hydroxide (lanes 6 to 8). Size markers were provided by the nucleotide sequence ladder shown in lanes 4 and 5. Other lanes show the extension products from virion DNA obtained from DNA transfections with pCMW82 (lanes 1 and 7), with p82DR12 (lanes 2 and 8), and with p82DR1/2C (lanes 3 and 6). (B) Primer extension products after treatment of virion DNA with RNase CL3 (lanes 3 to 5). Lanes show the reaction products from virion DNA from woodchuck W1451 (lane 3) and DNA transfection with pCMW82 (lane 4) and p82DR12 (lane 5). The nucleotide sequence ladder in lanes 1 and 2 provided size markers. The lengths of the extended products are indicated (nt, nucleotides).

by the relative length of the direct repeats, the 5' end of the plus-strand DNA would be shifted by 1 nucleotide in the 3' direction, initiating at position 3121 rather than at 3120. The analysis of virion DNA isolated from the supernatant of HepG2 cells transfected with p82DR12 was performed after removal of the RNA primer with either RNase C13 (Fig. 4B, lane 5) or sodium hydroxide (Fig. 4A, lane 8). The data show that in both cases the lengths of the extended products are identical with those obtained with virion DNA from CMW82 (Fig. 4B, lane 4, and Fig. 4A, lane 7). This result indicates that an extension of the homology between DR1 and DR2 has no effect on the initiation site of plus-strand DNA synthesis.

To confirm this result, we have constructed p82DR1C (Fig. 3B, line c). This mutant is identical with the previously described p82DR1/2C, except that the sequence pattern in DR2 remained unchanged. This mutant was assayed for its capacity to direct the synthesis of plus-strand DNA. If the RNA primer was determined prior to its transfer at DR1, it would not base pair properly at DR2 as a consequence of a nucleotide mismatch at the penultimate nucleotide between DR1 and DR2. In the absence of a nuclease activity that could remove the last 2 nucleotides of the primer, initiation of plus-strand DNA thus would be unlikely to occur. Transfection of HepG2 cells with p82DR1C led to the production of virion DNA as observed by dot blot analysis and agarose gel electrophoresis. However, primer extension analysis

with purified virion DNA did not reveal any extended products characteristic of initiation of plus-strand DNA at DR2 (data not shown). This observation confirms our previous result that the 3' end of the RNA primer is not modified after its transfer to DR2.

Thus far, we have demonstrated that the formation of the 3' end of the RNA primer is not dependent either on a distinct sequence motif at the 3' end of DR1 or on the extent of the homology between DR1 and DR2. In addition, our data indicate that correct annealing of the primer at its 3' end is a requirement for initiation of plus-strand DNA synthesis at DR2.

Identification of an alternative site for initiation of plus-strand DNA synthesis. The failure to map the 5' end of virion DNA from mutant p82DR1C could have resulted from the lack of plus-strand DNA synthesized or from the presence of a novel 5' end located 3' of the normal 5' end. In the latter case, primer extension analysis with the oligonucleotide CS2 used previously in the primer extension reaction would not have given any signal. Since dot blot analysis of virion DNA purified from the supernatant of cells transfected with p82DR1C revealed the presence of plus-strand DNA (data not shown), we attempted to identify the novel 5' end by primer extension with a different DNA oligomer (DR17P), derived from position 26 to 3316 (Fig. 5B). A comparison between primer-extended products from pCMW82 (Fig. 5A, lane 1) and mutant p82DR1C (lane 2) reveals the presence of

FIG. 5. Mapping the 5' end of plus-strand DNA of virion DNA from p82DR1C. (A) Primer extension was performed on virion DNA isolated from HepG2 cells transfected with pCMW82 (lane 1) and p82DR1C (lanes 2, 6, 7, and 8) as described in the legend to Fig. 4, except that DNA oligomer DR17P (position 26 to 3316) was used as a primer. Virion DNA was treated with RNase CL3 (lane 6), RNase T₁ (lane 7), or sodium hydroxide (lane 8) prior to primer extension. The sizes of the extended products were deduced from the sequencing ladder of M13 clone MWs shown in lanes 3 to 5. The sequencing reaction was primed with oligonucleotide DR17P. nt, Nucleotides. (B) Depiction of the primer extension results obtained as shown in panel A. The major extension products observed before and after removal of the RNA primer are lined up with the nucleotide sequence of WHV (14). Lane A shows the extended product after treatment of virion DNA with RNase T₁ or alkali, lane B shows the product after RNase CL3 treatment, and lane C shows the product without any treatment. Lane D indicates the length of the extended product characteristic for initiation of plus-strand DNA at DR2. The length of the extended products and the corresponding positions of their 3' ends on the WHV genome are indicated (NT, nucleotides; POS., position).

at least six additional bands 108 to 114 nucleotides in length with virion DNA from p82DR1C. This could indicate the presence of heterogeneous 5' ends of plus-strand DNA mapping to position 3233 to 3239, approximately 110 nucleotides 3' of the original 5' end at DR2. The bands routinely

observed in primer extension analysis of virion DNA that has not been treated with ribonuclease or alkali most likely represent extension products from residual pregenome RNA species present in virion DNA preparations. Upon treatment of virion DNA with RNase CL3, RNase T₁, or sodium hydroxide prior to primer extension (lanes 6 to 8, respectively), the sizes of the extended products were reduced to 97 and 95 nucleotides depending on the treatment. A summary of these results is shown in Fig. 5B.

On the basis of the results obtained after treatment of virion DNA either with RNase T₁ or with alkali, the 5' end of the plus DNA strand starts with dC at position 3252 on the WHV map. Since this dC residue is preceded by a G (guanosine), treatment with RNase T₁ and alkali showed the same band, provided that the origin of the RNA primer coincides with the priming site. This assumption is directly confirmed by the pattern obtained upon digestion with RNase CL3. This ribonuclease cleaves after the rC residues at position 3249 and hence leaves the two terminal ribonucleotides rArG of the RNA primer intact, resulting in a primer extension product 2 nucleotides longer (lane 6) than the reaction products obtained after treatment with RNase T₁ and alkali (lanes 7 and 8, respectively). On the basis of the most prominent band corresponding to position 3239, the RNA primer is 13 nucleotides in length. However, longer primers up to 19 nucleotides in length could be present, as judged from additional bands observed in lane 2. The nucleotide sequence of the RNA primer is 5'-GGGAGGAGG GCAG-3'. This purine-rich sequence motif is reminiscent of polypurine tracts (PPT) found at the priming sites for plus-strand DNA synthesis in retroviruses and other retroviral elements (22, 29, 38). Evidence based upon in vitro experiments suggests that in retroviruses polypurine sequences cannot be cleaved internally by RNase H or other ribonucleases and thus can serve as primers for DNA synthesis (29). In summary, our results suggest that a PPT-like sequence at position 3239 to 3251 can substitute for a defective priming site for plus-strand DNA synthesis at DR2 and can direct the synthesis of plus-strand DNA with a discrete 5' end at position 3252. Initiation of plus-strand DNA at this PPT reduces the length of the cohesive overlap region in virion DNA from p82DR1C from 215 to 83 nucleotides.

Initiation of plus-strand DNA synthesis from polypurine tracts. We have shown that priming of plus-strand DNA at the PPT occurs in a mutant that is defective for priming of plus-strand DNA at DR2. To investigate whether the PPT can serve as a priming site in the presence of a functional DR1-DR2 sequence motif, primer extension was performed on virion DNA isolated from a chronically infected woodchuck, W1451. As evident from Fig. 6 (lane 7), the vast majority of extended products are 227 nucleotides long, positioning the 5' end of plus-strand DNA to position 3120 (27), adjacent to DR2, as observed with p82DR1/2C (lane 6). However, a minor fraction (about 1%) of extension products are shorter, 97 or 95 nucleotides in length, depending on whether RNase CL3 (lane 7) or RNase T₁ (lane 8) was used for the removal of the RNA oligomer on plus-strand DNA. Identical extension products are obtained upon RNase CL3 digestion of virion DNA from woodchuck W1451 and from virion DNA from HepG2 cells transfected with p82DR1C (Fig. 6, lanes 5 and 7). Thus, the 5' end of the shorter plus-strand DNA cluster coincides with the position of the 5' end of plus-strand DNA mapped from p82DR1/C. These results demonstrate that even in wild-type virus, initiation of plus-strand DNA synthesis can occur at the PPT, albeit at a low level compared with initiation at DR2.

FIG. 6. Initiation of plus-strand DNA at the PPT. Primer extension analysis was performed as described in the legend to Fig. 5. Virion DNA was purified from the supernatant of HepG2 cells transfected with p82DR1C (lane 5) or p82DR1/2C (lane 6) and from virion DNA isolated from woodchuck W1451 (lanes 7 and 8). Virion DNA was treated with either RNase CL3 (lanes 5 to 7) or RNase T₁ (lane 8) prior to the primer extension. Lanes 1 to 4, A, C, G, and T tracks of the sequencing reaction performed with clone MWs by using oligomer DR17P as a primer. The lengths of the extended products are indicated. To visualize the 227-nucleotide (nt)-long extension product in lane 6 the portion of the dried gel containing the corresponding extension products in lanes 7 and 8 was physically separated from its original position after initial exposure.

DISCUSSION

Efficient expression of WHV in HepG2 cells. The development of cell culture systems for the expression of HBV from cloned DNA created an exciting path for a detailed genetic analysis of the hepadnavirus genome (1, 5, 13, 21, 33, 40). Genetically altered genomes can be assayed in a transient expression system for their ability to execute individual steps important for replication. A requirement for such a system is that sufficient virus is secreted into the culture medium to allow for a biochemical analysis of virion DNA without extensive purification steps.

The synthesis of pregenome RNA, the template for reverse transcription, occurs from a closed circular molecule and results in a RNA with terminally redundant ends. This terminally redundant portion contains several important functions for replication; among them are the initiation site for minus-strand DNA synthesis, the origin of the RNA primer for plus-strand DNA synthesis, and most likely signals important for packaging of the viral genome (9, 27). Therefore, we attempted to resolve this portion of pregenome RNA on the DNA level, in order to be able to

manipulate one element independently of the other. We have fused WHV sequences to the CMV-IE promoter unit such that the synthesis of WHV pregenome RNA is under the control of this regulatory element. The 3.6-kb RNA of hepadnaviruses has heterogeneous 5' ends, separated by approximately 30 nucleotides, bracketing the precore AUG codon (8). Previous reports have demonstrated that the shortest RNA species initiating 3' to the precore AUG is packaged into viral core particles and thus serves as the template for reverse transcription (9, 27). Furthermore, experiments conducted with DHBV showed that a functional precore open reading frame is not required for virus replication and infectivity (4, 24). Therefore, we designed the CMV-IE-WHV vector for the expression of a WHV transcript corresponding to pregenome RNA.

Our results demonstrate that sufficient virion DNA can be obtained from as little as 3 ml of culture medium for the biochemical analysis of plus-strand DNA. Comparison of the viral DNA produced *in vitro* with virion DNA produced in the animal revealed identical initiation sites of plus- and minus-strand DNA (Fig. 4B; J. Maragos and C. Seeger, unpublished results). In contrast, the RNA primer for plus-strand DNA synthesis made *in vitro* is shorter by 1 nucleotide than the primer made *in vivo*. This is the consequence of a 1-nucleotide-shorter pregenome RNA made from the CMV-IE promoter (Fig. 2B) compared with pregenome RNA made from the WHV promoter. This observation provides direct genetic evidence for the origin of the RNA primer from the 5' end of pregenome RNA and is in accord with previous results obtained from the biochemical and genetic analysis of this RNA primer (15, 16, 27). In agreement with this result is the observation that a mutation introduced into the 5' copy of DR1 (p82DR1C) can interfere with the priming of plus-strand DNA at DR2. Inoculation of woodchucks with culture supernatant from pCMW82-transfected HepG2 cells, as well as some cell lines of nonhepatic origin, can lead to productive WHV infection in these animals (J. Maragos, B. Baldwin, B. C. Tennant, and C. Seeger, unpublished results), emphasizing the utility of this system for a careful genetic analysis of the hepadnavirus replication strategy *in vitro* and *in vivo*.

Determination of the 3' end of the RNA primer. The primary objective of this study was to investigate the mechanism by which the RNA primer for plus-strand DNA synthesis of hepadnaviruses is generated; in particular, we addressed the question of whether the primer is determined prior to or after its transfer to DR2. In resemblance to the retrovirus model (38), it would seem likely that the 3' end of the RNA primer is created by an RNase H-like activity. However, unlike the sequence of retroviruses and other reovirus elements, the nucleotide sequence of the hepadnavirus RNA primer does not include the PPT involved in retrovirus primer formation. Instead, a sequence pattern, 5'-CCU(U/A)-3', flanking the cleavage site at the 3' end of the RNA primer is conserved in hepadnaviruses. We tested the hypothesis that this sequence motif serves as a recognition site for RNase H, by shifting this pattern by 1 nucleotide in the 5' direction. This change did not affect the relative length of the RNA primer, implying that cleavage occurred at the same position but between two different nucleotides. This result is not consistent with the hypothesis that the conserved sequence pattern 5'-CCUU-3' governs the determination of the exact 3' end of the RNA primer. Therefore, we tested whether the exact 3' end of the primer could be generated after its transfer to DR2, perhaps by an exonuclease-like activity. The observation that the extension of

the homology between the two repeats did not affect the position of the 3' end of the RNA primer strongly argues against the involvement of such an activity in primer formation. This result was confirmed with mutant p82DR1C (Fig. 3B), in which a mismatch at the penultimate nucleotide between the primer and the primer-binding site abolished priming from DR2. We conclude that proper annealing of the primer at its 3' end is mandatory for the priming event and that the virus lacks the necessary enzymatic activities to repair such a mismatch. In this context, it is noteworthy that a mutation changing nucleotide 4 from the 3' end of DR1 (position 18) does not impede priming at DR2 (27).

The data presented here support a model where the 3' end of the RNA primer is determined by an endonucleolytic ribonuclease activity prior to its transfer to DR2. It appears that the cleavage site for the creation of the 3' end does not coincide with the actual recognition site of the ribonuclease engaged in this process. It is conceivable that sequences 5' or 3' to the cleavage site serve as recognition sites for the ribonuclease. A possible recognition sequence could be the DR1 sequence motif itself. However, recent observations in our laboratory indicate that substantial nucleotide substitutions in DR1 do not interfere with the correct cleavage of the primer (A. Glaser and C. Seeger, unpublished results). These observations would argue against an involvement of the DR1 sequence pattern in primer formation. While the presented models consider the primary nucleotide sequence as the determining factor for primer formation, an alternative possibility is that the secondary structure of the RNA in the core particle is the determining factor. In this model, the cleavage site would be located in an accessible position for recognition and cleavage by RNase H and cleavage would be independent of a primary nucleotide sequence pattern.

A polypurine tract can serve as an additional plus-strand DNA priming site. An unexpected finding of our investigation is that plus-strand DNA synthesis can initiate at a purine-rich sequence independently of the DR1-DR2 priming mechanism. We have observed priming at this site in a mutant (p82DR1C) defective for initiation of plus-strand DNA at DR2, as well as in virion DNA isolated from an infected animal. In the latter DNA, about 1% of plus-strand DNA initiated at the PPT. Thus the pool of WHV DNA present in sera of infected animals consists of at least two populations, which differ at the 5' ends of plus-strand DNA. Since similar PPTs are present on the genomes of the closely related ground squirrel hepatitis virus and HBV, it is likely that this observation can be generalized for the mammalian hepadnaviruses (25, 37). Since the PPT maps to the cohesive overlap region of plus- and minus-strand DNA, virion DNA with plus-strand DNA initiating at this site should still form a relaxed circle and could be converted to CCC DNA upon infection (Fig. 7). Thus far, we have identified one PPT as an initiation site for DNA replication. However, scanning of the hepadnavirus genomes for similar sequence motifs revealed several additional sites scattered over the viral genomes, any of which could serve as a target for plus-strand DNA priming (10, 17, 25, 37).

The formation of the RNA primer by RNase H and the priming of plus-strand DNA at DR2 can occur only after the completion of minus-strand DNA synthesis. This is a consequence of the identity of the RNA primer with the 5' end of pregenome RNA. This mechanism differs from the strategy employed by other retroviral elements in which the origin of the RNA primer for plus-strand DNA synthesis coincides with the priming site and in which plus-strand DNA synthesis commences prior to completion of minus-strand DNA

FIG. 7. Initiation of plus-strand DNA at DR2 and at the PPT. The positions of the 5' ends of plus-strand DNA initiating at DR2 (position 3120) and at the PPT (position 3252) are indicated. The position of the 5' end of minus-strand DNA at position 14 is marked with the solid oblong symbol (27). Line A depicts the condition where priming of plus-strand DNA at DR2 is blocked (symbolized by dr2), and line B shows initiation of plus-strand DNA in the presence of a functional DR2 at both DR2 and the PPT.

synthesis (22, 29, 38). An important consequence of the priming event at the PPT in hepadnaviruses is that plus- and minus-strand DNA synthesis could occur simultaneously. It is not clear why only a minor fraction of plus strands initiates at the PPT. It is conceivable that the formation of the RNA primer is inefficient or, alternatively, that even though a primer is available, steric constraints prevent efficient initiation at this site. It will be interesting to determine whether the WHV with plus strands initiating exclusively at the PPT is infectious in the animal. Infection of woodchucks with virus originating from tissue culture cells transfected with the mutant p82DR1C should provide an answer to this question.

ACKNOWLEDGMENTS

We thank Bryan Cullen for the generous gift of plasmid pBC/CMV/IL2; Mary Ann Sells, Peter Price, and George Acs for help with the maintenance of HepG2 cells; Bud Tennant for woodchuck serum samples; William Mason, John Taylor, and Volker Vogt for helpful discussions and encouragement; and Jim Casey and Volker Vogt for valuable comments to the manuscript.

This work was supported by Public Health Service grant AI-24972 from the National Institutes of Health.

LITERATURE CITED

1. Acs, G., M. A. Sells, R. H. Purcell, P. Price, R. Engle, M. Shapiro, and H. Popper. 1987. Hepatitis B virus produced by transfected Hep G2 cells causes hepatitis in chimpanzees. *Proc. Natl. Acad. Sci. USA* **84**:4641-4644.
2. Aden, D. P., A. Fogel, S. Plotkin, I. Damjanov, and B. Knowles. 1979. Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature (London)* **282**:615-616.
3. Boshart, M., F. Weber, G. Jahn, K. Dorsch-Haesler, B. Fleckenstein, and W. Schaffner. 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* **41**:521-530.
4. Chang, C., G. Enders, R. Sprengel, N. Peters, H. E. Varmus, and D. Ganem. 1987. Expression of the precore region of an avian hepatitis B virus is not required for viral replication. *J. Virol.* **61**:3322-3325.
5. Chang, C. M., K. S. Jeng, C. P. Hu, S. J. Lo, T. Su, L.-P. Ting, C. K. Chou, S. Han, E. Pfaff, J. Salfeld, and H. Schaller. 1987. Production of hepatitis B virus in vitro by transient expression of cloned HBV DNA in a hepatoma cell line. *EMBO J.* **6**:675-680.
6. Chen, C., and H. Okayama. 1987. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**:2745-2752.
7. Cullen, B. R. 1987. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* **46**:972-982.

8. Enders, G. H., D. Ganem, and H. E. Varmus. 1985. Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. *Cell* **42**:297-308.
9. Enders, G. H., D. Ganem, and H. E. Varmus. 1987. 5'-Terminal sequences influence the segregation of ground squirrel hepatitis virus RNA into polyribosomes and viral core particles. *J. Virol.* **61**:35-41.
10. 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.
11. Ganem, D., and H. E. Varmus. 1987. The molecular biology of the hepatitis B viruses. *Annu. Rev. Biochem.* **56**:651-693.
12. Gerlich, W. H., and W. S. Robinson. 1980. Hepatitis B virus contains protein attached to the 5' terminus of its complete DNA strand. *Cell* **21**:801-809.
13. Junker, M., P. Galle, and H. Schaller. 1987. Expression and replication of the hepatitis B virus genome under foreign promoter control. *Nucleic Acids Res.* **15**:10117-10132.
14. Kodama, K., N. Ogasawara, H. Yoshikawa, and S. Murakami. 1985. Nucleotide sequence of a cloned woodchuck hepatitis virus genome: evolutionary relationship between hepadnaviruses. *J. Virol.* **56**:978-986.
15. Lien, J.-M., C. E. Aldrich, and W. S. Mason. 1986. Evidence that a capped oligoribonucleotide is the primer for duck hepatitis B virus plus-strand DNA synthesis. *J. Virol.* **57**:229-236.
16. Lien, J.-M., D. J. Petcu, C. E. Aldrich, and W. S. Mason. 1987. Initiation and termination of duck hepatitis B virus DNA synthesis during virus maturation. *J. Virol.* **61**:3832-3840.
17. 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.
18. Mason, W. S., M. S. Halpern, J. M. England, G. Seal, J. Egan, L. Coates, C. Aldrich, and J. Summers. 1983. Experimental transmission of duck hepatitis B virus. *Virology* **131**:375-384.
19. Mason, W. S., J. M. Taylor, and R. Hull. 1987. Retrovirus genome replication. *Adv. Virus Res.* **32**:35-96.
20. Molnar-Kimber, K. L., J. W. Summers, J. M. Taylor, and W. S. Mason. 1983. Protein covalently bound to minus-strand DNA intermediates of duck hepatitis B virus. *J. Virol.* **45**:165-172.
21. Pugh, J. C., K. Yaginuma, K. Koike, and J. Summers. 1988. Duck hepatitis B virus (DHBV) particles produced by transient expression of DHBV DNA in a human hepatoma cell line are infectious in vitro. *J. Virol.* **62**:3513-3616.
22. Resnick, R., C. A. Omer, and A. J. Faras. 1984. Involvement of retrovirus reverse transcriptase-associated RNase H in the initiation of strong-stop (+) DNA synthesis and the generation of the long terminal repeat. *J. Virol.* **51**:813-821.
23. Sattler, F., and W. S. Robinson. 1979. Hepatitis B DNA has cohesive ends. *J. Virol.* **32**:226-233.
24. Schlicht, H. J., J. Salfeld, and H. Schaller. 1987. The duck hepatitis B virus pre-C region encodes a signal sequence which is essential for synthesis and secretion of processed core proteins but not for virus formation. *J. Virol.* **61**:3701-3709.
25. Seeger, C., D. Ganem, and H. E. Varmus. 1984. The nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus. *J. Virol.* **51**:367-375.
26. Seeger, C., D. Ganem, and H. E. Varmus. 1984. The cloned genome of ground squirrel hepatitis virus is infectious in the animal. *Proc. Natl. Acad. Sci. USA* **81**:5849-5852.
27. Seeger, C., D. Ganem, and H. E. Varmus. 1986. Biochemical and genetic evidence for the hepatitis B virus replication strategy. *Science* **232**:477-484.
28. Seeger, C., D. Ganem, and H. E. Varmus. 1987. In vitro recombinants of ground squirrel and woodchuck hepatitis viral DNA produce infectious virus in squirrels. *J. Virol.* **61**:3241-3247.
29. Smith, J. K., A. Cywinski, and J. M. Taylor. 1984. Specificity of initiation of plus-strand DNA by Rous sarcoma virus. *J. Virol.* **52**:314-319.
30. Sprengel, 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.
31. Stenberg, R. M., D. R. Thomsen, and M. F. Stinski. 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. *J. Virol.* **49**:190-199.
32. 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.
33. Sureau, C., J. Romet-Lemonne, J. I. Mullins, and M. Essex. 1987. Production of hepatitis B virus by a differentiated human hepatoma cell line after transfection with cloned circular HBV DNA. *Cell* **47**:37-47.
34. Tagawa, M., M. Omata, and K. Okuda. 1986. Appearance of viral RNA transcripts in the early stage of duck hepatitis B virus infection. *Virology* **152**:477-482.
35. Taylor, J. W., J. Ott, and F. Eckstein. 1985. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucleic Acids Res.* **13**:8765-8784.
36. Tuttleman, J. S., J. C. Pugh, and J. Summers. 1986. In vitro experimental infection of primary duck hepatocyte cultures with duck hepatitis B virus. *J. Virol.* **58**:17-25.
37. Valenzuela, P., M. Quiroga, J. Zaldivar, P. Gray, and W. J. Rutter. 1980. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes, p. 57-70. *In* B. Fields, R. Jaenisch, and C. F. Fox (ed.), *Animal virus genetics*. Academic Press, Inc., New York.
38. Varmus, H. E., and R. Swanstrom. 1985. Replication of retroviruses, p. 75-134. *In* R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (ed.), *RNA tumor viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. Will, H., R. Cattaneo, H. Koch, G. Darai, H. Schaller, H. Schellekens, P. Van Eerd, and S. Deinhardt. 1982. Cloned HBV DNA causes hepatitis in chimpanzees. *Nature (London)* **299**:740-742.
40. Yaginuma, K., Y. Shirakata, M. Kobayashi, and K. Koike. 1987. Hepatitis B virus (HBV) particles are produced in a cell culture system by transient expression of transfected HBV DNA. *Proc. Natl. Acad. Sci. USA* **84**:2678-2682.