JOURNAL OF VIROLOGY, June 1991, p. 3309–3316
0022-538X/91/063309-08$02.00/0

Vol. 65, No. 6

cis-Acting Sequences Required for Encapsidation of Duck Hepatitis B Virus Pregenomic RNA

RUSSELL C. HIRSCH,[1] DANIEL D. LOEB,[2] JONATHAN R. POLLACK,[1] AND DON GANEM[2]*

Departments of Microbiology and Immunology[2] and Biochemistry and Biophysics,[1] University of California Medical Center, San Francisco, California 94143-0502

Received 27 December 1990/Accepted 22 March 1991

Hepadnavirus reverse transcription requires that pregenomic RNA first be selectively packaged into a cytoplasmic core particle. This process presumably requires the presence of specific recognition sequences on the pregenomic RNA. To define the *cis*-acting sequences required for pregenome encapsidation in the duck hepatitis B virus (DHBV), we assayed the packaging efficiency of a series of pregenomic RNA deletion mutants and hybrid DHBV/*lacZ* fusion transcripts. The 5′ boundary of the packaging signal lies within the precore region, starting approximately 35 nucleotides from the cap site of pregenomic RNA; thus, the DR1 sequence required for proper viral DNA synthesis is not included in this signal. To define the 3′ boundary of the encapsidation signal, fusion transcripts bearing foreign (*lacZ*) sequences fused to DHBV at different sites 3′ to the pregenomic RNA start site were examined. A surprisingly large region of the DHBV genome proved to be required for packaging of such chimeras, which are efficiently encapsidated only when they contain the first 1,200 to 1,400 nucleotides of DHBV pregenomic RNA. However, mutant genomes bearing insertions within this region are packaged efficiently, making it likely that the actual recognition elements for encapsidation are smaller discontinuous sequences located within this region.

Hepadnaviruses are small, hepatotropic DNA viruses that can produce persistent infections of hepatocytes in a variety of animal hosts (for review, see reference 10). In such infections, their DNA genomes are replicated via reverse transcription of an RNA intermediate (25). As in retroviruses, hepadnavirus reverse transcription does not occur in a soluble form in the cytoplasm but rather takes place in a subviral core particle composed of the major capsid (C) protein, the polymerase (P) protein, and the RNA template (9). Hence, a key step in the replicative cycle of the virus is the selective encapsidation of the appropriate viral RNA into subviral core particles.

The basic steps in the hepadnavirus life cycle are now well established. Following viral entry, the genome is delivered to the nucleus, where it is transcribed into two classes of RNAs (Fig. 1): subgenomic transcripts, which serve as mRNAs for viral surface proteins, and genome-length RNAs (4, 7). Genomic RNAs are bifunctional; they serve as both mRNA (for C and P proteins) and as the templates for reverse transcription (4, 10). Concomitant with or following the translation of these proteins, genomic transcripts are assembled into nascent subviral cores, whereupon reverse transcription begins (8).

Until recently, little was known about the proteins that mediate the packaging of viral RNA into cores and about the particular features of the genomic RNA that dictate its selective encapsidation. We have recently employed a genetic approach to demonstrate that in the duck hepatitis B virus (DHBV), the P gene product is required not only for reverse transcription but also for the efficient packaging of the genomic RNA into viral cores: mutant viruses unable to synthesize P gene products produce empty cores lacking viral RNA (13). Similar findings for the human hepatitis B virus (1, 12) have also been made. Presumably, the packaging of genomic RNA occurs as a result of specific interactions between the P gene product (or a protein complex containing the P gene product) and a *cis*-acting encapsidation signal on the genomic RNA.

The structure and location of the DHBV encapsidation signal are unknown. To map and characterize this signal, we have analyzed the encapsidation efficiency of RNAs produced from a series of DHBV deletion mutants and from gene fusions between DHBV and heterologous sequences. In this article, we report that this signal contains essential elements that are dispersed over a surprisingly large region, extending from the precore region through the first third of the P open reading frame.

MATERIALS AND METHODS

Materials. Enzymes were purchased from New England BioLabs and were used according to the manufacturer's protocols. Radionuclides were purchased from Amersham Corporation.

Plasmids and plasmid construction. pPC1 is a tandem dimer of the European strain of DHBV (24) bearing an 8-bp *Cla*I linker insertion in the *Acc*I site at nucleotide (nt) position 2577. To generate pPC1, an *Nco*I (DHBV nt position 2351) to *Eco*RI fragment (DHBV nt position 3020) of pD3-SP65, a unit-length DHBV genome inserted at the *Eco*RI site of pSP65 (Promega Biotec), was subcloned into pSP65 to yield pD3N/R. pD3N/R was linearized with *Acc*I and blunted with Klenow, and an 8-bp *Cla*I linker was inserted with T4 DNA ligase to yield pD3N/R-*Cla*I. The *Nco*I to *Eco*RI fragment of pD3N/R-ClaI was recloned into pD3-SP65 cut with *Nco*I and *Eco*RI to yield a unit-length DHBV genome bearing a *Cla*I linker insertion in the *Acc*I site at nt position 2577. This genome was reconstructed into a tandem dimer, pPC1. pPC2 was generated in precisely the same manner except that two *Cla*I linkers were cloned into the *Acc*I site.

For construction of deletion mutants (mutants Δ1 to Δ5) in the DHBV precore region, we generated pD3-Xho-PCS.

* Corresponding author.

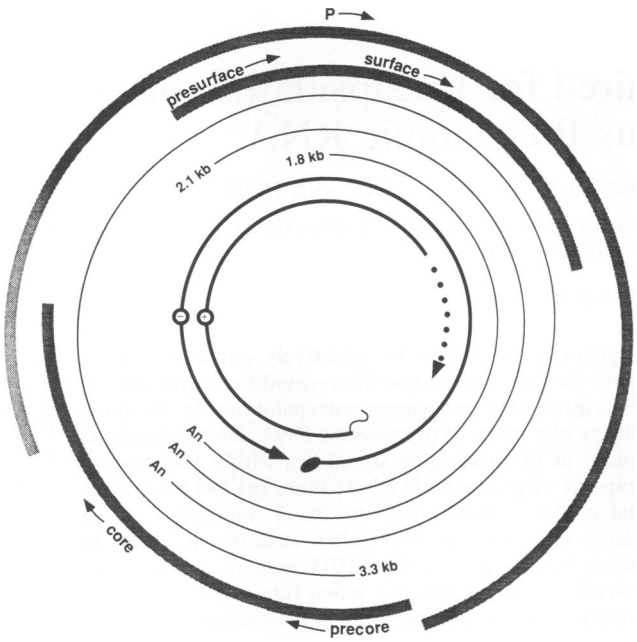

FIG. 1. Genetic and transcriptional map of DHBV. The inner circles represent the partially double-stranded DNA genome. The 5′ end of the full-length minus strand is linked to a terminal protein (solid oval). The 5′ end of the plus strand is in a fixed position and is linked to an oligoribonucleotide (wavy line), whereas the 3′ end of the plus strand is variably situated as indicated by the dotted line. The three thin lines represent the major transcripts seen in infected cells: the 2.1-kb mRNA encoding presurface antigen, the 1.8-kb mRNA encoding surface antigen, and the 3.3-kb RNA serving both as the template for reverse transcription and as the mRNA encoding the core antigen and the P gene product. The thick outer bars indicate the coding regions for the presurface/surface, precore/core, and P proteins. An indicates the mRNA poly(A) tail.

pD3-Xho-PCS was generated by cloning a synthetic oligonucleotide containing *Pst*I, *Cla*I, and *Spe*I sites (in that order) into a unique *Xho*I site that had been created at nt position 2548 of pD3-SP65 by using site-specific mutagenesis. pD3-Xho-PCS was digested with *Pst*I and *Spe*I and then subjected to digestion with exonuclease III and nuclease S1 (double-strand nested deletion kit; Pharmacia). The extent of the deletions was verified by DNA sequencing.

For construction of overlength linear genomes with deletions only in the 5′-terminal redundancy (e.g., Δ2-5′ and Δ5-5′), pD1.5G (an overlength DHBV genome between positions 1658 and 3020) was cut with *Sca*I, treated with calf intestinal alkaline phosphatase and digested with *Nsi*I. The resultant fragments were ligated with either Δ2 or Δ5 genomes that had been cut with *Nsi*I and treated with calf intestinal alkaline phosphatase, yielding a tandem dimeric genome bearing the appropriate deletion only in the region corresponding to the 5′-terminal redundancy. For construction of overlength linear genomes with deletions only in the 3′-terminal redundancy (e.g., Δ2-3′ and Δ5-3′), pD.5G (a partial DHBV genome between positions 1658 and 3020) was cut with *Eco*RI and treated with calf intestinal alkaline phosphatase. The resultant fragment was ligated with either Δ2 or Δ5 genomes that had been digested with *Sca*I, treated with calf intestinal alkaline phosphatase, and then cut with *Eco*RI. The ligation yielded overlength DHBV genomes (1.5 mers between positions 1658 and 3020) bearing the appropriate deletion only in the region corresponding to the 3′-terminal redundancy.

The DHBV/*lacZ* fusion transcripts (pseudopregenomes) were derived from pCMV-DHBV9 (6) as the source of the CMV-IE promoter-driven DHBV sequences and from pON3 (22) as the source of the *Escherichia coli lacZ* sequences and the simian virus 40 polyadenylation signal. Pseudopregenomes p136-X and p319-N were derived, respectively, from pCMV-DHBV9 from which sequences from *Xba*I (DHBV nt position 2662) or from *Nsi*I (DHBV nt position 2845) to *Bam*HI (polylinker) were removed and were replaced with a 3.0-kb *Hpa*I-*Xba*I fragment from pON3. p508-H was generated by removing the *Hin*dIII (DHBV nt position 14) to *Bam*HI (polylinker) fragment of pCMV-DHBV9 and replacing it with the 2.5-kb *Cla*I-*Xba*I fragment of pON3. p814-S was derived from pCMV-DHBV9 from which the sequence between *Sma*I (DHBV nt position 320) and *Bam*HI (polylinker) has been replaced with a 2.2-kb *Eco*RV-*Xba*I fragment of pON3. p1212-R was generated by removing the *Eco*RV (DHBV nt position 718) to *Bam*HI (polylinker) fragment of pCMV-DHBV9 and replacing it with the 1.8-kb *Ava*II-*Xba*I fragment of pON3. p1396-T and p1780-K were derived, respectively, from pCMV-DHBV9 from which sequences between *Tth*111I (DHBV nt position 902) or *Kpn*I (DHBV nt position 1290) and *Bam*HI (polylinker) were removed and replaced with the 1.4-kb *Sac*I-*Xba*I fragment of pON3. p1780-K-*src* was generated by excising the 821-bp *Bgl*II (DHBV nt position 391) to *Xho*I (DHBV nt position 1212) fragment and replacing it with the 800-bp *Mlu*I-*Bgl*II fragment of the chicken c-*src* gene (16).

pD3-ØX118 was constructed by cutting pD3-SP65 with *Nsi*I (DHBV nt position 2845), generating blunt ends by treatment with T4 DNA polymerase, and inserting a 118-bp *Hae*III fragment (nt 4758 to nt 4876) of ØX174 DNA.

Cell culture and transfections. LMH avian hepatoma cells (6) were grown in Dulbecco's minimum essential medium-Ham's nutrient mixture F12 supplemented with 10% fetal calf serum and were passaged every 3 to 4 days at a 1:3 dilution. DNA transfections were performed by the calcium phosphate coprecipitation method exactly as described previously (11).

RNA preparation and RNA analysis. Polyadenylated total cellular RNA was extracted from LMH cells 72 h posttransfection as previously described (11, 17). Encapsidated RNA was isolated from LMH cells 72 h posttransfection by purifying cytoplasmic core particles with polyethylene glycol precipitation followed by proteinase K digestion and phenol extraction as described previously (11, 17). Primer extension analysis of pregenomic RNA with a synthetic 18-base oligonucleotide (spanning positions 2658 to 2640 within the core gene) was carried out as described previously (17). RNase protection analysis of pregenomic RNA and synthesis of [α-^{32}P]CTP-labeled RNA probes was carried out as described previously (13). Probe 442linP was transcribed from plasmid p442linP, which has been described previously (13). Probe lacZ-P108 was transcribed from plasmid placZ-P108. Plasmid placZ-P108 was generated by cloning the 425-bp *Mlu*I-*Mlu*I fragment of pON3 into the *Hin*cII site of pGEM 3Z (Promega Biotec); linearization with *Hin*dIII and transcription with T7 RNA polymerase (Promega Biotec) yields a 475-nt transcript with about 50 nonhybridizing polylinker nucleotides and 425 nt of *lacZ* that are complementary to the fragments of *lacZ* cloned into all of the DHBV/*lacZ* fusion pseudopregenomes. Probe D3-X/R-P was transcribed from plasmid pD3-X/R-P. Plasmid pD3-X/R-P was generated by cloning the 358-bp *Xba*I (DHBV nt

position 2662) to *Eco*RI (DHBV nt position 3020) fragment into the *Hin*cII site of pGEM 3Z such that linearization with *Hin*dIII and transcription with T7 RNA polymerase yields a 414-nt transcript with 56 nonhybridizing polylinker nucleotides. Since this probe is derived from sequences within the terminal redundancy of pregenomic RNA, it is complementary to 358 nt at the 5′ end of pregenomic RNA and 140 nt at the 3′ end of pregenomic RNA.

Note that both the American (19) and European (24) strains of DHBV have been employed in these studies. In experiments where complementation is performed, the wild-type (WT) donor is always of the same strain as the mutant recipient.

RESULTS

Experimental strategy. To define the *cis*-acting site required for DHBV genomic RNA encapsidation, we constructed mutant DHBV genomes and assayed their ability to package genomic RNA. To assay for RNA packaging, the mutant DNAs were transfected into the avian hepatoma cell line LMH (6) by calcium phosphate-mediated transfection. Following transfection, we isolated total cellular poly(A)$^+$ RNA from one-half of each sample of transfected cells; from the other half, we purified cytoplasmic core particles and extracted their nucleic acids. The genomic RNA present in equal portions of each preparation was then quantitated by primer extension or RNase mapping; the ratio of encapsidated to total RNA is a reflection of the packaging efficiency of the mutant pregenome (13, 17).

Several of the mutations we have constructed affect not only the *cis*-acting regions of pregenomic RNA but also the coding regions for the C and/or P gene products, both of which are essential for genomic RNA encapsidation (1, 3, 13). Accordingly, in these cases, the mutant genomes were cotransfected with a WT genome to provide (in trans) the viral proteins required for encapsidation (17). Although this *trans* complementation is less efficient than WT packaging (in which P proteins act preferentially in *cis* to encapsidate their own pregenomic RNAs [13]), it is sufficient to allow analysis of such constructions.

Sequences residing within the terminal redundancy of genomic RNA are required for encapsidation. Our first clue to the location of *cis*-acting encapsidation sequences within the pregenome came from mutants PC1 and PC2 (Fig. 2). Mutants PC1 and PC2 bear, respectively, a single or double (tandem) insertion of an 8-bp *Cla*I linker at an *Acc*I site within the terminal redundancy of the genomic RNA. Accordingly, as depicted in Fig. 2b, these insertions were represented at both ends of the mutant pregenome. Transfection of cells with these mutants results in levels of progeny DNA synthesis that are drastically reduced compared with those of the WT (data not shown). Primer extension analysis of total cellular poly(A)$^+$ RNA from cells transfected with PC1 (Fig. 2a, lane 3) and with PC2 (data not shown) revealed that they accumulate levels of properly initiated genomic RNA equivalent to that of the WT (lane 1; note that the PC1 extension product is 8 nt longer than the WT extension product because of the linker insertion). But when we examined the encapsidated RNA from cells transfected with these mutants, we found that PC1 (lane 4) is significantly impaired compared with the WT (lane 2) in pregenome encapsidation and that encapsidated PC2 RNA is essentially undetectable (data not shown).

Mutants PC1 and PC2 are able to complement core-defective mutants with WT efficiency, and cotransfection

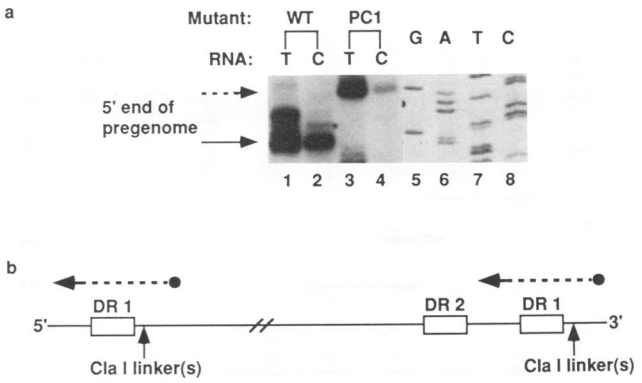

FIG. 2. Primer extension analysis of total and encapsidated pregenomic RNA in transfected cells. (a) Cells were transfected with either WT DHBV DNA (lanes 1 and 2) or PC1 DNA (lanes 3 and 4). From one half of the transfected cells, total cellular poly(A)$^+$ RNA (T) was isolated, and from the other half, cytoplasmic cores (C) were purified and the nucleic acid was extracted. The genomic RNA present in equal portions of each nucleic acid preparation was quantitated by primer extension of an end-labeled oligonucleotide primer complementary to the 5′ region of the C open reading frame. Lanes 5 through 8, sequence ladder of WT DHBV DNA genome generated with the same primer utilized for primer extension. The bands representing the 5′ ends of WT pregenomic RNA and PC1 pregenomic RNA are indicated by a solid arrow and a broken arrow, respectively. The PC1 extension product is 8 nt longer than the WT extension product because of the linker insertion. (b) Schematic representation of mutants PC1 and PC2 and of the strategy for primer extension. Mutants PC1 and PC2 harbor, respectively, single and tandem insertions of an 8-bp *Cla*I linker at the *Acc*I site within the terminal redundancy of pregenomic RNA. Primer extension was performed with a 5′-end-labeled primer (solid circle) complementary to the 5′ region of the C gene and to the identical sequence in the downstream terminal redundancy of the RNA. Extension products other than the authentic 5′ end of pregenomic RNA are due to strong stops to extension from the downstream priming site. The intensity of the extension product that is longer than the authentic 5′ end of pregenomic RNA is greatly decreased in the pool of encapsidated RNA relative to the pool of total cellular poly(A)$^+$ RNA because of the 3′-to-5′ degradation of pregenomic RNA by RNase H during reverse transcription.

with WT genomes ruled out the possibility that precore insertions had created a dominant negative phenotype. Additionally, both mutants can complement a P$^-$ mutant genome as effectively as WT DHBV, indicating that levels of P gene expression are normal for this mutant (data not shown). Since the failure of PC1 and PC2 to encapsidate their genomic RNA is not attributable to aberrant C or P protein synthesis, we conclude that their phenotype is due to a defect in a *cis*-acting signal for encapsidation.

To further define the extent of this signal, we engineered a set of mutants, Δ1 to Δ5, bearing deletions within the region of the genome that comprises the terminal redundancy of genomic RNA (Fig. 3a). Recircularized monomeric genomes bearing these lesions were transfected into LMH cells; thus, in these mutants (as in PC1), the lesions are represented in both ends of pregenomic RNA. Total or encapsidated RNA was then examined by RNase protection (Fig. 3b and c). The uniformly labeled RNA probe used in this RNase protection assay, 442linP, contains a linker not found in pregenomic RNA (Fig. 3c). Therefore, following annealing to genomic RNA and removal of single-stranded regions by RNase treatment, two product fragments, of 430 and 300 nt, will be

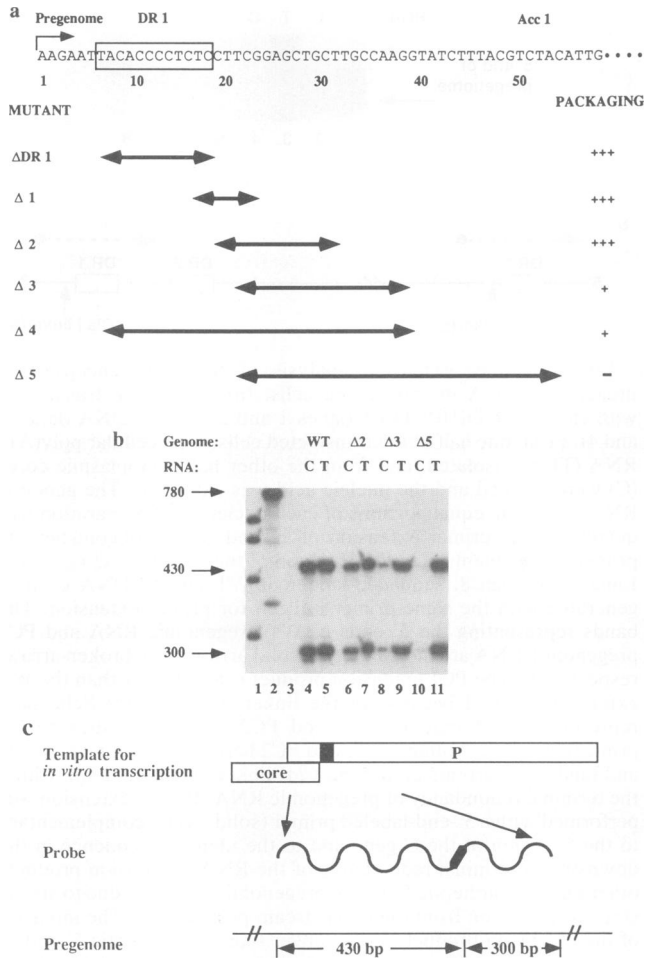

FIG. 3. RNA packaging by mutants bearing deletions in both terminal redundancies. (a) Schematic representation of the 5′ end of pregenomic RNA. The extent of the deletion in mutant genomes ΔDR1 and Δ1 to Δ5 are indicated by the double-headed arrows. Mutant ΔDR1 is a precise deletion of the DR1 sequence in the 5′-terminal redundancy of pregenomic RNA. Since mutants Δ1 to Δ5 bear deletions in the terminal redundancy of the genomic RNA, the lesion is represented twice in the RNA: once in the 5′-terminal redundancy (as depicted in this figure) and once in the 3′-terminal redundancy (data not shown). The encapsidation efficiency of the mutant genomes relative to that of the WT is indicated by + and − signs as follows: +++, equivalent to WT; +, ≤30% of WT; −, no detectable encapsidated RNA. (b) Cells were transfected with WT DHBV DNA or mutant DNA; from one-half of each sample, total cellular poly(A)$^+$ RNA was isolated, and from the other half, cytoplasmic cores were purified and the nucleic acid was extracted. The genomic RNA present in equal portions of each preparation was quantitated by nuclease protection as described for panel c. Lanes: 1, size standards; 2, full-length (780 nt) undigested probe; 3, probe digested in absence of added sample RNA; 4, 6, 8, and 10, fragments protected by core RNA (C) from cells transfected with WT, Δ2, Δ3, and Δ5 DNA, respectively; 5, 7, 9, and 11, fragments protected by total cellular poly(A)$^+$ RNA (T) from cells transfected with WT, Δ2, Δ3, and Δ5 DNA, respectively. (c) Nuclease protection assay. SP6 transcription of p442linP generates a uniformly labeled 780-nt RNA probe (wavy line) that contains a linker (black box) not found in pregenomic RNA. Following annealing to pregenomic RNA and removal of single-stranded regions by nuclease treatment, two product fragments of 430 and 300 nt are generated.

generated. As expected, in WT-transfected cells, abundant quantities of genomic RNA are present in the pool of total cellular poly(A)$^+$ RNA (Fig. 3b, lane 5). Analysis of encapsidated RNA from the same cells confirms that a large fraction of the genomic RNA is found within core particles (lane 4). Similar analysis of deletion mutants Δ2, Δ3, and Δ5 reveals that Δ2 encapsidates its genomic RNA at WT efficiency (lanes 6 and 7), while Δ3 is greatly decreased in packaging efficiency (lanes 8 and 9), and in Δ5 (lanes 10 and 11), encapsidated RNA is nearly undetectable. Deletion mutant Δ1, like Δ2, encapsidates at WT efficiency, and deletion mutant Δ4, like Δ5, is greatly decreased in encapsidation efficiency (data not shown). Furthermore, since deletion mutants Δ1 to Δ5 are able to complement core-defective mutants as efficiently as WT, we conclude that the failure of mutants Δ3, Δ4, and Δ5 to package their genomic RNA at WT efficiency is due to a defect in their *cis*-acting encapsidation signal. These data indicate that sequences within the terminal redundancies of the genomic RNA are required for encapsidation.

Sequences residing in the 5′-terminal redundancy are required for encapsidation. In order to determine whether sequences within one or both of the terminal redundancies are recognized by the packaging apparatus, deletion mutants Δ2 and Δ5 were reconstructed into overlength linear genomes such that the effects of deletions in the 5′ and 3′ copies of the terminal redundancy could be assayed independently (Fig. 4a). LMH cells were transfected with mutants Δ2 or Δ5 bearing deletions in either the 5′-terminal redundancy (Δ2-5′ and Δ5-5′; Fig. 4a) or in the 3′-terminal redundancy (Δ2-3′ and Δ5-3′; Fig. 4a), and total or encapsidated RNA was examined by RNase protection assay with the same uniformly labeled RNA probe, 442linP, depicted in Fig. 3b and c.

Recall that mutant Δ5, bearing a deletion in both the 5′- and 3′-terminal redundancies, is defective for RNA encapsidation. Analysis of mutants Δ5-5′ and Δ5-3′ reveals that both mutants accumulate abundant quantities of genomic RNA in the pool of total cellular poly(A)$^+$ RNA (Fig. 4b, lanes 8 and 10). (The 5′ deletions were cloned as tandem dimers, while the 3′ deletions were cloned as 1.5-mers. For unknown reasons, we consistently observe lower levels of pregenomic RNA in total cellular poly(A)$^+$ RNA in cells transfected with dimeric genomes than in cells transfected with 1.5-mer genomes (compare lanes 4 and 8 with lanes 6 and 10). This is also true for WT DHBV genomes.) Examination of the encapsidated RNA from the same cells reveals that mutant Δ5-5′ (Fig. 4b, lane 7), like mutant Δ5, is defective for RNA encapsidation, while mutant Δ5-3′ (lane 9) encapsidates at WT efficiency. Analysis of encapsidated and total cellular poly(A)$^+$ RNA from mutants Δ2-5′ (lanes 3 and 4) and Δ2-3′ (lanes 5 and 6) reveals that, like that of mutant Δ2, their genomic RNA is encapsidated at WT efficiency. This demonstrates that sequences required in *cis* for pregenome encapsidation lie in the 5′-terminal redundancy within the precore region.

The results obtained from the previous experiments suggested that the DR1 sequence is not required for encapsidation. To explicitly address this, we generated mutant ΔDR1 by engineering a precise deletion of the DR1 sequence within the 5′-terminal redundancy (Fig. 3a). Analysis of encapsidated and total cellular poly(A)$^+$ RNA from cells transfected with ΔDR1 revealed that it is encapsidated at WT efficiency (data not shown). Taken together, the data suggest that the 5′ boundary of the packaging signal lies within the precore region approximately 35 nt downstream of the genomic RNA

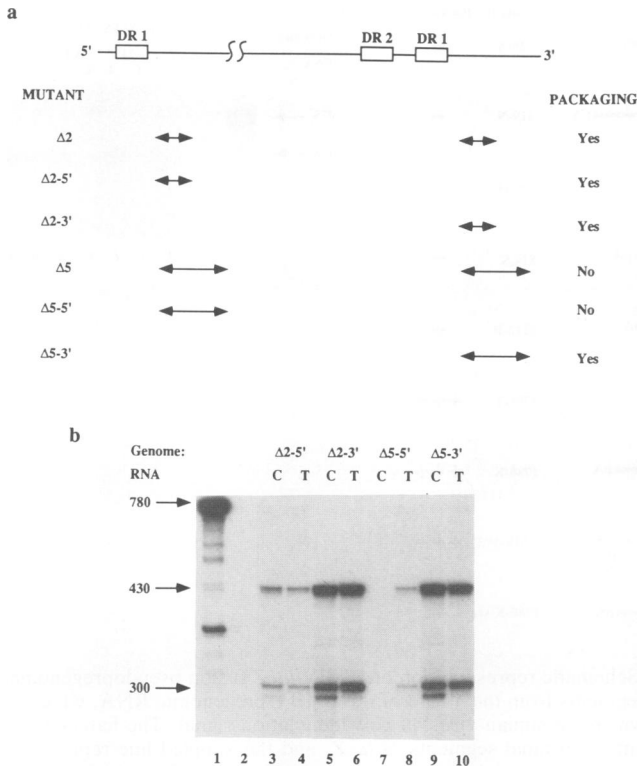

FIG. 4. RNA encapsidation by mutant DNAs bearing deletions in either the 5′ or 3′ copy of the genomic RNA terminal redundancy. (a) Schematic diagram of pregenomic RNA. Deletion mutants Δ2 and Δ5 (Fig. 3a) were reconstructed into cloned overlength genomes bearing the corresponding deletion in either the 5′-terminal redundancy (Δ2-5′ and Δ5-5′) or the 3′-terminal redundancy (Δ2-3′ and Δ5-3′). The position and relative size of the deletions in these mutants are depicted by the double-headed arrows. The ability of the pregenomic RNA of these mutants to be encapsidated is indicated by a Yes or a No. (b) RNase protection analysis was performed exactly as described in the legend for Fig. 3b and c. Lanes: 1, full-length undigested probe (780 nt); 2, probe digested in the absence of added sample RNA; 3, 5, 7, and 9, fragments protected by core RNA (C) from cells transfected with Δ2-5′, Δ2-3′, Δ5-5′, or Δ5-3′ DNA, respectively; 4, 6, 8, and 10, fragments protected by total cellular poly(A)$^+$ RNA (T) from cells transfected with Δ2-5′, Δ2-3′, Δ5-5′, and Δ5-3′ DNA, respectively.

cap site and that sequences 5′ of this position, including DR1, are not required in *cis* for encapsidation.

Sequences distributed over a large segment of the pregenomic RNA are required to mediate encapsidation of a foreign RNA sequence into viral cores. Having defined the 5′ boundary of the packaging signal, we sought to identify the 3′ boundary. To this end, we constructed a series of DHBV/*lacZ* fusion pseudopregenomes (Fig. 5a) and tested their ability to be encapsidated. These are pregenome analogs in which various segments from the 5′ end of the DHBV pregenomic RNA are fused to 3′-terminal fragments of *lacZ*; the constructs initiate at the cap site of pregenomic RNA but are promoted by a CMV-IE promoter and are polyadenylated at a downstream poly(A) site derived from simian virus 40. The pseudopregenomes were constructed such that the total size of the RNA is nearly equivalent to the size of WT DHBV pregenomic RNA. They are designated on the basis of the number of nucleotides of the 5′ end of WT DHBV pregenome contained in the construct. (The constructs are numbered such that nucleotide 1 corresponds to the 5′ end of authentic pregenomic RNA.) Plasmids encoding pseudopregenomes were cotransfected with WT DHBV DNA as a donor for C and P gene products, and their ability to be encapsidated was assayed by nuclease protection using a *lacZ*-specific probe. Annealing of the *lacZ*-specific probe to any of the DHBV/*lacZ* fusions results in protection of a 425-nt fragment. In cells transfected with DHBV/*lacZ* fusions containing 508, 814, 1,212, and 1,396 bp of DHBV sequence, normal quantities of *lacZ*-containing RNA appear in the total cellular poly(A)$^+$ RNA (Fig. 5b, lanes 4, 6, 8, and 10). Analysis of encapsidated RNA from the same transfectants (lanes 3, 5, 7, and 9) reveals that DHBV/*lacZ* fusions containing up to 814 nt of DHBV sequence are not encapsidated (lane 5), fusions with 1,212 nucleotides of DHBV are inefficiently encapsidated (lane 7), and fusions with 1,396 nucleotides of DHBV are encapsidated with WT efficiency (lane 9).

In light of what is known about the encapsidation signal in HBV (15) and in retroviruses (2, 20, 21), we were surprised by the large extent of the DHBV genome required to mediate efficient packaging of a foreign RNA into subviral core particles. We were initially concerned that the additional *lacZ* sequences present in constructs possessing less than 1,396 bp of the 5′ end of the DHBV pregenome might specifically inhibit the encapsidation machinery. To address this concern, we constructed mutant 1780-K-*src* (Fig. 5a), which is analogous to mutant 319-N except that *lacZ* sequences between nt 319 and 1705 have been replaced with a fragment of the same size from the chicken c-*src* gene (1780-K-*src* also retains 78 nt of DHBV sequence between the *src* and *lacZ* sequences). Analysis of total cellular poly(A)$^+$ RNA and encapsidated RNA from cells transfected with 1780-K-*src* reveals that it, like mutant 319-N, is defective for encapsidation (data not shown). Therefore, the failure of chimeric RNAs possessing less than 1,212 nt of DHBV sequence to be encapsidated is not due to an inhibitory effect of *lacZ* sequences but to some defect in the encapsidation signal.

Finally, to determine whether the continuity of the entire 5′ 1,396 nt is required for a functional packaging signal, we interrupted it by inserting a fragment of foreign DNA. Specifically, we inserted a 118-nt fragment of ØX174 DNA into the C gene at position 318 of the pregenomic RNA to yield mutant D3-ØX118 (Fig. 6a). Since this insertion is 3′ of the DHBV polyadenylation signal (represented by the solid diamond in Fig. 6a), it is found only at the 5′ end of pregenomic RNA. To determine whether D3-ØX118 could be encapsidated, it was cotransfected with WT as a donor for the C gene product, and pregenomic RNA in total cellular poly(A)$^+$ RNA and cytoplasmic cores was quantitated by RNase protection. For this experiment, we used a 414-nt uniformly labeled RNA probe, D3-X/R-P, which overlaps the site of insertion of the ØX174 DNA in mutant D3-ØX118. Following annealing to this probe and digestion with RNase, the 5′ end of WT genomic RNA protects an RNA of 358 nt, whereas the 5′ end of D3-ØX118 RNA protects fragments of 183 and 175 nt. Since probe D3-X/R-P overlaps the terminally redundant part of the pregenome, it also anneals to the 3′ ends of both WT and D3-ØX118 pregenomic RNA, resulting in protection of a fragment of approximately 140 nt.

Examination of the total cellular poly(A)$^+$ RNA in cells transfected with WT and D3-ØX118 DNA reveals that both WT (Fig. 6b, lane 5) and mutant (lane 9) genomes are abundantly transcribed and that their RNAs generated protected fragments of appropriate size. As expected, a sub-

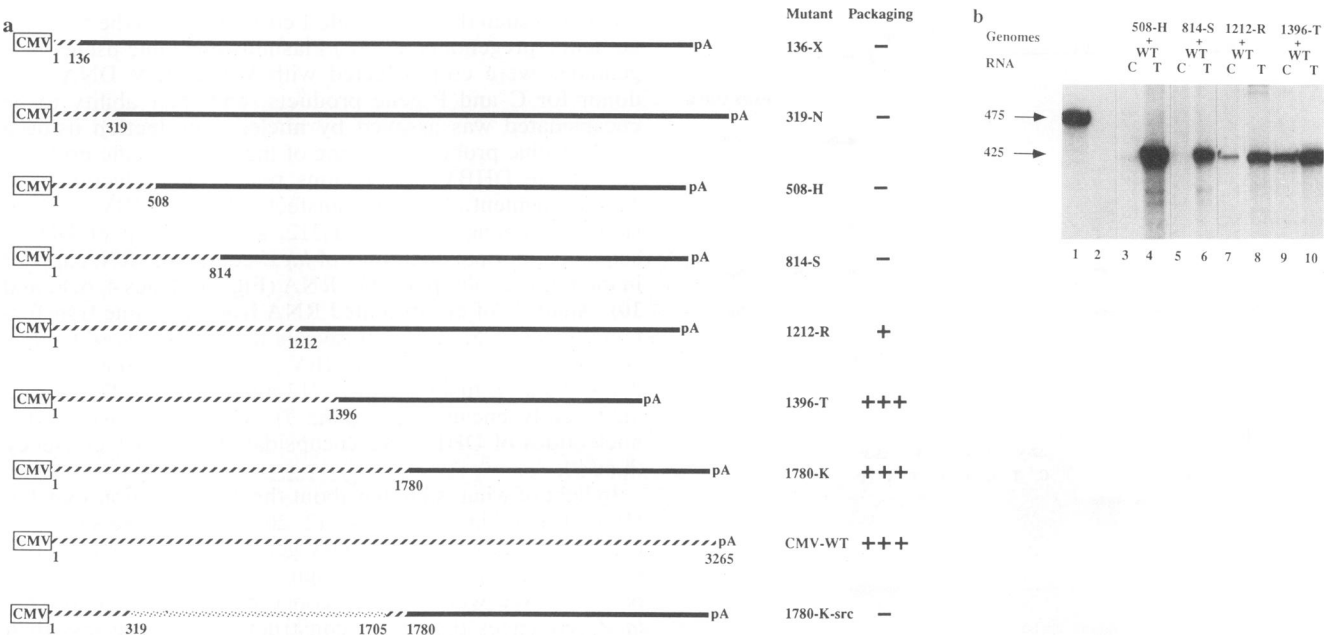

FIG. 5. Encapsidation of DHBV/*lacZ* fusion pseudopregenomes. (a) Schematic representation of DHBV/*lacZ* fusion pseudopregenomes. The pseudopregenomes consist of a CMV-IE promoter driving various segments from the 5′ end of the DHBV pregenomic RNA, which are fused to 3′-terminal fragments of *lacZ* and are polyadenylated by a downstream simian virus 40 polyadenylation signal. The hatched lines represent segments of DHBV pregenomic RNA, the solid lines represent 3′-terminal segments of *lacZ*, and the stippled line represents a segment of the chicken c-*src* gene. The pseudopregenomes are numbered such that nucleotide 1 corresponds to the 5′ end of authentic pregenomic RNA and are designated on the basis of the number of nucleotides of the 5′ end of WT DHBV pregenome that they contain. The encapsidation efficiency of the pseudopregenomes relative to that of the WT is indicated by + and − signs as follows: +++, equivalent to WT; +, ≤30% of WT; −, no detectable encapsidated RNA. (b) RNase protection data. Cells were cotransfected with WT DNA and with DNA encoding pseudopregenomes, and the genomic RNA present in equal portions of preparations of total cellular poly(A)$^+$ RNA and encapsidated RNA were quantitated by nuclease protection with a *lacZ*-specific probe. Annealing of the *lacZ*-specific probe to any of the DHBV/*lacZ* fusions results in protection of 425 nt of the 475-nt full-length *lacZ* probe. Lanes: 1, full-length undigested probe; 2, probe digested in absence of added sample RNA; 3, 5, 7, and 9, fragment protected by core RNA (C) from cells cotransfected with WT DNA plus 508-H, 814-S, 1212-R, and 1396-T DNA, respectively; 4, 6, 8, and 10, fragments protected by total cellular poly(A)$^+$ RNA from cells cotransfected with WT DNA plus 508-H, 814-S, 1212-R, and 1396-T DNA, respectively.

stantial amount of WT RNA was found in cores (lane 4), but no D3-ØX118 RNA was packaged since D3-ØX118 is defective for core biosynthesis (lane 8). Analysis of assembled cores (lane 6) and total cellular poly(A)$^+$ RNA (lane 7) from cells cotransfected with WT and D3-ØX118 DNA reveals that the WT and mutant genomic RNAs are encapsidated with identical efficiency. These data suggest that although a surprisingly large portion of the DHBV genome appears to be required to mediate efficient packaging, it is likely that the *cis*-acting encapsidation signals of DHBV are composed of smaller discontinuous sequences located within this region of the genome.

DISCUSSION

The finding that genomic but not subgenomic viral RNAs are encapsidated suggested early on that structural features of these RNAs mediated their selective recognition and packaging (8). In this article, we have further characterized the *cis*-acting recognition sequences in DHBV. Our results indicate that they function to mediate encapsidation at very high efficiency (~50%) and that they are distributed over a ca. 1,200-nt region at the 5′ end of pregenome, a region that is not present in the subgenomic RNAs.

We have previously shown that a mutant DHBV genome bearing a 2-bp insertion at the *Acc*I site in the precore region is fully infectious and replicates with WT efficiency (5). It therefore came as a surprise that mutants bearing 8-bp and 16-bp insertions at the same *Acc*I site were drastically reduced in their encapsidation efficiency. Interestingly, Junker-Niepmann et al. (15) have pointed out that this particular part of the precore region contains a potential stem-loop structure that is also found at a similar position in the mammalian hepadnaviruses (Fig. 7). As shown in Fig. 7, the *Acc*I site, in boldface, is found within the stem of this putative stem-loop structure. While insertion of 2 bp at this site would likely not significantly alter the stem-loop structure, the 8- and particularly the 16-bp insertions might be expected to more drastically disrupt the structure of this region, particularly that of the bulge region in the 5′ stem. While we have no direct evidence to prove that this structure exists in the RNA, it is interesting to note that mutant Δ2-5′, which deletes nucleotides up to but not including the base of the stem-loop, is encapsidated at WT efficiency (Fig. 4 a and b), while mutants Δ3 and Δ5-5′, which delete increasing amounts of the 5′ half of the stem, have increasingly deleterious effects on encapsidation (see Fig. 7). Further studies are required to investigate the possible role played by this predicted stem-loop structure in pregenome encapsidation.

Recent work of Junker-Niepmann et al. (15) indicates that for HBV, a 137-nt sequence from the 5′ end of pregenomic RNA is sufficient to mediate encapsidation of heterologous

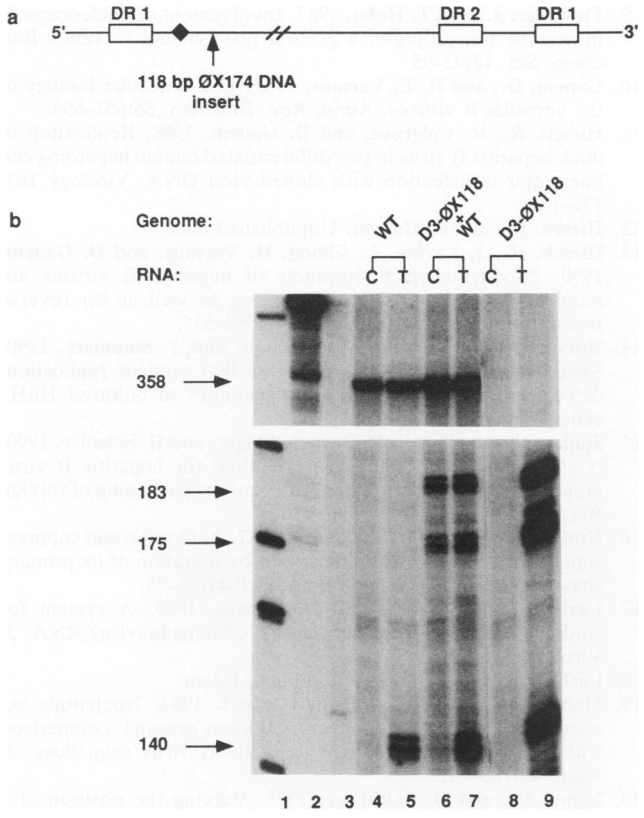

FIG. 6. RNase protection analysis of a mutant with an insertion in the C gene. (a) Schematic representation of pregenomic RNA. Mutant D3-ØX118 contains an insertion of a 118-nt fragment of ØX174 DNA into the C gene at position 318 of the pregenomic RNA. Since this insertion is 3′ of the polyadenylation signal (represented by the solid diamond), it is found only at the 5′ end of the pregenome. (b) Cells were transfected with WT DHBV DNA, D3-ØX118 DNA, or both, and the genomic RNA present in equal portions of preparations of total cellular poly(A)$^+$ RNA and encapsidated RNA were quantitated by nuclease protection using a 414-nt uniformly labeled RNA probe which overlaps the site of insertion of the ØX174 DNA in mutant D3-ØX118. Following annealing to this probe and digestion with RNase, the 5′ end of WT genomic RNA protects a fragment of 358 nt, whereas the 5′ end of D3-ØX118 RNA protects fragments of 183 and 175 nt. Since this probe overlaps the terminally redundant part of the pregenome, it also anneals to the 3′ end of both WT and D3-ØX118 pregenomic RNA, resulting in protection of a fragment of 140 nt. The intensity of the 140-nt protected fragment that corresponds to the 3′ end of pregenomic RNA is greatly decreased in the pool of encapsidated RNA relative to the pool of total cellular poly(A)$^+$ RNA, because of the 3′-to-5′ degradation of genomic RNA by RNase H during reverse transcription. Lanes: 1, molecular size standards; 2, full-length undigested probe; 3, probe digested in the absence of added sample RNA; 4, 6, and 8, fragments protected by core RNA (C) from cells transfected with WT DNA, WT and D3-ØX118 DNA, and D3-ØX118 DNA, respectively; 5, 7, and 9, fragments protected by total cellular poly(A)$^+$ RNA (T) from cells transfected with WT DNA, WT and D3-ØX118 DNA, and D3-ØX118 DNA, respectively.

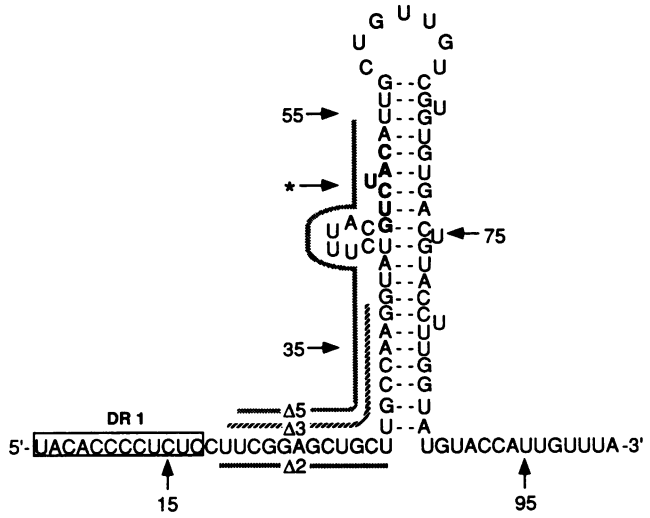

FIG. 7. Potential stem-loop structure in the precore region of DHBV. The sequence is numbered starting at the cap site of the pregenomic RNA as nucleotide 1. The *Acc*I site is in boldface, and the site of insertion of the *Cla*I linker(s) in mutants PC1 and PC2 is designated by the asterisk. The extents of the deletions in mutants Δ2, Δ3, and Δ5 are indicated by the appropriately labeled patterned lines.

RNA into HBV cores. In view of this, we were surprised that sequences distributed over 1,200 nt of the DHBV pregenome seemed to be required to facilitate encapsidation of *lacZ* RNA into DHBV cores. However, while the overall strategies of encapsidation in DHBV and HBV are similar, there is a substantial difference in the efficiency of the process between the two viruses. While DHBV encapsidates 50 to 60% of the pregenomic RNA transcripts it produces, HBV typically encapsidates only about 10% of its pregenomic RNA products (1, 12, 15). It seems likely that DHBV may employ a more complex signal in order to achieve this higher encapsidation efficiency.

How do DHBV sequences participate in genomic RNA packaging? We envision an interaction between this signal and the P gene product, either alone or in association with core antigen (or other proteins). Since the data of Fig. 6 suggest that the packaging signal is discontinuous, it is possible that the P gene product (or a complex containing the P protein) interacts with two or more widely separated sequence elements on the genomic RNA. Alternatively, P protein binding to one site on the genomic RNA may facilitate an alteration of the secondary structure of the RNA such that a second sequence required for packaging is exposed. Such a sequence might also be a target for P or C-P complex binding or might interact with C subunits or other factors required for encapsidation.

When we initially noted the requirement for P protein in encapsidation, we wondered whether the same P-RNA (or C-P-RNA) complex involved in packaging might also be involved in minus-strand initiation (13). But our finding (Fig. 4 and 5) that the essential packaging elements are in the 5′ portion of the RNA makes this simple model unlikely, since minus-strand initiation occurs at the 3′ end of the pregenome (18, 23). Also a complete deletion of DR1 does not impair packaging but strongly affects reverse transcription (14, 18), again implying that the packaging and priming steps are distinct and have different sequence requirements. But it remains possible that these steps could be coupled in some way. For example, proteins bound to the packaging signal might alter the structure of the RNA to expose the reverse transcriptase initiation site at the other end of the molecule.

Our exploration of the *cis*-acting sequences required for DHBV pregenome encapsidation also has implications for

the potential of this virus to serve as a vector for the delivery of foreign genes to the liver. In addition to its obvious hepatotropism and its known ability to produce persistent noncytocidal infections, DHBV encapsidates RNA extremely efficiently; these factors are clearly desirable attributes of a potential vector. On the other hand, the complexity of its encapsidation signal(s) and the limited size of the viral genome may severely limit the amount of foreign DNA that could be carried by the vector. Clearly, further definition of the minimal sequences required for efficient RNA packaging by DHBV will be needed to optimize strategies for gene transfer mediated by this virus.

ACKNOWLEDGMENTS

We thank Roland Russnak for many helpful discussions during the course of this work.

This work was supported by grants from the National Institutes of Health.

REFERENCES

1. **Bartenschlager, R., M. Junker-Niepmann, and H. Schaller.** 1990. The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. J. Virol. **64**:5324–5332.
2. **Bender, M. A., T. D. Palmer, R. E. Gelinas, and A. D. Miller.** 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region. J. Virol. **61**:1639–1646.
3. **Birnbaum, F., and M. Nassal.** 1990. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. J. Virol. **64**:3319–3330.
4. **Buscher, M., W. Reiser, H. Will, and H. Schaller.** 1985. Transcripts and the putative RNA pregenome of duck hepatitis B virus: implications for reverse transcription. Cell **40**:717–724.
5. **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.
6. **Condreay, L. D., C. E. Aldrich, L. Coates, W. S. Mason, and T.-T. Wu.** 1990. Efficient duck hepatitis B virus production by an avian liver tumor cell line. J. Virol. **64**:3249–3258.
7. **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.
8. **Enders, G. H., D. Ganem, and H. E. Varmus.** 1987. 5′-terminal sequences influence the segregation of ground squirrel hepatitis virus RNAs into polyribosomes and viral core particles. J. Virol. **61**:35–41.
9. **Fuetterer, J., and T. Hohn.** 1987. Involvement of nucleocapsids in reverse transcription: a general phenomenon? Trends Biochem. Sci. **12**:92–95.
10. **Ganem, D., and H. E. Varmus.** 1987. The molecular biology of the hepatitis B viruses. Annu. Rev. Biochem. **56**:651–694.
11. **Hirsch, R., R. Colgrove, and D. Ganem.** 1988. Replication of duck hepatitis B virus in two differentiated human hepatoma cell lines after transfection with cloned viral DNA. Virology **167**:136–142.
12. **Hirsch, R., and D. Ganem.** Unpublished data.
13. **Hirsch, R., J. Lavine, L. Chang, H. Varmus, and D. Ganem.** 1990. Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. Nature (London) **344**:552–555.
14. **Horwich, A. L., K. Furtak, J. Pugh, and J. Summers.** 1990. Synthesis of hepadnavirus particles that contain replication-defective duck hepatitis B virus genomes in cultured HuH7 cells. J. Virol. **64**:642–650.
15. **Junker-Niepmann, M., R. Bartenschlager, and H. Schaller.** 1990. A short *cis*-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. EMBO J. **9**:3389–3396.
16. **Kmiecik, T. E., and D. Shalloway.** 1987. Activation and suppression of $pp60^{c\text{-}src}$ transforming ability by mutation of its primary sites of tyrosine phosphorylation. Cell **49**:65–73.
17. **Lavine, J., R. Hirsch, and D. Ganem.** 1989. A system for studying the selective encapsidation of hepadnavirus RNA. J. Virol. **63**:4257–4263.
18. **Loeb, D., and D. Ganem.** Unpublished data.
19. **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.
20. **Mann, R., and D. Baltimore.** 1985. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs. J. Virol. **54**:401–407.
21. **Mann, R., R. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retroviruses. Cell **33**:153–159.
22. **Manning, W. C., and E. S. Mocarski.** 1988. Insertional mutagenesis of the murine cytomegalovirus genome: one prominent alpha gene (ie2) is dispensable for growth. Virology **167**:477–484.
23. **Seeger, C., and J. Maragos.** 1990. Identification and characterization of the woodchuck hepatitis virus origin of DNA replication. J. Virol. **64**:16–23.
24. **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.
25. **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.