32. P. H. Nixon and F. R. Boyd, in *Lesotho Kimberlites*, P. H. Nixon, Ed. (Lesotho National Development Corporation, Maseru, Lesotho, 1973), pp. 67–75.
33. J. J. Gurney, W. R. O. Jakob, J. B. Dawson, in *Proceedings of the 2nd International Kimberlite Conference*, F. R. Boyd and H. O. A. Meyer, Eds. (American Geophysical Union, Washington, DC, 1979), vol. 2, pp. 227–243.
34. R. V. Danchin and F. R. Boyd, *Carnegie Inst. Washington Yearb.* **75**, 531 (1976).
35. M. J. O'Hara, M. J. Saunders, E. L. P. Mercy, *Phys. Chem. Earth* **9**, 571 (1975).
36. R. G. Coleman, *Ophiolites* (Springer-Verlag, Berlin, 1977).
37. H. J. B. Dick and R. L. Fisher, in *Kimberlites II: The Mantle and Crust-Mantle Relationships*, J. Kornprobst, Ed. (Elsevier, Amsterdam, 1984), pp. 295–308.
38. T. H. Jordan, in *Proceedings of the 2nd International Kimberlite Conference*, F. R. Boyd and H. O. A. Meyer, Eds. (American Geophysical Union, Washington, DC, 1979), vol. 2, pp. 1–14.
39. H. G. Wilshire *et al.*, *U.S. Geol. Surv. Open-File Rep. 85-139* (1985).
40. K. G. Cox, J. J. Gurney, B. Harte, in *Lesotho Kimberlites*, P. H. Nixon, Ed. (Lesotho National Development Corporation, Maseru, Lesotho, 1973), pp. 76–98.
41. A. P. Jones, J. V. Smith, J. B. Dawson, *J. Geol.* **90**, 435 (1982).
42. P. H. Nixon, F. R. Boyd, N. Z. Boctor, *Trans. Geol. Soc. S. Afr.* **86**, 221 (1983).
43. C. S. Kennedy and G. C. Kennedy, *J. Geophys. Res.* **81**, 2467 (1976).
44. F. R. Boyd, *Mineral. Soc. Am. Spec. Pap.* **3**, 63 (1970).
45. J. J. Gurney, unpublished data.
46. We have benefited from helpful remarks on this manuscript by P. H. Nixon and H. S. Yoder, Jr. We thank J. B. Hawthorne for his assistance in arranging field studies. J.J.G. acknowledges financial support for research from DeBeers Consolidated Mines, Ltd. Research by F.R.B. was supported under NSF grants EAR-7924567, 8120832, 8418447, and 8417437.

Research Articles

Biochemical and Genetic Evidence for the Hepatitis B Virus Replication Strategy

CHRISTOPH SEEGER, DON GANEM, HAROLD E. VARMUS

Hepatitis B viruses synthesize their open circular DNA genomes by reverse transcription of an RNA intermediate. The details of this process have been examined with the use of mammalian hepatitis B viruses to map the sites for initiation and termination of DNA synthesis and to explore the consequences of mutations introduced at short, separated direct repeats (DR1 and DR2) implicated in the mechanisms of initiation. The first DNA strand to be synthesized is initiated within DR1, apparently by a protein primer, and the completed strand has a short terminal redundancy. In contrast, the second DNA strand begins with the sequence adjacent to DR2, but its 5′ end is joined to an oligoribonucleotide that contains DR1; thus the putative RNA primer has been transposed to the position of DR2. It is now possible to propose a detailed strategy for reverse transcription by hepatitis B viruses that can be instructively compared with that used by retroviruses.

RNA-DIRECTED DNA SYNTHESIS, FIRST DESCRIBED FOR retroviruses (*1, 2*), is now recognized as the probable means for transfer of genetic information in various other settings, such as the replication of hepatitis B (*3*) and cauliflower mosaic viruses (*4*); the transposition of some eukaryotic mobile elements, including the yeast Ty elements (*5–7*) and the *copia*-like elements of *Drosophila* (*8*); and the generation of several repeated components of eukaryotic genomes, including processed pseudogenes (*9, 10*), some repeated dispersed sequences (*11*), and possibly the abundant Alu-like repeats (*10*).

The synthesis of double-stranded DNA from a single-stranded RNA template requires (i) an enzyme for synthesizing the first DNA strand from an RNA template and the second DNA strand from the first; (ii) primers for each of the two strands; and (iii) a means for removing the RNA template after reverse transcription, to allow synthesis of the second DNA strand. In some cases, strategies for duplicating sequences are required. For example, retrotransposons must not only perpetuate all of the single copy sequences present in the RNA template, but also generate two copies of sequences present once in the RNA to form long terminal repeats (LTR's) in DNA (*2, 5, 12*).

For retroviruses, the cardinal features of reverse transcription have been elucidated: (i) the first strand is primed by a host transfer RNA (tRNA) base-paired to the viral RNA genome near its 5′ end; (ii) the second strand is primed by a viral RNA oligomer that is produced from a polypurine region by a ribonuclease (RNase) activity associated with reverse transcriptase and specific for RNA-DNA hybrids (RNase H); (iii) a terminal redundancy (R) in viral RNA and complementarity between the ends of nascent strands facilitate transfer of nascent strands twice between templates, without loss of genetic material; and (iv) the finished product, a linear DNA duplex, displays LTR's composed mainly of sequences present only once in viral RNA (*2*). Structural analyses of transposable elements of yeast and *Drosophila* suggest that their replication strategies are likely to be fundamentally similar to those used by retroviruses (*5, 6, 13*).

The hepatitis B viruses (hepadnaviruses) differ fundamentally from retroviruses in that the form of the genome present in mature virus particles is DNA rather than RNA. These strongly hepatotropic agents have now been isolated from man, woodchucks, ground squirrels, and ducks, and they are important causes of liver disease in man and woodchucks (*14*). The discovery by Summers and Mason and colleagues that hepadnaviruses replicate by reverse transcription of RNA intermediates was based on the analysis of virus-specific nucleic acids and subviral particles from infected duck livers (*3, 15*). In contrast to the products of semiconservative DNA replication, hepatitis B virus DNA is asymmetric, with the strand complementary to viral RNA (the minus strand) longer and more abundant than the plus strand (*15, 16*). Isolated particles synthesize minus strand DNA in the presence of actinomycin D, and the product of

C. Seeger is a postdoctoral fellow in the Department of Microbiology and Immunology, D. Ganem is an assistant professor in the Departments of Medicine and Microbiology and Immunology, and H. E. Varmus is a professor in the Departments of Microbiology and Immunology and Biochemistry and Biophysics, University of California, San Francisco, CA 94143.

the reaction is an RNA-DNA hybrid (*3*). These findings and others imply that reverse transcription of hepadnavirus RNA generates minus strands, with plus strands produced from templates of minus strand DNA. In support of this proposal, the predicted amino acid sequence of an unassigned long open reading frame in hepadnavirus genomes displays significant homology with the sequences of retroviral reverse transcriptases (*17*).

Several features of hepadnaviruses relevant to studies of the replication cycle are illustrated in Fig. 1A, with the ground squirrel hepatitis virus (GSHV) as an example (*18*). The genome is composed of two open strands of fixed polarity, a full-length minus strand and an incomplete plus strand; the two strands form an open circle maintained by base-pairing between overlapping 5′ ends. The 5′ end of the minus strand is linked to an uncharacterized protein (*19*) that appears likely to serve as its primer (*20*). As previously shown for duck hepatitis B virus (DHBV) (*21*) and below for mammalian hepadnaviruses, the 5′ ends of both strands are positioned near short (10 to 12 nucleotides) direct repeats known as DR1 and DR2. Two major classes of RNA transcripts are found in infected cells (*22, 23*): a subgenomic class that serves as messenger RNA (mRNA) for viral surface antigens and larger RNA's that include the putative mRNA for viral core protein and the template for DNA synthesis (also called "pregenomic RNA"). The latter species contains a terminal redundancy (R) that encompasses DR1 (*22, 23*).

In addition to virion DNA and virus-specific RNA, two other important forms of hepadnaviral nucleic acid have been detected in infected liver cells and incorporated into a tentative replication cycle for these viruses (Fig. 1B): (i) superhelical, closed circular duplex DNA (*16, 24*), believed to arise directly from virion DNA (*25, 26*) and proposed to serve as template for synthesis of viral RNA by host RNA polymerase II; and (ii) incomplete forms of viral DNA, composed of protein-linked minus strands alone or paired with nascent plus strands. The latter forms, in conjunction with RNA-DNA hybrids (*3*), may represent intermediates in DNA synthesis from the pregenomic RNA template.

A detailed model for the replication cycle requires better definition of the 5′ and 3′ ends of the critical strands of viral nucleic acid and deeper insight into the roles of the direct repeats, DR1 and DR2. Accordingly, we have supplemented our earlier studies of pregenomic GSHV RNA (*23*) with precise physical mapping of the 5′ ends of plus and minus strands and the 3′ end of minus strand DNA from two mammalian hepadnaviruses, and we have constructed mutant GSHV genomes with altered DR1 and DR2 sequences. The results specify the priming sites for the two strands (the minus strand beginning within DR1, the plus strand at the boundary of DR2); reveal a short terminal redundancy in minus strand DNA; prove that differences between DR1 and DR2 are tolerated and not resolved by the replication machinery; and confirm that the oligoribonucleotide primer for the second strand at the DR2 position contains a translocated DR1 sequence (*27*). These findings allow construction of a more informative model for synthesis of hepatitis B virus DNA that can be provocatively contrasted with the strategy for synthesis of retroviral DNA.

The 5′ and 3′ ends of minus strand DNA. The initial steps in reverse transcription are the priming and extension of minus strand DNA. The 5′ end of the minus strand of DHBV DNA has been mapped to a site within DR1 (*21*), a region of the hepadnavirus genome that is represented twice in pregenomic RNA since it resides within R, the terminally repeated sequence (Fig. 1A) (*22, 23*). This means that an entire complement of genomic sequence, or slightly more than one complement, could be transcribed into minus strand DNA without transfer of nascent DNA between templates.

To determine the precise termini of minus strand DNA in the genomes of the mammalian hepadnaviruses, GSHV (*18*) and woodchuck hepatitis virus (WHV) (*28, 29*), we performed primer extension and S1 nuclease protection studies on deproteinized DNA extracted from virions. For identification of the 5′ terminus we used a synthetic oligomer, of plus polarity and 20 nucleotides in length, representing a region of the GSHV genome [positions 3220 to 3239, see Fig. 2A and (*18*)] that is identical in the closely related WHV genome (*29*). Two major extension products, 105 to 106 nucleotides in length, were observed after priming synthesis with the Klenow fragment of *Escherichia coli* DNA polymerase I on either GSHV or WHV DNA (Fig. 2B, lanes 3 and 4). The sequence at the 5′ end of the minus strand was deduced precisely by comparing the position of the bands in lanes 3 and 4 with the positions of bands in a sequencing ladder; the ladder was generated by use of the same

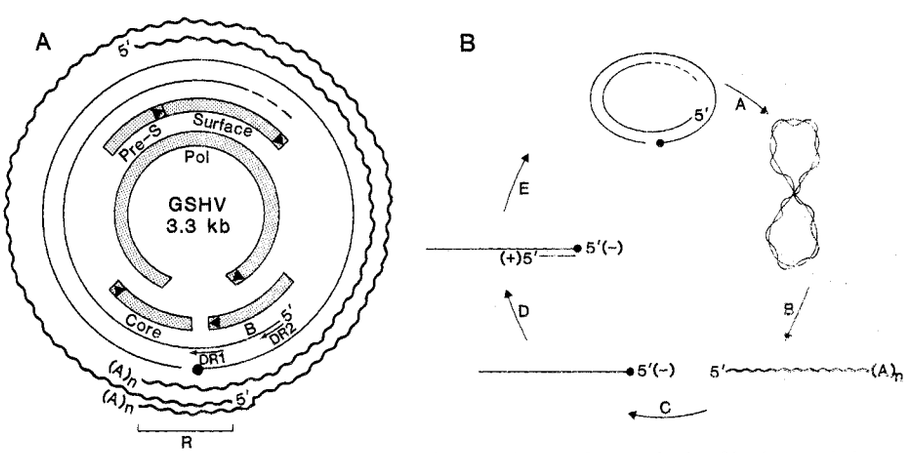

Fig. 1. (A) The transcriptional and translational maps of the genome of a mammalian hepadnavirus, GSHV. Positions of the four continuous open reading frames are represented by stippled boxes, with arrowheads denoting the direction of translation (Pre-S and Surface indicate surface antigens, Core indicates core antigen, Pol indicates the viral DNA polymerase, and B indicates a protein of unknown function). The reading frames are placed in relation to the open circular DNA genome found in virus particles (DNA strands shown as thin lines); protein covalently linked to the 5′ end of the longer (minus) strand is represented by a solid circle, 5′ indicates the 5′ end of the incompleted plus strand (dashes imply variability at the 3′ end). The 11-bp direct repeats, DR1 and DR2, are not to scale and their direction is arbitrary. The wavy lines in the outer portion of the figure represent the two major classes of intracellular viral RNA (*23*). The subgenomic RNA is the major mRNA for surface antigen, the RNA longer than the DNA genome consists of the template for DNA synthesis (pregenome) and the presumptive mRNA for core antigen. The RNA's are polyadenylated, $(A)_n$, at a common 3′ terminus; heterogeneity at the 5′ ends is not shown; and the terminal redundancy in the larger class of RNA is indicated by R. Maps for other mammalian hepadnaviruses (human and woodchuck) are indistinguishable (*18, 28, 29, 31, 32*). (B) The replication cycle of hepadnavirus genomes. Virion DNA (at top) is converted after infection to closed circular DNA (step A), which serves as template for synthesis of pregenomic RNA (step B). The RNA template is copied into protein-primed minus strand DNA (step C), which in turn templates synthesis of plus strand DNA (step D); synthesis of both strands is presumably catalyzed by a viral DNA polymerase encoded in the pol frame. Further extension of the plus strand following circularization (step E) produces the open circular form with cohesive 5′ termini found in virus particles. The model is based on that first proposed by Summers and Mason (*3*).

oligomer to prime synthesis on an M13 clone of GSHV DNA spanning the relevant region. The position of the upper band indicates that the dA (deoxyadenylate) residue of DR1, at position 14 in the GSHV sequence, is the first nucleotide in the minus strand (Fig. 2D). The result is in remarkable agreement with that of Molnar-Kimber *et al.* (*21*) who mapped the 5′ end of the DHBV minus strand to the third residue in DR1, and it implies, as anticipated, that initiation of minus strand could occur within either copy of R, near the 3′ or 5′ end of pregenomic RNA (*22, 23*).

Termination of the minus strand could occur in response to some internal signal or, more likely, by exhaustion of the pregenomic template. Since three closely spaced 5′ termini have been identified in full-length species of GSHV RNA (*23*), multiple 3′ ends of minus strand DNA would be expected if synthesis proceeded to the end of the template and if all three RNA's were packaged in subviral particles and used as pregenomes. To locate the exact 3′ end of minus strand DNA, a 240-nucleotide minus strand fragment of GSHV DNA, produced by cleavage with Nsi I and Cla I (positions 3244 to 173 on the circular map), was labeled with ^{32}P at its 3′ end, annealed to purified virion DNA, and subjected to digestion with single-strand-specific (S1) nuclease and polyacrylamide gel electrophoresis (PAGE) (Fig. 2C). From results with increasing amounts of S1 nuclease, the major protected fragment was approximately 167 nucleotides in length (lanes 5 to 7), placing the 3′ terminus of minus strand DNA at position 6 ± 1 on the GSHV map (*18*), as determined by comparison with a nucleotide sequence of known origin (lanes 1 to 4). Digestion beyond this end point with higher concentrations of S1 nuclease (lane 7) is probably due to the presence of five thymidine residues at positions 8 to 12; this phenomenon was not observed when the experiment was repeated with mung bean nuclease (*30*). Furthermore, no protection of the labeled probe was evident in the absence of virion DNA (lanes 8 and 9).

Since the 5′ end of the shortest species of the three candidate pregenomic GSHV RNA's was mapped to position 6 ± 3 (*23*), our results suggest that only one of the three RNA species is likely to be used for DNA synthesis and that synthesis of minus strand DNA terminates at the 5′ end of the templating RNA. Finally, placement of the 5′ end of minus strand DNA at position 14 and its 3′ end at position 6 provides an overall length of 3320 nucleotides and a short terminal redundancy (r) of 9 nucleotides (Fig. 2D).

The 5′ end of plus strand DNA. The plus strands of mammalian hepadnaviruses span only 40 to 60 percent of the length of the minus strand, and their 3′ ends are heterogeneous; thus precise mapping of the 3′ ends is unlikely to provide insight into the replication mechanism. The 5′ end, however, appears to be homogeneous and is presumed to reflect the priming site for the second strand on a template of minus strand DNA. The 5′ end of DHBV plus strand DNA was initially mapped to the vicinity of the 5′ side of DR2 (*21*). More recently, Lien *et al.* (*27*) demonstrated that RNA is covalently linked to the 5′ end of DHBV plus strand DNA and that the 5′ end remaining after removal of ribonucleotides is positioned at the 3′ boundary of DR2. Partial sequencing of the oligoribonucleotide at the 5′ end of the plus strands indicated that it includes the sequence common to DR1 and DR2, as well as nucleotides expected from the 5′ side of DR1. Thus it was concluded that an oligomer containing DR1 from the 5′ end of DHBV pregenomic RNA was translocated to the DR2 site to prime synthesis of plus strand DNA.

In order to generalize these observations to include the mammalian hepadnaviruses, we mapped the 5′ ends of GSHV and WHV plus strands by primer extension before and after digestion of virion plus strands with RNase. For this analysis, we used a synthetic oligomer complementary to the one used for mapping the 5′ end of

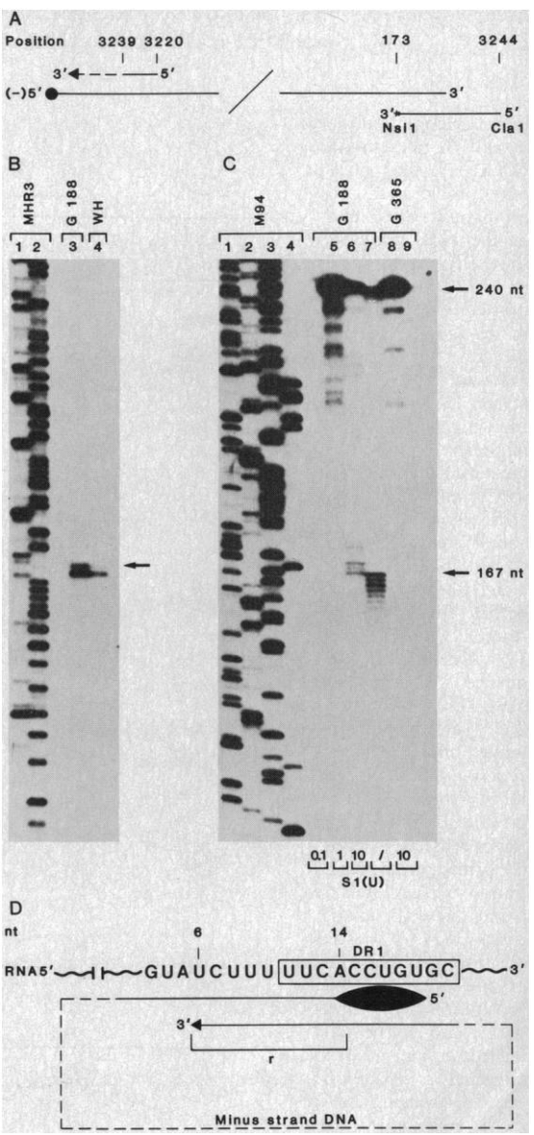

Fig. 2. Mapping the ends of minus strand DNA. (A) The position of the oligonucleotide (3220 to 3239) used as a primer for mapping the 5′ end of minus strand DNA by primer extension is shown together with the position of the Nsi I–Cla I fragment (3244 to 173) employed for S1 mapping of the 3′ end. The positions are numbered according to the sequenced GSHV genome (*18*). The figure is not to scale. (B) Primer extension analysis. Virion DNA was purified from the serum of an infected Beechey ground squirrel (G188) and woodchuck (WH). The template for the sequencing reaction was M13 clone MHR3 with a 1.5-kb GSHV Eco RI–Hinc II fragment cloned into M13mp8. For primer extension analysis and sequencing reactions, the chemically synthesized primer annealing at position 3220 to 3239 was used as in (A) and (*41, 42*). (Lanes 1 and 2) The C and T tracks of the sequencing reaction with MHR3 DNA; (lanes 3 and 4) primer extended products from virion DNA of ground squirrel G188 and woodchuck WH, respectively. (C) S1 mapping of the 3′ end of minus DNA strand with the 3′ end-labeled Nsi I–Cla I fragment shown in (A) (*43, 44*). A sequencing ladder derived from M13 clone M94, containing a 1.5-kb Eco RI to Hinc II GSHV fragment cloned into M13mp9, provided size markers. (Lanes 1 to 4) The A, C, G, and T tracks from M94 primed with the oligomer used for primer extension on plus strand DNA (Fig. 3A). (Lanes 5 to 7) The fragments of 3′ end–labeled DNA protected from S1 digestion by annealing to DNA from animal G188 after treatment with 0.1, 1, and 10 units of nuclease S1. (Lanes 8 and 9) Controls with animal G365, which had no detectable virion DNA in the serum. (Lane 8) No nuclease; (lane 9) 1 unit of nuclease S1. (D) The initiation and termination sites of minus strand DNA. The initiation site at position 14 and the termination site at position 6 are shown in relation to the position of DR1 on the pregenome RNA. The terminal redundancy of the minus strand DNA (r) is indicated. The solid oblong symbol represents protein covalently bound to the 5′ terminus of minus strand DNA. The positions on the RNA are in accordance with the GSHV sequence in (*18*).

minus strand DNA (Fig. 3A). Extension of the primer on GSHV and WHV plus strands not treated with RNase produced a major fragment 147 ± 1 nucleotides in length, indicating that the 5′ end of the plus strands corresponded to a position 7 ± 1 nucleotides upstream of DR2 [Fig. 3B, lanes 5 and 8; position 3093 on the GSHV map (*18*)]. Addition of pancreatic RNase A (lanes 6 and 9) or RNase T1 (lanes 7 and 10) reduced the length of the extended fragments by 18 and 17 nucleotides, respectively. These findings showed that the plus strands of mammalian hepadnaviruses are also linked to oligoribonucleotides and that the DNA sequence of the plus strands must begin near the 3′ boundary of DR2 (Fig. 3C).

The RNA sequence across the 3′ boundary of DR2 reads 5′-UGUGCAGA-3′, with C representing the final residue in the DR sequence. On the basis of the altered mobilities of the primer extension products and the known specificities of the RNases, RNase A appears to have cut the plus strand 3′ of the C residue and RNase T1, 3′ of the preceding G residue. This suggests that plus strand DNA begins exactly 3′ of the DR2 sequence. However, since neither of the nucleases recognizes rA (riboadenosine) residues, we cannot exclude the possibility that the DNA sequence initiates with dG at position 3112, rather than dA at position 3111.

Viability of DR1 and DR2 mutants. Our results to this point indicate that the 5′ ends (and hence the priming sites) for both minus and plus strand DNA's of mammalian hepadnaviruses map to fixed sites near or within the two DR sequences, as also reported for DHBV (*21, 27*). Since DR1 is identical to DR2 in all sequenced clones of hepadnavirus DNA, despite small differences between the DR sequences of virus isolates from different species (*18, 28–32*), we designed experiments to determine whether the replication mechanism would tolerate changes in DR1 or DR2 and, if so, whether the differences would be corrected to restore identity during replication.

A genetic analysis of this type has recently become possible for hepadnaviruses, despite the lack of a tissue culture system for virus propagation, because cloned genomes can initiate production of infectious virus when introduced directly into the liver of susceptible animals (*33, 34*). We therefore used a synthetic oligomer to produce a G to C transversion in either DR1 or DR2 of a GSHV DNA clone previously shown to be infectious (*34*) (Fig. 4A). This nucleotide change was chosen because it represents the difference between the DR sequences in GSHV and human hepatitis B virus (*18*).

Recircularized viral DNA from each of the two mutated plasmids, DR1* and DR2*, was injected into the livers of three Beechey ground squirrels, and sera were screened periodically for the appearance of GSHV surface antigen (see legend to Fig. 4). Two of the three animals that received DR2* (G315 and G316) and one of the three that received DR1* (G307), became seropositive within 13 to 19 weeks, consistent with our experience with wild-type GSHV DNA (*30, 34*); active virus replication was confirmed in all three animals by detection of GSHV DNA in serum samples.

To define the genotypes of the viruses produced by the infected animals, we isolated virion DNA and subjected it to molecular cloning and partial nucleotide sequence determination (Fig. 4B). The DNA from animal G307 showed a mutant DR1 sequence and a wild-type DR2 sequence, whereas DNA's from animals G315 and G316 showed the reverse pattern (wild-type DR1 and mutant DR2). Thus the original mutant genotypes were preserved during all three infections, indicating that identical DR's are not required for successful replication and that the replication scheme does not convert a mutant DR to the wild-type sequence or vice versa.

The plus strand primer in DR mutants. Since DR1 and DR2 are distinguishable in the DR1* and DR2* mutants, the viability of these mutants allowed us to perform a direct test of the contention that the oligoribonucleotide linked to plus strand DNA at the 3′ boundary of DR2 is derived from a portion of viral RNA that contains DR1 (*27*). The G to C transversions in the DR mutants produce RNA substrates distinguishable by RNase T1 (rG-specific) and RNase CL3 (rC-specific). Plus strands were prepared from virion DNA of DR1* and DR2* mutants and analyzed by the primer extension procedure used for wild-type plus strands (Fig. 3) after partial digestion with RNase T1 and RNase CL3.

Figure 5A indicates the products expected if the DR1* sequence is present as RNA at the 5′ end of plus strands in DR1* virus or if the DR1 sequence is present similarly in DR2* virus, both as predicted by the findings of Lien *et al.* (*27*). Extension products ending at positions 20, 18, and 4 should be generated after RNase T1 digestion of virion DNA with a wild-type DR1 (from animals G188 and G316), but the product ending at position 18 should be missing from the analysis of DNA with the DR1* genotype (from animal G307). Conversely, products observed after digestion with RNase CL3 should include one that ends at position 18 only with the DR1* genotype and not with wild-type and DR2* genotypes.

These predictions were all confirmed in the analysis shown in Fig. 5, B and C. Most important, partial digestion with RNase T1 (Fig. 5B, lanes 8 to 10) produces extension products consistent with cleavage at position 18 only in samples from viruses with the DR1 genotype (lanes 8 and 10), not with the sample from DR1* virus (lane 9), and digestion with RNase CL3 produces an extension product at that position only in the sample from DR1* virus (Fig. 5C, lane 10). Thus the genotype of the DR sequence of the

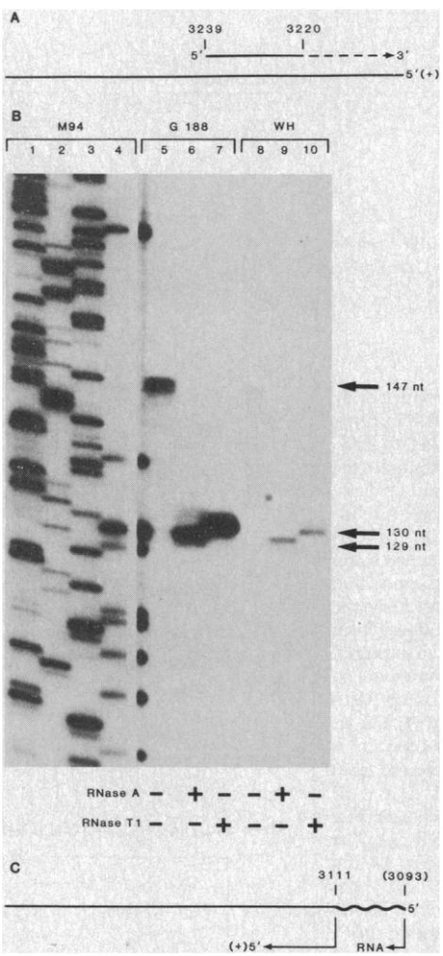

Fig. 3. Mapping of the 5′ end of plus strand DNA. (A) The position of the primer used for mapping the 5′ end of the plus strand is shown (position 3239 to 3220). (B) For primer extension on minus strand DNA, the same protocol was followed as described in the legend to Fig. 2B, except that an oligomer of complementary sequence was used, and RNase A (1.5 μg/μl) (lanes 6 and 9) or RNase T1 (0.8 unit/μl) (lanes 7 and 10) were added before the extension reaction. (Lanes 1 to 4) Nucleotide sequence analysis of M94 DNA (as in Fig. 2C); (lanes 5 to 7) extended primer from reactions with viral DNA from infected ground squirrel G188; (lanes 8 to 10) extended primer from reactions with viral DNA from infected woodchuck WH. (C) The positions of the 5′ end of plus strand DNA (position 3111) and the 5′ end of the 18-nucleotide-long RNA oligomer (position 3093) covalently attached to it.

presumptive RNA primer is determined by the DR1 genotype, confirming evidence from the analysis of the sequence upstream of the DR sequence (*27*) that an oligomer containing DR1 has been transposed to the position of DR2.

Implications for hepadnavirus replication. The analyses of mammalian hepadnaviruses presented here, viewed in concert with recent studies of DHBV (*22, 26, 27*) and earlier findings (*3, 15, 16, 25*), allow us to propose a relatively complete scheme for the synthesis of hepadnavirus DNA (Fig. 6A) and to compare it with corresponding features of retroviral reverse transcription (Fig. 6B).

1) *The template.* Although hepatitis B RNA is exclusively an intracellular form of the genome, whereas retroviral RNA is found in virus particles, both templates for reverse transcription are terminally redundant, polyadenylated, and probably equivalent to mRNA for the major core proteins of each virus (Fig. 6, panel 1). The pregenomic RNA of hepatitis B viruses is longer than the DNA to be made from it (and thus longer than the DNA template from which it is transcribed), whereas the retroviral RNA genome is shorter than its unintegrated DNA products and its proviral DNA template. These inverted length relationships are due to the presence of LTR's in the DNA products of retroviral reverse transcription (*2, 12*). Two features of hepadnavirus DNA synthesis figure in the failure to produce LTR's: the position of the priming sites and the lack of displacement synthesis (see below).

2) *The first strand.* The first strand of hepadnaviral DNA begins within the RNA template's terminal redundancy (R), four nucleotides from the 3′ end of DR1, presumably initiated by an uncharacterized protein primer (Fig. 2 and Fig. 6A, panel 2). Retroviral minus strand, in contrast, is initiated outside R, at a position roughly 100 nucleotides downstream, at the end of an 18-nucleotide

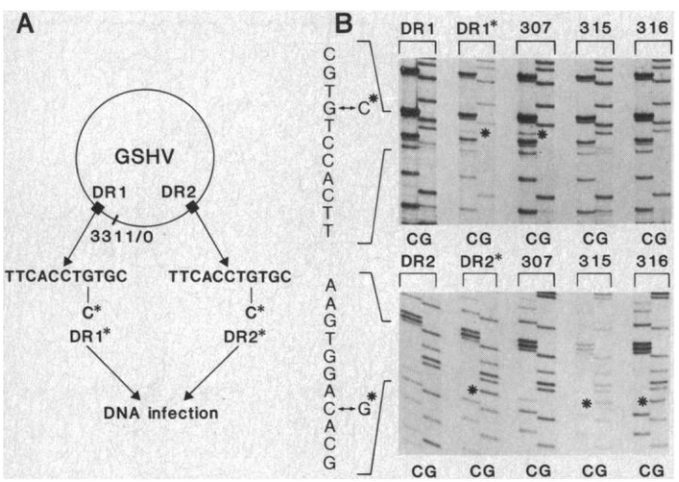

Fig. 4. Generation and analysis of DR1 and DR2 mutants. (A) Oligomer-directed, site-specific mutagenesis (*45*) was used to introduce the indicated G to C transversion independently into DR1 and DR2 of an infectious clone of GSHV DNA. The resulting mutants DR1* and DR2* were assayed separately for infectivity by injection of monomeric circles into livers of susceptible ground squirrels (*34*). (B) Partial nucleotide sequence analyses of DR1 (top) and DR2 (bottom) are shown for the wild-type GSHV clone 27 (left pair of lanes), for the mutated clones DR1* and DR2*, and for the DR sequences present in clones of viral DNA derived from animals that received DR1* DNA (ground squirrel 307) or DR2* DNA (ground squirrels 315 and 316). Only the analysis of relevant nucleotides (G and C) is shown. Identification of mutant residues is highlighted by asterisks. The deduced nucleotide sequences of the DR's on the analyzed strand are indicated at the left margin. The DR1 sequence from the plus strand is read 5′ to 3′ from the bottom of the panel, and the DR2 sequence from the minus strand is read from the top. Cloning of virion DNA from animals 307, 315, and 316 into λgt10 and dideoxy nucleotide sequence analyses were performed as described (*42, 46*).

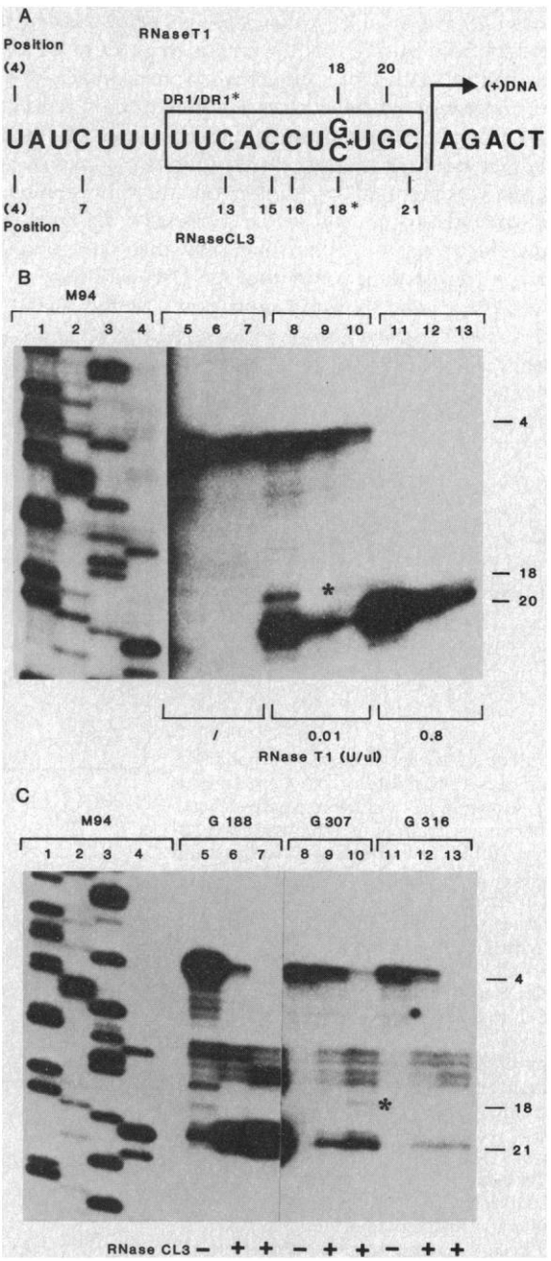

Fig. 5. Partial nucleotide sequence analysis of the RNA primer on the 5′ end of plus strand DNA in DR1* and DR2* mutants. (A) The positions of the RNase T1 (positions 18 and 20) and the RNase CL3 recognition sites (positions 7, 13, 15, 16, and 21) are shown with respect to the nucleotide sequence of the anticipated RNA primer for plus strand DNA synthesis. The DR1 or DR1* sequence is boxed and shows the G to C conversion at position 18 of the DR1* mutant; the mutant has acquired an additional RNase CL3 recognition site (denoted as 18*) and lost the RNase T1 site at the same position. Position 4 is the predicted 5′ end of the RNA primer (see text). (B) Primer extension analysis was done on plus strand templates from virion DNA of the designated animals as described in Fig. 3B, except that RNase T1 at 0.01 unit/μl was used for partial digestion in lanes 8, 9, and 10. (Lanes 5, 8, and 11) Animal G188; (lanes 6, 9, and 12) animal G307; (lanes 7, 10, and 13) animal G316. Evidence for the G to C transversion at position 18 in virion RNA from animal G307 (DR1*) is indicated. (C) Primer extensions were performed as in (B), except that RNase CL3 was used. The concentration of RNase CL3 was 0.2 unit/μl in lanes 6, 9, and 12 and 2 unit/μl in lanes 7, 10, and 13. (Lanes 5 to 7) Animal G188; (lanes 8 to 10) animal G307; (lanes 11 to 13) animal G316. Evidence for the G to C transversion at position 18 in virion RNA from animal G307 (DR1*) is indicated.

primer binding site (PBS) to which the 3′ end of a host tRNA is base-paired (*2*) (Fig. 6B, panel 2). The sequence between R and PBS—a unique sequence known as U5—is one of the two regions that are duplicated during DNA synthesis to produce LTR's.

Other than the nature of the putative protein primer, two issues about the initiation of hepadnaviral minus strands remain unresolved: Which of the three potential pregenomic RNA's of GSHV (*23*) are used as template? and which of the two copies of DR1 in pregenomic RNA is used as an initiation site? Since minus strand DNA extends just to the end of the shortest of the three possible templates (Fig. 2), we assume that only that one is used. For simplicity, we favor initiation within the DR1 site close to the 3′ end of the pregenome; however, initiation near the 5′ end would produce a nascent strand capable of base-pairing to the 3′ end of pregenomic RNA, in a manner similar to that used to transfer nascent retroviral minus strands (*2*) (Fig. 6B, panel 3). In either case, fully extended minus strands would have the short terminal redundancy (r) that appears pivotal in a later stage of synthesis (see below).

3) *The second strand.* Primers for the second DNA strands of hepadna- and retroviruses are similar in kind (oligomers of viral RNA) and position (a few hundred nucleotides upstream from the 3′ end of viral RNA). The retroviral primer, however, appears to be generated by the action of virus-coded ribonuclease H at specific sites near a polypurine tract (PPT) adjacent to the unique sequence (U3) destined to be duplicated during DNA synthesis; synthesis of plus strands commences when minus strands are still relatively short; and the primer initiates synthesis at its native position (*2*) (Fig. 6B, panels 3 and 4).

To generate the second hepadnavirus primer (*27*) (Figs. 3 and 5), an approximately 18-nucleotide RNA oligomer containing DR1 and several 5′ nucleotides must be released from viral RNA and transported by undefined mechanisms to the position previously occupied by RNA containing DR2 (Fig. 6A, panels 3 and 4). The primer could be derived from the DR1 sequence near either end of viral RNA, but two findings argue for derivation from the 5′ end of the RNA, as shown in Fig. 6A, panel 3; Lien *et al.* (*27*) found that a phosphodiesterase specific for linkage of cap nucleotides to RNA is active on the DHBV primer, and we have found concordance between the 5′ end of the plus strand prior to RNase digestion and the 5′ end of a candidate pregenomic RNA (Fig. 3) (*23*). We cannot explain how or why hepadnaviruses move a distant oligonucleotide to a partly complementary priming site, instead of being primed by endogenous RNA in the manner of retroviruses.

Regardless of priming mechanism, synthesis of both retro- and hepadnavirus plus strands must encounter an exhaustion of minus

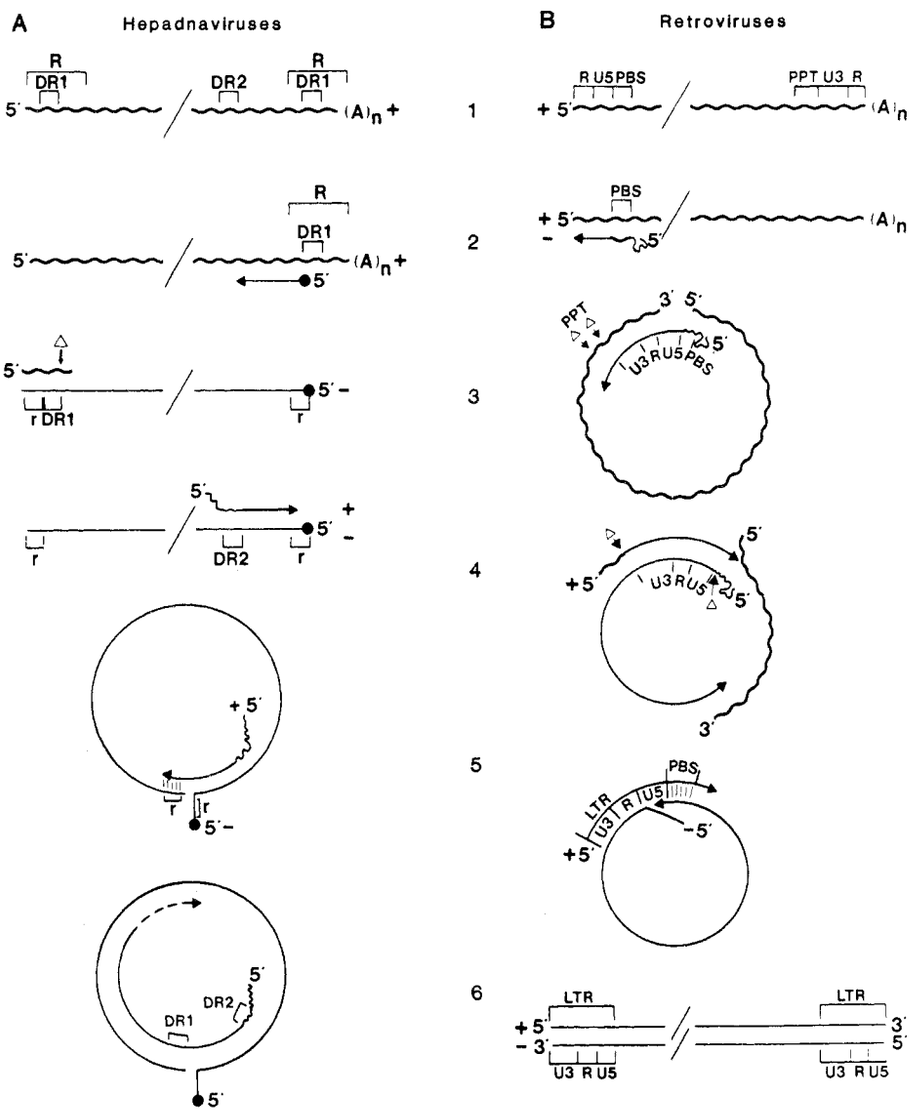

Fig. 6. Models for the synthesis of viral DNA from an RNA template. The hepadnavirus scheme is shown in (A) and the retrovirus scheme in (B). The composition of the strands are indicated by wavy (RNA) and straight (DNA) lines, and the chemical polarity is indicated by + and − signs. (1) The RNA templates, with terminal redundancies (R), polyadenylic acid tract, (A_n), hepadnavirus direct repeats (DR1, DR2), and retrovirus tRNA primer binding site (PBS) and polypurine tract (PPT) flanking U5 and U3. (2) Priming of the first (minus) strands, by a protein at a position within DR1 or by host tRNA at the boundary of U5 and PBS. (3) Extension of the minus strand and generation of the second (plus) strand primer. Completion of the minus strand produces a short terminal redundancy (r) in hepadnavirus DNA; an open circular retroviral RNA-DNA hybrid is formed after transfer of the nascent DNA from the 5′ to the 3′ end of an RNA subunit. (An alternative model in which the transfer occurs to the second subunit in the dimeric retroviral genome is possible but not further considered.) Open triangles indicate presumptive scissions of viral RNA by DNA polymerase-associated RNase H activities to produce oligoribonucleotides for priming the second strand. (4) Priming and extension of the plus strand. Movement of the DR1 containing oligoribonucleotide to the DR2 position precedes initiation of hepadnavirus plus DNA; retroviral plus strand is a copy of U3, R, U5, and a portion of the tRNA primer. (5) Transfers of the nascent plus strands are accomplished using complementary sequences (r for hepadnaviruses, PBS for retroviruses) at the ends of the strands to form short duplexes (accented with short vertical lines). (6) Extension of strands to form the final product. Partial synthesis of hepadnaviral plus strands yields the open circular form of virion DNA with protein and RNA primers still attached. Full extension of both plus and minus retroviral strands by displacement synthesis yields a linear duplex with long terminal repeats (LTR's).

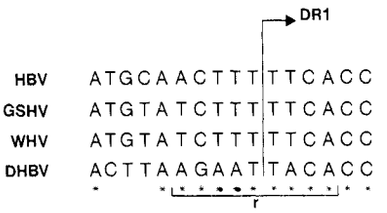

Fig. 7. The nucleotide sequences at the 5′ boundary of DR1 in the genomes of the four hepadnaviruses HBV (*31*), GSHV (*18*), WHV (*28, 29*), and DHBV (*32*) are aligned to show sequence conservation in the region destined to form the short terminal redundancy (r) in minus strand DNA. Asterisks indicate the positions at which only A or T or only G or C residues are observed.

DNA template shortly after initiation (Fig. 6, panel 4). The resulting retroviral plus strand, known as plus strong stop, is believed to end at the site of the first methylated base in the tRNA primer, after copying the first 18 nucleotides of the tRNA (*2*). Presumably the hepadnaviral plus strong stop terminates near or at the end of the r sequence at the 5′ end of the minus strand.

Further extension of either retroviral or hepadnaviral plus strands requires transfer of the nascent strands to new templates, and in both cases the transfer appears to be facilitated by base-pairing between the 3′ regions of plus strong stop and minus strand DNA. Retroviral DNA's can form an 18-nucleotide duplex composed of the PBS sequence (*2*) (Fig. 6B, panels 4 and 5); hepadnaviral DNA's can form a 9-nucleotide duplex composed of the r sequence, assuming that the existing duplex at the 5′ end of the minus strand can dissociate (Fig. 6A, panels 4 and 5). Two features might expedite the dissociation: protein linked to the 5′ end of the minus strand (perhaps in concert with DNA polymerase) and the high A+T content (and hence low T_m) of the r sequence. There is full conservation of the positions of A or T compared to G or C residues in this region of all four hepatitis B virus genomes, despite some minor sequence variations (Fig. 7). Such conservation would be consistent with an absolute requirement to transfer the second strand by a mechanism dependent upon local melting and duplex formation.

4) *Completion*. The retroviral intermediate generated by transfer of the second strand, most likely the open circular form (Fig. 6B) is presumed to be transient; it is converted to a linear duplex, with long terminal repeat units composed of U3, R, and U5, when the plus and minus strands are fully extended by displacement synthesis (*2*) (Fig. 6B, panel 6). In contrast, the open circle formed by transfer of the nascent plus strand of hepadnaviral DNA is altered only by the partial and variable extension of plus strand along the minus DNA template before export in virions (Fig. 6A, panel 6); there is no further elongation of minus or plus strands by displacement synthesis, no disruption of the open circle, and hence no production of long terminal repeats in the DNA. The form of viral DNA isolated from virions retains its protein and oligoribonucleotide primers still joined to the 5′ ends of minus and plus strands, and the r sequence remains at both ends of the minus strand.

5) *Fate of DNA*. Linear retroviral DNA, completed in the cytoplasm of infected cells, is later circularized after transport to the nucleus, forming covalently closed molecules with either one or two LTR's; the latter is the apparent substrate for integrative recombination with the host chromosome (*2*). Further events in the hepatitis B life cycle may be postponed until virus particles containing open circular DNA enter a susceptible cell, or they may occur as part of an intracellular replication scheme. In either case, the open circular DNA is converted to closed circular DNA, which in turn serves as template for synthesis of viral RNA, without a requirement for integration. Several steps are necessary for the transition to a closed circular form: removal of the RNA and protein primers, trimming of the r sequence from minus DNA, completion of the plus strand, ligation, and introduction of superhelical turns. It is not known which of these steps, if any, are catalyzed by virus-coded proteins.

Prospects. The life cycles of hepadnaviruses and retroviruses exhibit alternative solutions to some of the problems inherent in DNA synthesis from an RNA template: the composition and placement of primers, the lengths of the DNA strands, and the transfer of nascent strands between templates. The steps described provide guides to the study of other genetic elements that are subject to RNA-directed synthesis of DNA, and, when the mechanisms are better understood, they may offer targets for the design of antiviral therapy. At least two features of the hepatitis B viruses invite wider comparisons. The protein linked to the first strand of HBV DNA, the presumptive primer, is reminiscent of the protein primer for adenovirus DNA (*35*) and the protein joined to the 5′ end of poliovirus RNA (*36*). The priming of the second strand occurs by a mechanism that produces a polynucleotide whose adjacent sequences, composed of RNA and DNA, are discontinuous in minus strand DNA or in pregenomic RNA; in this sense, the product is analogous to several other polynucleotides, including the mRNA's of coronaviruses (*37*), influenza viruses (*38*), and trypanosomes (*39*), as well as conventional eukaryotic mRNA's produced by splicing (*40*). Development of a soluble system for the synthesis of hepatitis B virus DNA is likely to be necessary for further insight into these intriguing aspects of the virus life cycle.

REFERENCES AND NOTES

1. D. Baltimore, *Nature (London)* **226**, 1209 (1970); H. M. Temin and S. Mizutani, *ibid.*, p. 1211.
2. H. E. Varmus and R. Swanstrom, in *The Molecular Biology of RNA Tumor Viruses*, R. Weiss, N. Teich, H. E. Varmus, J. Coffin, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1982) p. 369; *ibid.* (1985), p. 75.
3. J. Summers and W. S. Mason, *Cell* **29**, 403 (1982).
4. P. Pfeiffer and T. Hohn, *ibid.* **33**, 781 (1983); R. Hull and S. N. Covey, *Trends Biochem. Sci.* **8**, 119 (1983). M. Volovitch, N. Modjtahedi, P. Yot, G. Brun, *EMBO J.* **3**, 309 (1984).
5. J. D. Boeke, D. J. Garfinkel, D. A. Styles, G. R. Fink, *Cell* **40**, 491 (1985).
6. D. J. Garfinkel, J. D. Boeke, G. R. Fink, *ibid.* **42**, 507 (1985).
7. J. Mellor *et al.*, *Nature (London)* **318**, 583 (1985).
8. T. Shiba and K. Saigo, *ibid.* **302**, 119 (1983); A. J. Flavell and D. Ish-Horowicz, *ibid.* **292**, 591 (1981).
9. Y. Nishioka, A. Leder, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* 77, 2806 (1980).
10. P. A. Sharp, *Nature (London)* **301**, 471 (1983).
11. L. DiGiovanni, S. R. Haynes, R. Misra, W. R. Jelinek, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6533 (1983).
12. P. R. Shank *et al.*, *Cell* **15**, 1383 (1978); T. W. Hsu, J. L. Sabran, E. G. Mark, R. V. Guntaka, J. M. Taylor, *J. Virol.* **28**, 810 (1978).
13. K. Saigo *et al.*, *Nature (London)* **312**, 659 (1984); S. M. Mount and G. M. Rubin, *Mol. Cell. Biol.* **5**, 1630 (1985); J. Clare and P. Farabaugh, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2829 (1985).
14. P. Tiollais, C. Pourcel, A. Dejean, *Nature (London)* **317**, 489 (1985).
15. W. S. Mason, C. Aldrich, J. Summers, J. M. Taylor, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3997 (1982).
16. B. Weiser, D. Ganem, C. Seeger, H. E. Varmus, *J. Virol.* **48**, 1 (1983).
17. H. Toh, H. Hayashida, T. Miyata, *Nature (London)* **305**, 827 (1983).
18. C. Seeger, D. Ganem, H. E. Varmus, *J. Virol.* **51**, 367 (1984).
19. W. H. Gerlich and W. S. Robinson, *Cell* **21**, 801 (1980); D. Ganem, L. Greenbaum, H. E. Varmus, *J. Virol.* **44**, 374 (1982).
20. K. L. Molnar-Kimber, J. Summers, J. M. Taylor, W. S. Mason, *J. Virol.* **45**, 165 (1983).
21. K. L. Molnar-Kimber, J. W. Summers, W. S. Mason, *ibid.* **51**, 181 (1984).
22. M. Buscher, W. Reiser, H. Will, H. Schaller, *Cell* **40**, 717 (1985).
23. G. H. Enders, D. Ganem, H. E. Varmus, *ibid.* **42**, 297 (1985).
24. N. Ruiz-Opazo, P. R. Chakrabarty, D. A. Shafritz, *Cell* **29**, 129 (1982).
25. W. S. Mason *et al.*, *Virology* **131**, 375 (1983).
26. J. Tuttleman, C. Pourcel, J. Summers, personal communication.
27. J. M. Lien, C. E. Aldrich, W. S. Mason, *J. Virol.* **57**, 229 (1986).
28. F. Galibert, T. N. Chen, E. Mandart, *ibid.* **41**, 51 (1982).
29. K. Kodama, N. Ogasawara, H. Yoshikawa, S. Murakami, *ibid.* **56**, 978 (1985).
30. C. Seeger, unpublished results.
31. F. Galibert, E. Mandart, F. Fitoussi, P. Tiollais, P. Charnay, *Nature (London)* **281**, 646 (1979); P. Valenzuela, M. Quiroga, J. Zaldivar, P. Gray, W. J. Rutter, in *Animal Virus Genetics*, B. N. Fields and R. Jaenisch, Eds. (Academic Press, New York, 1980), p. 57.
32. E. Mandart, A. Kay, F. Galibert, *J. Virol.* **49**, 782 (1984). R. Sprengel, C. Kuhn, H. Will, H. Schaller, *J. Med. Virol.* **15**, 323 (1985).
33. H. Will *et al.*, *Nature (London)* **299**, 740 (1982); R. Sprengel, C. Kuhn, C. Manso, H. Will, *J. Virol.* **52**, 932 (1984).

34. C. Seeger, D. Ganem, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5849 (1984).
35. D. M. K. Rekosh, W. C. Russell, A. J. D. Bellet, A. J. Robinson, *Cell* **11**, 283 (1977).
36. E. Wimmer, *ibid.* **28**, 199 (1982).
37. M. M. C. Lai, R. S. Baric, P. R. Brayton, S. A. Stohlman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3626 (1984).
38. R. M. Krug, *Cell* **41**, 651 (1985).
39. P. Borst, *Ann. Rev. Biochem.*, in press.
40. P. A. Sharp, *Cell* **42**, 397 (1985).
41. Virus was purified from the serum from an infected Beechey ground squirrel (G188) and woodchuck WH by cesium chloride equilibrium density centrifugation. Serum (2.5 ml) was centrifuged in a total volume of 13 ml at $d = 1.18$ g/cm^3 in a Beckman type 70.1 rotor at 50,000 rev/min for 24 hours. Fractions (300 μl) were collected from the top of the gradient and assayed for GSHV-specific DNA by the dot-blot procedure. Positive fractions (density, 1.357 ± 0.002 g/cm^3) were pooled, diluted fourfold with 150 m*M* NaCl and 20 m*M* tris-HCl (*p*H 7.5), and centrifuged for 16 hours at 30,000 rev/min in a Beckman SW 50.1 rotor. Virion DNA was purified from virus pellets described (*19*). Virion DNA (1 ng) was denatured at 100°C for 2 minutes in the presence of at least a 100-fold molar excess of primer in a volume of 4.5 μl. Samples were chilled immediately on ice and 0.5 μl of TM (100 m*M* tris-HCl, 100 m*M* $MgCl_2$) was added. For a primer extension reaction, 2 μl of virion DNA and primer were mixed with 1 μl of a solution containing dGTP, dCTP and TTP (300 μ*M* each). The reaction was started by addition of 2 μl of a solution containing ^{32}P-dATP at 1.5 μCi/μl (3000 Ci/mmol, Amersham), AMV reverse transcriptase at 2 unit/μl (Seikagaku America Inc.), 80 m*M* tris-HCl (*p*H 8.4), 16 m*M* $MgCl_2$, 80 m*M* KCl, and 1.6 m*M* dithiothreitol (Cleland's reagent, Calbiochem Inc.). The reaction mixture was incubated at 37°C for 10 minutes and dATP (40 μ*M* final concentration) was added. After 10 minutes, the reaction was stopped by addition of 5 μl of formamide. Samples were denatured for 10 minutes at 70°C before loading on a 6 percent polyacrylamide sequencing gel. Dideoxynucleotide sequence analysis was done as described (*42*).
42. M. D. Biggin, T. J. Gibson, G. F. Hong, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963 (1983); F. Sanger, S. Nicklen, A. R. Coulsen, *ibid.* **74**, 5463 (1977).
43. S1 mapping was performed essentially as described (*44*). Purified virion DNA (2 ng, Fig. 2A) was denatured in 50 percent formamide, 10 m*M* Hepes (*p*H 7) in the presence of gel-purified, 3′ end-labeled Nsi I–Cla I fragment (position 3244 to 173) in a total volume of 8.5 μl at 90°C for 2 minutes. After addition of 1.5 μl 5*M* NaCl, the sample was incubated at 40°C for 15 to 18 hours. The 3′ end of the Nsi I site was labeled by incubating 1 μg of Nsi I–digested plasmid DNA (0.4 p*M* ends) with 7.5 units of the Klenow fragment of DNA polymerase 1 (P-L Biochemicals) in the absence of dNTP's at 37°C for 15 minutes before the addition of dCTG, dGTP, TTP, and [^{32}P]dATP (3000 Ci/mmol, Amersham) to the reaction mix. After a further 15 minutes of incubation at 37°C, the reaction was stopped by heating at 65°C for 10 minutes, and the RNA was then digested with Cla I to release a 240-bp fragment. The fragment was purified by gel electrophoresis, the specific activity of the fragment was 0.5×10^7 to 2×10^7 Cerenkov count/min per microgram of DNA. For an annealing reaction 2×10^4 to 4×10^4 Cerenkov count/min was used. To 5 μl of the annealing mixture, 2 μl of yeast RNA (10 mg/ml) and 90 μl of S1 buffer (50 m*M* sodium acetate, *p*H 4.5, 300 m*M* NaCl, 1 m*M* $ZnCl_2$) were added. After addition of 0.1 to 10 units of S1 nuclease (P-L Biochemicals), samples were incubated for 20 to 30 minutes at room temperature. The reaction was stopped, and the DNA was precipitated by addition of 250 μl of ethanol. Samples were denatured in 50 percent formamide at 70°C before loading on a 6 percent polyacrylamide sequencing gel.
44. A. J. Berk and P. A. Sharp, *Cell* **12**, 721 (1977).
45. R. B. Wallace, M. Schold, M. J. Johnson, P. Dembrek, K. Itakura, *Nucleic Acids Res.* **9**, 3647 (1981).
46. T. Maniatis, E. F. Fritsch, J. Sambrook, in *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), p. 264. N. E. Murray, in *Lambda 11*, R. W. Hendrix, J. W. Roberts, F. W. Stahl, R. A. Weisberg, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983).
47. We thank M. Urdea (Chiron Corp., Emeryville, CA) and J. Barnett (UCSF) for the synthesis of two oligomer primers, B. Mason for open exchange of information prior to publication, D. Westaway and C. Nottenburg for their help with phage cloning, and R. Sprengel, T. De Lange, and G. Enders for comments about the manuscript. Supported by grants from NIH. D.G. is a scholar of the John A. and George L. Hartford foundation and H.E.V. is an American Cancer Society Research Professor.

6 January 1986; accepted 26 February 1986