JOURNAL OF VIROLOGY, June 1993, p. 3254–3263
0022-538X/93/063254-10$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 67, No. 6

An RNA Stem-Loop Structure Directs Hepatitis B Virus Genomic RNA Encapsidation

JONATHAN R. POLLACK[1] AND DON GANEM[2,3]*

Department of Biochemistry and Biophysics,[1] Howard Hughes Medical Institute,[2] and Departments of Microbiology and Immunology and of Medicine,[3] University of California Medical Center, San Francisco, California 94143-0502

Received 29 December 1992/Accepted 27 February 1993

Selective encapsidation of hepatitis B virus (HBV) genomic RNA within cytoplasmic core particles requires recognition of the *cis*-encapsidation signal, (termed ε) located at the 5′ end of genomic RNA. By transfecting plasmids expressing chimeric RNAs bearing HBV sequences fused to *lacZ*, we have mapped the minimal region of ε to the 5′ 94 nucleotides (nt) of genomic RNA. Enzymatic probing of the RNA secondary structure in this region (by using either in vitro transcripts or RNA extracted from HBV core particles) reveals a stem-loop structure containing a lower stem, a 6-nt bulge, an upper stem with a single unpaired U residue, and a 6-nt loop. The functional role of this structure in encapsidation was explored by examining the effects of mutations in ε on encapsidation of RNA in vivo. These studies reveal that (i) in the lower stem, base pairing but not specific primary sequence is required for function; (ii) there is no requirement for base pairing in the lower portion of the upper stem, but base pairing elsewhere in this stem contributes to packaging efficiency; (iii) the presence of the 6-nt bulge, but not its primary sequence, is important for function; and (iv) specific nucleotide sequences in the loop and in regions of the upper stem are critical for RNA encapsidation.

Hepadnaviruses are small hepatotropic viruses that cause acute and chronic hepatitis and strongly predispose to the development of hepatocellular carcinoma (11). They include the prototype human hepatitis B virus (HBV) and several related viruses of mammals and birds. All hepadnaviruses contain 3- to 3.3-kb DNA genomes whose replication proceeds by reverse transcription of an RNA intermediate (26). Reverse transcription occurs principally within subviral particles (or cores) composed of the viral nucleocapsid protein (C), the polymerase (P), and the RNA template for the reaction. The latter is also called pregenomic RNA (pgRNA), in view of its role as the precursor to the DNA form of the genome. Thus, these DNA viruses encapsidate their genome as RNA, and this encapsidation reaction is an essential early step in the pathway of genomic replication.

The RNA packaging reaction displays remarkable selectivity for pgRNA. HBV-infected cells produce a variety of viral transcripts of both genomic and subgenomic lengths. The genomic RNAs include the true pgRNA as well as minor additional transcripts with start sites up to 31 nucleotides (nt) 5′ to the pgRNA cap site (9, 28). These RNAs, otherwise identical in structure to pgRNA, are not encapsidated (10); they function only as mRNAs for a secreted variant of the C protein known as preC. The subgenomic mRNAs (of 2.4, 2.1, and possibly 0.7 kb) are 3′ coterminal with pgRNA but also are not packaged into cores (10).

Selective encapsidation of pgRNA is now known to be dependent on both C and P gene products (2, 14, 21, 23) and on a *cis*-acting encapsidation signal, termed ε, on pgRNA. This site has been localized to an 85- to 137-nt stretch near the 5′ end of the HBV pgRNA transcript (2, 16), while a much larger region is required to direct efficient pgRNA encapsidation in the avian hepadnaviruses (15). The ε sequences are also present on preC mRNAs, but ribosomes translating the preC region (which overlaps with ε) apparently preclude the functioning of the ε sequences to direct encapsidation of these transcripts (22). Because of the terminal redundancy of the pgRNA transcript and the coterminal nature of all viral transcripts, the ε sequences are also present at the 3′ ends of pgRNA and subgenomic RNAs but do not direct encapsidation from this position (15); the basis of this positional effect is not understood.

Although computer-based secondary-structure prediction and phylogenetic analyses have suggested potential secondary structures in the ε region (16), the existence of such structures has not been demonstrated and their potential roles in encapsidation are unknown. In this paper, we further define the minimal ε element, determine the secondary structure of this RNA region, and explore the functional role of this RNA structure in encapsidation in vivo.

MATERIALS AND METHODS

Materials. Restriction enzymes were purchased from New England BioLabs and used as specified by the manufacturer. RNase T_1 and RNase T_2 were purchased from Bethesda Research Laboratories, Inc.; RNase CV_1 was purchased from Pharmacia; and avian myeloblastosis virus reverse transcriptase was purchased from Seikagaku America, Inc. Radionucleotides were purchased from Amersham Corp.

Plasmids. All HBV nucleotide positions are numbered from the unique *Eco*RI site of HBV adw2 (27). In this numbering scheme, nt 1815 is the transcription initiation site of pgRNA.

Helper plasmid pCMV-CP was constructed by inserting HBV sequences from the *Sty*I site (nt 1880) to the *Taq*I site (nt 2012) into the polylinker of pCDNA1 (Invitrogen), downstream of the cytomegalovirus (CMV) immediate-early promoter. pCMV-CP contains the *env*(−) termination mutation in the S gene (6) and therefore does not express HBV envelope proteins.

pE-BS was constructed by cloning the 172-nt polymerase chain reaction-amplified fragment containing nt 1815 to 1986

* Corresponding author.

of HBV (with linearized HBV monomer as a template) into the *Hin*dIII and *Bam*HI sites of pBS(−) (Stratagene). pLacZ was constructed from pON249 (12) by removing a 364-nt *Pvu*II fragment within the *lacZ* gene so that the resultant CMV promoter-driven *lacZ* transcript approximates the length of HBV pgRNA. The *Hin*dIII-*Kpn*I (*Kpn*I site in polylinker) HBV fragment of pE-BS was inserted to replace the *Hin*dIII-*Kpn*I fragment at the 5′ end of *lacZ* in pLacZ to generate pE-LacZ. In pE-LacZ, transcription from the CMV immediate-early promoter initiates at nt 1815 of the HBV sequence.

The other HBV-*lacZ* fusions, used in mapping the limits of ε (see Fig. 3), were created by cloning the appropriate polymerase chain reaction-amplified fragment into the *Hin*dIII-*Kpn*I sites of pLacZ. pb-LacZ, pc-LacZ, pd-LacZ, pe-LacZ, and pf-LacZ contain HBV nt 1842 to 1986, 1842 to 1927, 1829 to 1986, 1815 to 1927, and 1815 to 1908, respectively. pDHBV-LacZ was created by cloning the polymerase chain reaction-amplified fragment containing the 5′ 172 nt of duck HBV type 3 (DHBV3) (25) pgRNA into the *Hin*dIII-*Kpn*I sites of pLacZ.

Small deletions and all nucleotide changes in the stem-loop structure sequences (see Table 1) were constructed by site-directed mutagenesis by standard methods (18). First, the mutation was introduced into pE-BS. After confirmation by DNA sequencing, the *Hin*dIII-*Kpn*I fragment containing the mutated stem-loop sequence was cloned into the *Hin*dIII-*Kpn*I sites of pLacZ.

Computer prediction of secondary structure. Prediction of the RNA secondary structure for wild-type and mutant constructs was performed with the software of Martinez (20) and Abrahams et al. (1).

In vitro structure determination. RNA was transcribed in vitro from a *Sma*I-linearized pE-BS template by using T3 RNA polymerase. The resultant RNA contains the 5′ 172 nt of HBV genome plus 14 nt of polylinker upstream and 2 nt of polylinker downstream. The RNA was gel purified by electrophoresis on a 4% (wt/vol) (3% NuSieve, 1% SeaPlaque [FMC BioProducts]) agarose gel in Tris-acetate buffer and recovered by phenol extraction. For structure mapping of mutated RNAs, the relevant mutation introduced into pE-BS was used as a template.

RNA renaturation and enzymatic digestion with RNase T_1, RNase T_2, and RNase CV_1 were performed exactly as described by Skinner et al. (24). The RNA from one-fifth of each digestion reaction was used for the primer extension reaction, performed exactly as described by Black and Pinto (5) with a [γ-^{32}P]ATP-labeled oligonucleotide primer (5′-GAAGTCAGAAGGCAAAAACG-3′) complementary to nucleotides at the 3′ end of the RNA. Primer extension products were then separated on a 6% denaturing polyacrylamide gel and detected by autoradiography.

For the enzymatic structure determination of RNA encapsidated within core particles, RNA extracted from core particles purified from half of a 100-mm plate was used for enzymatic digestion as described above.

Cell culture and transfections. HepG2 human hepatoma cells were grown in HME16 medium supplemented with 10% fetal calf serum, 0.14% sodium bicarbonate, and 2 mM glutamine and passaged every 2 to 3 days at a 1:3 dilution. DNA transfections were performed by the calcium phosphate coprecipitation method exactly as described previously (13).

RNA preparation and RNase protection assay. Poly(A)$^+$ total-cell RNA was purified 72 h posttransfection as previously described (13, 19). RNA within cytoplasmic core particles was isolated 72 h posttransfection by polyethylene glycol precipitation of core particles followed by proteinase K digestion, phenol extraction, and precipitation of nucleic acid as previously described (13, 19). RNase protection analysis was performed as previously described (14) on poly(A)$^+$ RNA or core RNA prepared from equivalent numbers (half of a 100-mm plate) of transfected HepG2 cells. [α-^{32}P]CTP-labeled RNA probes were synthesized as previously described (14); probe lacZ-P108 (15) is a 475-nt riboprobe with about 50 nonhybridizing polylinker nucleotides and 425 nt of *lacZ* complementary to the *lacZ* fragment contained in pLacZ. The intensity of RNase-protected bands was quantitated by using a phosphorimager (Molecular Dynamics). At least two independent transfections and RNase protection assays were performed for each mutant construct tested.

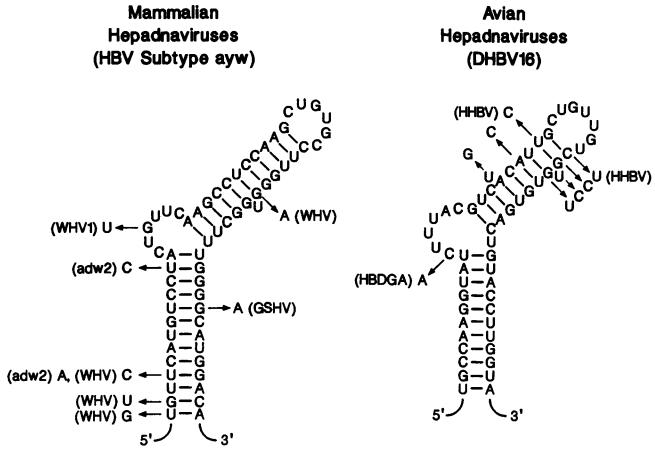

FIG. 1. Phylogenetic analysis of the ε region showing the RNA secondary structure proposed for the mammalian and avian hepadnaviruses based on computer prediction and phylogenetic analysis. HBV subtype ayw and DHBV16 sequences are shown, with nucleotide changes found in selected hepadnavirus isolates indicated by arrows and the corresponding hepadnavirus isolate indicated in parentheses. HBDGA is a DHBV isolate. All HBV, DHBV, woodchuck hepatitis virus (WHV), and ground squirrel hepatitis virus (GSHV) sequences were obtained from GenBank. HHBV, heron HBV.

RESULTS

Computer and phylogenetic analysis of the ε region. Computer-based secondary-structure prediction suggests several potential secondary structures within the region of the HBV encapsidation signal, ε. One plausible structure, first pointed out by Junker-Niepmann et al. (16) and shown in Fig. 1, is supported by phylogenetic comparisons. As shown in Fig. 1, despite significantly divergent primary sequences, grossly similar secondary structures can be drawn for both the mammalian and avian hepadnaviruses. Note, however, that the upper stem is less well conserved within the avian grouping. In the heron HBV, as indicated by the nucleotide substitutions in Fig. 1, and in two infectious DHBV isolates (DHBV-S5 and DHBV-S31; not shown), there are nucleotide changes which in each case disrupt three potential base pairs within the upper stem. These findings emphasize the need for direct experimental characterization of the RNA structure in this region and its relation to RNA packaging in vivo.

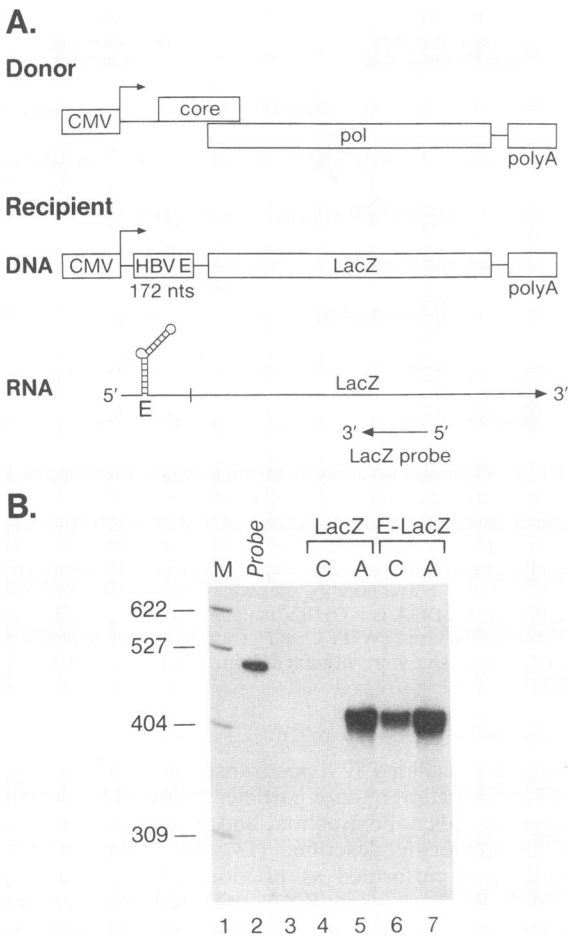

FIG. 2. Encapsidation assay. (A) Schematic representation of the helper plasmid pCMV-CP (Donor), which uses the CMV immediate-early promoter to drive the expression of HBV C and P proteins, and plasmid pE-LacZ (Recipient), from which is transcribed a chimeric RNA bearing ε (HBV E) fused to *lacZ*. (B) RNase protection analysis of total poly(A)$^+$ and encapsidated RNA. From equivalent numbers of transfected cells, we prepared either total poly(A)$^+$ RNA or RNA extracted from purified cytoplasmic core particles. RNA from equal portions of each preparation was quantified by RNase protection with a LacZ riboprobe schematically depicted by the arrow. When annealed to complementary *lacZ* sequences and digested with RNase, this probe generates a 425-nt protected fragment. Lanes: 1, size standards; 2, undigested probe; 3, probe digested in the absence of added RNA sample; 4 and 6, core RNA (C) from cells cotransfected with helper plasmid pCMV-CP along with pLacZ and pE-LacZ, respectively; 5 and 7, total poly(A)$^+$ RNA (A) from cells cotransfected with helper plasmid pCMV-CP along with pLacZ and pE-LacZ, respectively. Lanes 4 to 7 are a reproduction of a section of the autoradiograph (lanes 1 to 4) in Fig. 3.

Encapsidation of heterologous RNA directed by the HBV ε region. To study the function of the packaging signal in vivo, we used an assay based on the ability of ε sequences to mediate the encapsidation of heterologous RNAs. HepG2 hepatoma cells were transfected with plasmid pE-LacZ (Fig. 2A), which expresses a chimeric RNA of genomic length bearing the 5′ 172 nt of HBV pgRNA (which should contain the entire ε region [16]) fused to *lacZ*. This RNA is encapsidated by C and P proteins supplied in *trans* from helper plasmid pCMV-CP (Fig. 2A), which does not contain ε and which donates P protein efficiently (2). Following transfection, we harvested from equivalent numbers of transfected cells either the total-cell poly(A)$^+$ RNA or the RNA contained within purified cytoplasmic core particles. The RNA present in equal portions of each preparation was quantified by RNase protection with a riboprobe derived from *lacZ* sequences; the ratio of encapsidated to total poly(A)$^+$ RNA is a measure of the packaging efficiency.

Figure 2B, lanes 4 and 5, shows the results from a cotransfection of the helper plasmid and plasmid pLacZ, which expresses only *lacZ* sequences. The *lacZ* RNA accumulates in the cytoplasm, since we found a protected band of the expected size (425 nt) in the sample containing poly(A)$^+$ RNA (lane 5). However, this RNA is not encapsidated, since no protected band was found in the sample of RNA extracted from core particles (lane 4). By contrast, transfection of the chimeric plasmid pE-LacZ bearing the ε region fused to *lacZ* results in efficient RNA encapsidation within cores (lanes 6 and 7). The efficiency of encapsidation (ca. 50% of the total RNA) is similar to that of wild-type HBV pgRNA, indicating that all the relevant *cis*-acting information is contained within these 172 nt of HBV.

Mapping the limits of ε. To better define the limits of the ε region, and in particular to determine whether HBV sequences containing only the proposed stem-loop structure are sufficient to direct encapsidation, we generated the constructs shown in Fig. 3A. These constructs were cotransfected with helper plasmid pCMV-CP, and levels of encapsidated RNA were determined by RNase protection (Fig. 3B). As before, the 5′ 172-nt region directs efficient packaging of *lacZ* RNA sequences (pE-LacZ [construct a]; lanes 3 and 4), whereas the RNA with *lacZ* alone is not encapsidated (lanes 1 and 2). Among these constructs, construct f (lanes 13 and 14) represents the smallest HBV region sufficient to direct RNA encapsidation with nearly wild-type efficiency. Thus, ε is contained within the 5′ 94 nt of pgRNA found in construct f; this includes the proposed stem-loop structure and the 30 nt upstream of it. Constructs b, c, and d (lanes 5 to 10) harbor deletions of various extents upstream of the stem-loop and are packaged 5- to 10-fold less efficiently, indicating that sequences in this region, although not absolutely essential for encapsidation, contribute measurably to its efficiency.

Since the proposed stem-loop structure is conserved phylogenetically, we sought to determine whether a similar structure with a different primary sequence could still be packaged by HBV C and P proteins. We therefore tested construct g (pDHBV-LacZ), which contains the putative DHBV stem-loop structure fused to *lacZ* RNA. As shown in Fig. 3B (lanes 15 and 16), this RNA is not encapsidated by HBV C and P proteins.

Probing the structure of ε in RNA synthesized in vitro. To obtain direct biochemical evidence for the presence of secondary structure within ε, we probed the structure of this region in RNA transcripts generated in vitro. A 188-nt RNA containing the ε region was transcribed in vitro, self-annealed, and then subjected to cleavage by RNases specific for either single-stranded or base-paired regions. The positions of cleavage sites were then determined by primer extension with an end-labeled primer homologous to sequences at the 3′ end of the RNA. Figure 4B, lanes 1 to 4, shows typical results obtained by cleavage with RNase T_2, an RNase specific for single-stranded regions (8); products of primer extension on these cleaved RNAs are run next to the corresponding sequencing ladder generated from the same

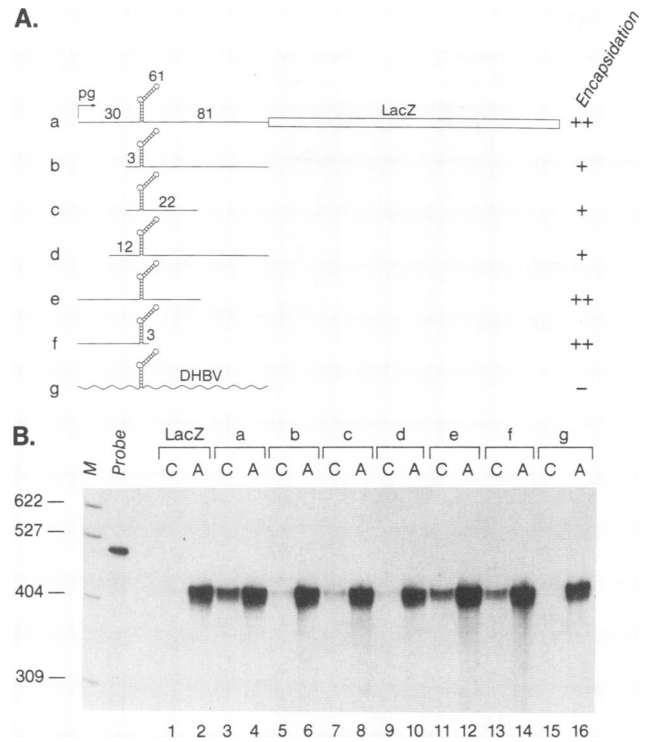

FIG. 3. Mapping the limits of ε. (A) Schematic representations of the chimeric plasmid series containing various extents of the 5′ region of HBV pgRNA fused to *lacZ*. pg indicates transcription initiation from the authentic pgRNA start site. The number of nucleotides contained in each construct is indicated by the numbers above the constructs. pE-LacZ appears here as construct a. pDHBV-LacZ (construct g) contains the 5′ sequences of DHBV pgRNA. The encapsidation efficiency of constructs compared with pE-LacZ (wild type) is indicated as follows: ++, 50 to 100% of wild type; +, 10 to 50% of wild type; +/−, 1 to 10% of wild type; −, no detectable encapsidated RNA. (B) RNase protection analysis of total poly(A)$^+$ and encapsidated RNA. Methods are identical to those described in the legend to Fig. 2B. Lanes: M, size standards; Probe, undigested probe; 1, 3, 5, 7, 9, 11, 13, and 15, core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ (a), pb-LacZ (b), pc-LacZ (c), pd-LacZ (d), pe-LacZ (e), pf-LacZ (f), and pDHBV-LacZ (g), respectively; lanes 2, 4, 6, 8, 10, 12, 14, and 16, poly(A)$^+$ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pb-LacZ, pc-LacZ, pd-LacZ, pe-LacZ, pf-LacZ, and pDHBV-LacZ, respectively.

primer (lanes 5 to 8). Increasing the concentration of the enzyme yields two sets of prominent bands (lanes 1 to 3) not found in the absence of enzyme (lane 4), representing positions of single-stranded nucleotides. A single structure is accommodated by this information; this structure is identical to that predicted by phylogenetic analysis and is summarized in Fig. 4A (the RNase T_2 cleavage sites are indicated by arrows in Fig. 4A). The two sets of prominent bands correspond to cleavages in the loop and bulge regions. Digestion with RNase T_1, which specifically cleaves single-stranded RNA after G residues (8), yields cleavages only after G residues present in the loop and bulge (data not shown), in close agreement with the RNase T_2 results.

Figure 4B also shows the results of similar experiments with RNase CV_1 (lanes 9 to 12), which is specific for base-paired or stacked regions of RNA (8). Two prominent sets of bands appear, corresponding to cleavages in the upper and lower stems of this structure (depicted as arrowheads in Fig. 4A).

The in vitro transcript used in these experiments contains 14 nt of polylinker sequences upstream of the HBV sequences. To exclude the possibility that these extraneous sequences altered the folding of the RNA, 12 of these nucleotides were deleted from the DNA template by oligonucleotide mutagenesis (leaving the 5′ GG required for efficient transcription by T3 RNA polymerase); the RNase cleavage patterns on this transcript are identical to those of its parent (data not shown). Therefore, the RNA corresponding to the ε region synthesized in vitro adopts the stable secondary structure depicted in Fig. 4A, with a lower stem, a 6-nt bulge, an upper stem with a single unpaired U residue, and a 6-nt loop.

Structure of ε in RNA encapsidated within core particles. We next sought to determine whether the secondary structure observed in transcripts generated in vitro actually forms in vivo. ε-*lacZ* chimeric RNA was therefore extracted from cytoplasmic core particles purified from HepG2 cells cotransfected with pE-LacZ and helper plasmid pCMV-CP. This deproteinized RNA was directly subjected to cleavage by RNase T_2 (Fig. 5, lane 6) and RNase CV_1 (Fig. 5, lane 7), without prior denaturation and refolding. The characteristic patterns of cleavage sites for the loop and bulge (lane 6) and for the upper stem and lower stem (lane 7) are evident. Although we cannot rule out the existence of higher-order structures resulting from RNA-protein interactions in vivo, our data strongly suggest that the stem-loop secondary structure, as depicted in Fig. 4A, exists in vivo in RNAs known to have been packaged.

Mutational analysis of stem-loop structure and function. To obtain genetic evidence for the existence of the stem-loop structure in vivo and to explore its functional role in RNA packaging, we next introduced numerous mutations into the ε region of chimeric ε-*lacZ* constructs and assayed encapsidation in vivo, using the transfection assay described in the legend to Fig. 2. The exact nucleotide sequences of these mutations are detailed in Table 1; the mutational lesions are also schematically depicted above the results of their corresponding packaging assays in Fig. 6, 7, and 9.

(i) **Stem mutations.** First we examined the functional effects of gross deletions within the stem-loop structure (Fig. 6). Complete deletion of either the upper stem (pΔupper [lanes 11 and 12]) or the lower stem (pΔlower [lanes 13 and 14]) region abolishes RNA packaging. Therefore, the encapsidation function is not localized to one subdomain of the structure.

We next introduced less radical mutations into the lower stem to determine the importance of base pairing there for RNA packaging (Fig. 6). When we changed four consecutive nucleotides in either the left (pLowerL [lanes 5 and 6]) or right (pLowerR [lanes 7 and 8]) side of the stem to disrupt base pairing in this region, in each case RNA was no longer packaged. In the doubly mutated construct (pLowerL/R [lanes 9 and 10]), which contains the changes in both the left and right sides of the stem to restore base pairing in this region, RNA became encapsidated, albeit at three- to four-fold-reduced levels compared with the wild-type construct (pE-LacZ [lanes 3 and 4]). This confirms that the lower stem of the structure forms in vivo and indicates that base pairing in this region is a structural requirement for encapsidation.

We next introduced mutations to test the role of base pairing in the upper stem for RNA packaging (Fig. 7A). When we altered three consecutive residues in the lower left side of this stem to disrupt base pairing, RNA was still

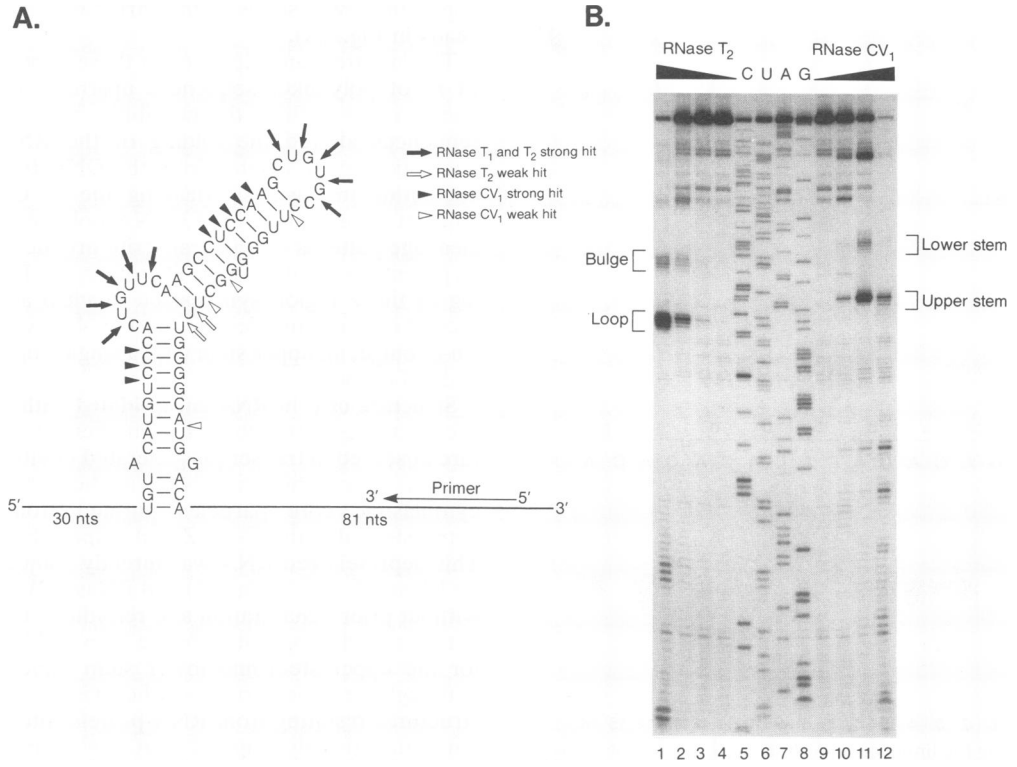

FIG. 4. Enzymatic probing of ε secondary structure in RNA synthesized in vitro. (A) Schematic representation of the RNase cleavage sites drawn onto the RNA secondary structure of the ε region as derived from enzymatic probing. (B) A 188-nt RNA spanning the ε region was transcribed in vitro, renatured, and cleaved with RNases specific for single-stranded (RNase T_1 and RNase T_2) or base-paired (RNase CV_1) regions. The positions of cleavage sites were determined by primer extension from an end-labeled primer homologous to sequences at the 3′ end of the RNA. Lanes: 1, 2, 3, and 4, RNase T_2 digestion with 60, 6.0, 0.6, and 0 U/ml, respectively; 5 to 8, sequencing ladder generated from the same primer; 9, 10, 11, and 12, RNase CV_1 digestion with 0, 0.25, 2.5, and 25 U/ml, respectively. Positions of nucleotide cleavage corresponding to the upper and lower stems, the loop, and the bulge are indicated alongside the lanes.

encapsidated (pUpper1L[lanes 5 and 6]) at levels only two- to threefold lower than in the wild-type construct (lanes 3 and 4); this suggests that base pairing in this region of the upper stem is not essential for encapsidation. When we altered three consecutive bases in the lower right side of this stem to disrupt base pairing, RNA was no longer encapsidated (pUpper1R [lanes 7 and 8]); given the lack of importance of base pairing here, this suggests that these specific nucleotides are critical for encapsidation. The doubly mutated construct, which contains both sets of changes to restore base pairing, did not direct RNA encapsidation (pUpper1L/R [lanes 9 and 10]); this is consistent with the fact that changing the bases in the right side is not tolerated. To prove that base pairing is indeed restored in the doubly mutated construct, we enzymatically probed the structure of these three mutated RNAs synthesized in vitro (Fig. 8). In each case the structure behaved as predicted for the introduced changes: mutating the left side (lane 12) or right side (lane 14) of the stem disrupted the structure locally, whereas the wild-type structure was restored in the doubly-mutated construct (lane 16).

A similar set of mutations was introduced in the upper region of the upper stem (Fig. 7B). When four residues were altered in the left side (pUpper2L [lanes 5 and 6]) or in the right side (pUpper2R [lanes 7 and 8]) to disrupt base pairing, in each case RNA was encapsidated at 4- to 10-fold-reduced levels compared with the wild-type construct (lanes 3 and 4). In the doubly mutated construct (pUpper2L/R [lanes 9 and 10]), which contains both sets of mutations and should restore base pairing, RNA was encapsidated at nearly wild-type levels. This reaffirms the existence of the upper stem in vivo and suggests that, although base pairing in the lower portion of this stem is dispensable for packaging (Fig. 7A), base pairing in its upper region increases packaging efficiency, though it is not absolutely essential.

Interestingly, when we altered the uppermost five consecutive residues of the upper stem (one more than was altered in the pUpper2 constructs of Fig. 7B), RNA was no longer packaged in either of the singly mutated constructs, and restoration of base pairing in the corresponding doubly mutated construct did not restore encapsidation (data not shown). Given that in the doubly mutated pUpper2L/R construct of Fig. 7B, four consecutive base changes that maintain base pairing are tolerated, this finding suggests that the specific nucleotides of the uppermost base pair are a primary sequence requirement for encapsidation; i.e., they cannot be altered.

(ii) Loop and bulge mutations. In the constructs of Fig. 9A, mutations were introduced to determine the functional role of specific sequences in the loop and bulge regions of the ε structure. When the first four nucleotides in the loop were changed, RNA was no longer encapsidated (pLoop1–4 [lanes 5 and 6]). In contrast, when the same four base changes were introduced in the bulge, RNA was encapsidated (pBulge1–4 [lanes 7 and 8]) at an efficiency near that of the wild-type construct (lanes 3 and 4). Changing the last two bases in the

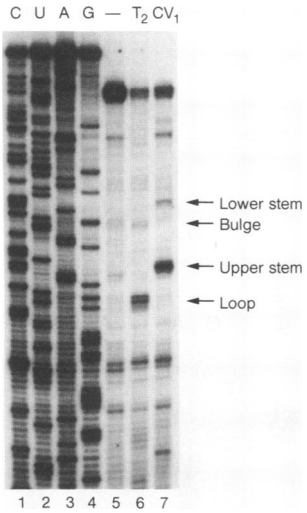

FIG. 5. Enzymatic probing of ε secondary structure in RNA from core particles. ε-*lacZ* chimeric RNA was extracted from core particles purified from HepG2 cells cotransfected with pE-LacZ and helper plasmid pCMV-CP. RNA was directly subjected to cleavage by RNases, and the positions of cleavage sites were determined by primer extension from an end-labeled primer homologous to sequences at the 3′ end of the RNA. Lanes: 1 to 4, sequencing ladder generated from the end-labeled primer; 5, 6, and 7, core RNA digested with no enzyme, 600 U of RNase T_2 per ml, and 25 U of RNase CV_1 per ml, respectively. Positions of nucleotide cleavage corresponding to the upper and lower stems, the loop, and the bulge are indicated alongside the lanes.

bulge also did not reduce encapsidation efficiency (data not shown). However, when the bulge was deleted entirely, RNA was not packaged (pΔbulge [lanes 9 and 10]). To rule out the possibility that the above mutations altered the overall structure of the stem-loop, we again examined the structures of these RNAs synthesized in vitro, by using nuclease accessibility experiments (Fig. 8). In each case the structure behaved as predicted for the introduced changes: base changes in the loop (Fig. 8, lane 8) and bulge (Fig. 8, lane 10) preserved the overall structure, and deletion of the bulge eliminated the bulge without having other gross effects on the stems or loop (data not shown). Thus, it appears that the primary sequence in the loop is important for encapsidation; by contrast, the bulge plays principally a structural role, with its primary sequence being relatively unimportant.

Since specific primary sequences in the loop appear to be important for encapsidation, we next mutated each of the six nucleotides in the loop individually, making primarily purine-to-purine and pyrimidine-to-pyrimidine changes (Fig. 9B). Alterations at either of the first two positions of the loop had no effect on encapsidation efficiency (lanes 5 and 6 and lanes 7 and 8); however, changing the base at the third position abolished packaging (lanes 9 and 10), and alterations at the fourth or fifth position diminished the level of encapsidated RNAs over 20-fold (lanes 11 and 12 and lanes 13 and 14). Altering the nucleotide at the sixth position (lanes 15 and 16) also reduced encapsidation efficiency, but less dramatically (ca. threefold). Therefore, specific nucleotides in the loop appear to play a crucial role in encapsidation.

Finally, we have examined the role of the single unpaired U residue in the upper stem in encapsidation (Fig. 9A). Deletion of this residue (pΔU [lanes 11 and 12]) reduced the level of packaged RNAs at least 10-fold compared with the wild-type construct (lanes 3 and 4); this demonstrates the

TABLE 1. Stem-loop mutations

Mutation	Sequence[a] (Lower stem / Bulge / Upper stem / Loop / Upper stem / Lower stem)	Encapsidation efficiency[b]
WT	UGUACAUGUCCCACUGUUCAAGCCUCCAAGCUGUGCCUUGGGUGGCUUUGGGGCAUGGACA	++
Δupper	----------.....----------	−
Δlower	.....----------.....	−
LowerL	.....CAGG.....	−
LowerR	.....CCUG.....	−
lowerL/R	.....CAGG.....CCUG.....	+
Upper1L	.....UCG.....	+
Upper1R	.....CGA.....	−
Upper1L/R	.....UCG.....CGA.....	−
Upper2L	.....GGUU.....	+
Upper2R	.....AACC.....	+
Upper2L/R	.....GGUU.....AACC.....	++
Loop1-4	.....UCUG.....	−
Bulge1-4	.....UCUG.....	++
Δbulge	.....------.....	−
ΔU	.....-.....	+/−
Loop1	.....U.....	++
Loop2	.....A.....	++
Loop3	.....A.....	−
Loop4	.....C.....	+/−
Loop5	.....A.....	+/−
Loop6	.....U.....	+

[a] The primary sequence of the wild-type (WT) stem-loop region is shown at the top, annotated above by arrows indicating the sequence components of the upper stem, lower stem, loop, and bulge. Subsequent rows show the mutant designations along with their nucleotide changes. Symbols: ., no change of nucleotide; −, deletion of a nucleotide.

[b] The encapsidation efficiency of constructs compared with pE-LacZ (wild type [WT]) is indicated as follows: ++, 50 to 100% of wild type; +, 10 to 50% of wild type; +/−, 1 to 10% of wild type; −, no detectable encapsidated RNA.

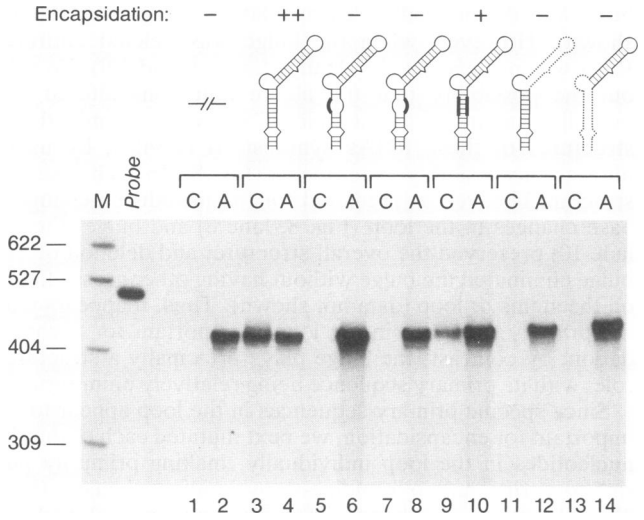

FIG. 6. Encapsidation assay of RNAs with lower-stem mutations. Methods are exactly as described in the legend to Fig. 2B. Lanes: M, size standards; Probe, undigested probe; 1, 3, 5, 7, 9, 11, and 13, core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLowerL, pLowerR, pLowerL/R, pΔupper, and pΔlower, respectively; 2, 4, 6, 8, 10, 12, and 14, poly(A)$^+$ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLowerL, pLowerR, pLowerL/R, pΔupper, and pΔlower, respectively. The encapsidation efficiency of constructs compared with pE-LacZ (wild type) is indicated as follows: ++, 50 to 100% of wild type; +, 10 to 50% of wild type; +/−, 1 to 10% of wild type; −, no detectable encapsidated RNA.

importance of the unpaired U in encapsidation. The integrity of the overall stem-loop structure in this mutated RNA was confirmed by enzymatic probing of the RNA synthesized in vitro (data not shown). We have not changed the unpaired U to other residues (to determine whether it plays primarily a structural role or is recognized specifically), because other bases at this position are predicted by computer secondary-structure analysis to alter the base-pairing partners within the upper stem.

DISCUSSION

In this study we demonstrate that the 5′ 94 nt of HBV pgRNA that defines the ε region form a stable stem-loop structure in vitro and in vivo and that this structure is functionally important for encapsidation. Mutational analysis leads to a picture of the functional encapsidation signal that includes (i) structural requirements for base pairing in the lower stem and the upper regions of the upper stem and for the presence of the bulge (these are relatively independent of the primary sequence in these regions); (ii) a requirement for a single unpaired U residue in the upper stem; and (iii) a requirement for specific nucleotide sequences in the loop and in regions of the upper stem. These key features are summarized schematically in Fig. 10.

Although enzymatic structure probing demonstrates the existence of the upper stem both in vitro and in vivo, only a single set of mutations in this stem (pUpper2 constructs [Fig. 7B]) supports a functional role for base pairing in the upper stem in encapsidation, and this role was relatively modest; disrupting base pairing reduced but did not eliminate packaged RNA. Mutations in the lower region of the upper stem (pUpper1 constructs [Fig. 7A]) suggest that primary se-

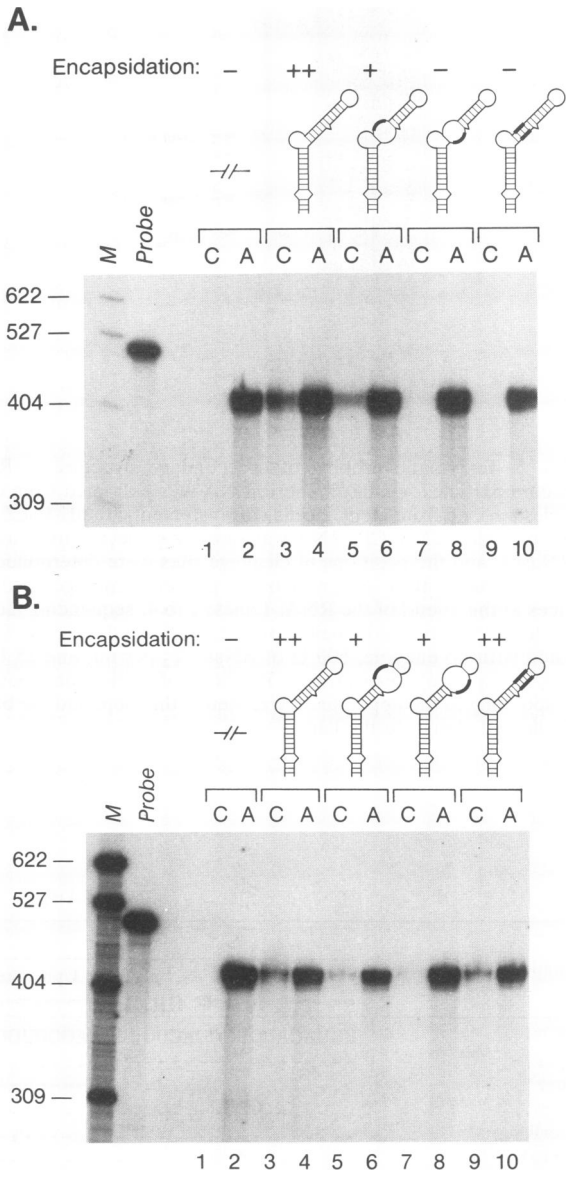

FIG. 7. Encapsidation assay of RNAs bearing base changes in the upper stem. Methods are identical to those described in the legend to Fig. 2B. (A) Lanes: M, size standards; Probe, undigested probe; 1, 3, 5, 7, and 9, core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper1L, pUpper1R, and pUpper1L/R, respectively; 2, 4, 6, 8, and 10, poly(A)$^+$ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper1L, pUpper1R, and pUpper1L/R, respectively. (B) Lanes: M, size standards; Probe, undigested probe; 1, 3, 5, 7, and 9: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper2L, pUpper2R, and pUpper2L/R, respectively; 2, 4, 6, 8, and 10: poly(A)$^+$ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper2L, pUpper2R, and pUpper2L/R, respectively. The encapsidation efficiency of constructs compared with pE-LacZ (wild type) is indicated as follows: ++, 50 to 100% of wild type; +, 10 to 50% of wild type; +/−, 1 to 10% of wild type; −, no detectable encapsidated RNA.

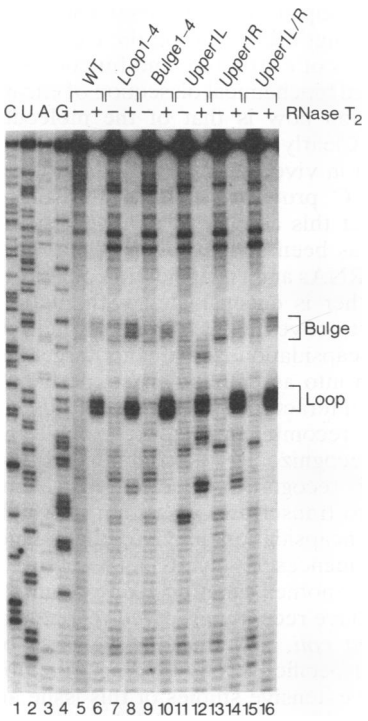

FIG. 8. Enzymatic structure probing of mutated RNAs generated in vitro. RNAs containing the indicated stem-loop mutations were synthesized in vitro, renatured, and cleaved with RNase T_2, specific for single-stranded regions. The positions of cleavage sites were determined by primer extension from an end-labeled primer homologous to sequences at the 3′ end of the RNA. Lanes: 1 to 4, sequencing ladder generated from the end-labeled primer; 6, 8, 10, 12, 14, and 16, RNase T_2 (60 U/ml) digestion of wild-type (WT), Loop1–4, Bulge1–4, Upper1L, Upper1R, and Upper1L/R RNAs, respectively; 5, 7, 9, 11, 13, and 15, no-enzyme digestion control of wild-type, Loop1–4, Bulge1–4, Upper1L, Upper1R, and Upper1L/R RNAs, respectively. Positions of nucleotide cleavage corresponding to the loop and bulge are indicated alongside the lanes.

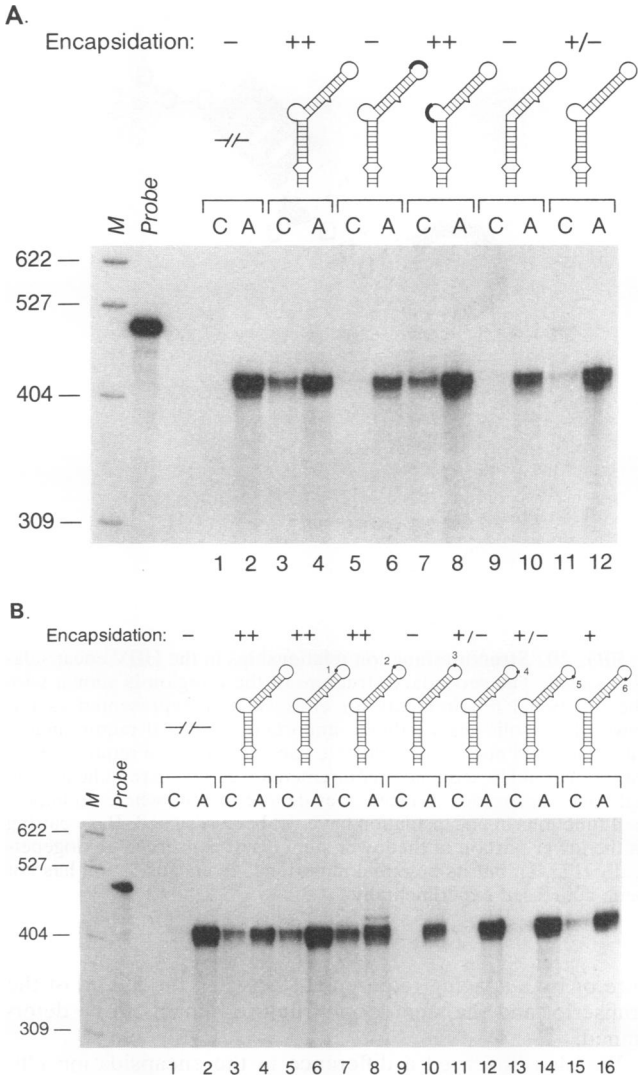

FIG. 9. Encapsidation assay of RNAs with nucleotide changes in the loop and bulge. Methods are identical to those described in the legend to Fig. 2B. (A) Lanes: M, size standards; Probe, undigested probe; 1, 3, 5, 7, 9, and 11, core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1–4, pBulge1–4, pΔbulge, and pΔU, respectively; 2, 4, 6, 8, 10, and 12: poly(A)$^+$ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1–4, pBulge1–4, pΔbulge, and pΔU, respectively. (B) Lanes: M, size standards; Probe, undigested probe; 1, 3, 5, 7, 9, 11, 13, and 15, core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1, pLoop2, pLoop3, pLoop4, pLoop5, and pLoop6, respectively; 2, 4, 6, 8, 10, 12, 14, and 16: poly(A)$^+$ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1, pLoop2, pLoop3, pLoop4, pLoop5, and pLoop6, respectively. The encapsidation efficiency of constructs compared with pE-LacZ (wild type) is indicated as follows: ++, 50 to 100% of wild type; +, 10 to 50% of wild type; +/–, 1 to 10% of wild type; –, no detectable encapsidated RNA.

quence and not base pairing is essential there. Therefore, although base pairing in some regions of the upper stem appears to increase packaging efficiency, perhaps by presenting essential bases in the loop in a favorable orientation, disruptions in base pairing in the upper stem appear better tolerated overall than do those in the lower stem. This is entirely consistent with the phylogenetic analysis, in which apparent disruptions of base pairing in the upper stem seem to be tolerated in several infectious avian hepadnavirus isolates (Fig. 1).

Our results in mapping the 5′ and 3′ extents of the HBV ε region differ in two minor respects from those reported by other investigators (2, 16). First, those studies reported that 29 nt that lie immediately 3′ of the stem-loop structure are essential for encapsidation, whereas we found no such requirement. Because the foreign sequences which lie 3′ of HBV sequences in their chimeric RNAs differ from ours, it is possible that the differences in these flanking sequences are responsible for this discrepancy. Second, in those studies, sequences upstream of the stem-loop structure were not found to increase packaging efficiency, whereas we found that their inclusion increases encapsidation efficiency about fivefold. In this context we note that Chiang et al. (7) found that deletion of these upstream sequences from wild-type pgRNA also resulted in a clear reduction of packaging efficiency; these results agree well with our findings and support the inference that sequences 5′ to the stem-loop contribute to RNA packaging. Whether this effect is mediated by specific sequences upstream of the stem-loop struc-

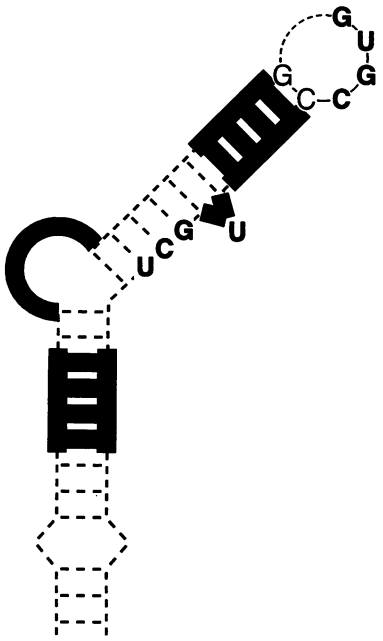

FIG. 10. Structure-function relationships in the HBV encapsidation signal. The secondary structure of the ε region is shown with the functional requirements for encapsidation represented as follows: thick solid lines indicate important structural requirements, and boldfaced nucleotides indicate specific sequence requirements. Additional likely sequence requirements (see text) are indicated by lightface nucleotides. Dashed lines denote regions where no important functions in encapsidation have yet been assigned. Base pairing in the lower portion of the lower stem is not conserved phylogenetically (Fig. 1), but its potential importance in encapsidation has not been addressed experimentally.

ture or by a spacing requirement between the 5′ end of the transcript and the stem-loop structure remains to be determined.

We also observed a difference in the encapsidation efficiency of HBV pgRNA from that reported by others (2), who found that HBV encapsidates only ca. 10% of its pgRNA. This efficiency is considerably lower than that previously observed (14) for the avian hepadnavirus DHBV (ca. 50%). However, in our hands HBV and DHBV encapsidated their pgRNAs with comparable efficiency (ca. 30 to 50%). We are uncertain of the reason for these differing estimates of HBV packaging efficiency; we note that we used a different HBV subtype from that used in the earlier experiments.

Because ε resides in the terminally redundant portion of pgRNA and because of the coterminal nature of all viral RNA transcripts, ε sequences are found at the 3′ ends of pgRNA and all subgenomic RNAs. However, at the 3′ position these sequences do not function to direct encapsidation (15). One possible explanation for this positional discrimination is that sequences upstream of the 3′ copy of ε interfere with the formation of the stem-loop structure (perhaps by favoring the formation of an alternate structure); these inhibitory sequences are not present upstream of the 5′ copy of ε. To test this idea, we generated an in vitro transcript that mimics the major subgenomic mRNA, containing a single ε region 3′ to 1,850 nt of HBV sequence. Enzymatic probing of this RNA shows that the characteristic ε stem-loop structure forms stably in such transcripts (unpublished data). However, we cannot exclude the possibility that this stem-loop structure forms in shorter RNAs derived from the full-length RNA during the cleavage reaction.

These studies of ε structure and function set the stage for a more detailed biochemical dissection of encapsidation. The central question now is that of the molecular basis of ε recognition. Clearly, C and P proteins are required for encapsidation in vivo, as judged by mutational analyses (2, 14, 21, 23). C protein is known to have RNA-binding properties, but this activity is relatively sequence nonspecific (4). It has been shown recently that encapsidation of ε-containing RNAs and encapsidation of P protein are tightly coupled; neither is encapsidated without the other (3). A simple and attractive model is that P protein directly recognizes the encapsidation signal and targets the RNA for encapsidation into assembling cores, perhaps through noncovalent C-P interactions. In support of this, a recent report suggests that recombinant P protein made in bacteria can specifically recognize RNAs containing ε sequences (17). However, this recognition operated on ε elements at the 3′ end of in vitro transcripts, a position in which they do not function in encapsidation in vivo; the bound RNAs also contained sequences involved in the initiation of reverse transcription, another function likely to involve P protein binding. We have recently prepared recombinant P proteins in *Escherichia coli*, and thus far we have observed only sequence-nonspecific RNA binding (unpublished data); clearly, more extensive studies of this issue are required.

Of course, this simple model is only one of several possibilities. The genetic data is equally compatible with a complex of C and P proteins being involved in the direct binding to ε or with even more complex models. For example, P protein might function to modify C protein to convert it to a sequence-specific RNA recognition protein. Distinguishing among these possibilities will require development of biochemical assays for ε recognition in vitro, and attempts to accomplish this are under way.

ACKNOWLEDGMENTS

We thank Dan Loeb and Sophie Roy for many helpful discussions throughout the course of this work.

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

REFERENCES

1. **Abrahams, J. P., M. van den Berg, E. van Batenburg, and C. Pleij.** 1990. Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. Nucleic Acids Res. **18:**3035–3044.
2. **Bartenschlager, R., M. Junker-Niepmann, and H. Schaller.** 1990. The P gene product of the hepatitis B virus is required as a structural component for genomic RNA encapsidation. J. Virol. **64:**5324–5332.
3. **Bartenschlager, R., and H. Schaller.** 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. EMBO J. **11:**3413–3420.
4. **Birnbaum, F., and M. Nassal.** 1990. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. J. Virol. **64:**3319–3330.
5. **Black, D. L., and A. L. Pinto.** 1989. U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. Mol. Cell. Biol. **9:**3350–3359.
6. **Bruss, V., and D. Ganem.** 1991. The role of envelope proteins in hepatitis B virus assembly. Proc. Natl. Acad. Sci. USA **88:** 1059–1063.
7. **Chiang, P.-W., K.-S. Jeng, C.-P. Hu, and C. Chang.** 1992. Characterization of a *cis* element required for packaging and replication of the human hepatitis B virus. Virology **186:**701–711.

8. **Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J.-P. Ebel, and B. Ehresmann.** 1987. Probing the structure of RNAs in solution. Nucleic Acids Res. **15:**9109–9128.
9. **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.
10. **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.
11. **Ganem, D., and H. E. Varmus.** 1987. The molecular biology of the hepatitis B viruses. Annu. Rev. Biochem. **56:**651–693.
12. **Geballe, A.-P., R. R. Spaete, and E. S. Mocarski.** 1986. A *cis*-acting element within the 5′ leader of a cytomegalovirus β transcript determines kinetic class. Cell **46:**865–872.
13. **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.
14. **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.
15. **Hirsch, R., D. D. Loeb, J. R. Pollack, and D. Ganem.** 1991. *cis*-acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. J. Virol. **65:**3309–3316.
16. **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.
17. **Kochel, H. G., M. Kann, and R. Thomssen.** 1991. Identification of a binding site in the hepatitis B virus RNA pregenome for the viral pol gene product. Virology **182:**94–101.
18. **Kunkel, T. A., J. D. Roberts, and R. A. Zabour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154:**376–382.
19. **Lavine, J., R. Hirsch, and D. Ganem.** 1989. A system for studying the selective encapsidation of hepadnaviral RNA. J. Virol. **63:**4257–4263.
20. **Martinez, H. M.** 1988. An RNA secondary structure workbench. Nucleic Acids Res. **16:**1789–1798.
21. **Nassal, M.** 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. J. Virol. **66:**4107–4116.
22. **Nassal, M., M. Junker-Niepmann, and H. Schaller.** 1990. Translational inactivation of RNA function: discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. Cell **63:**1357–1363.
23. **Roychoudhury, S., A. F. Faruqui, and C. Shih.** 1991. Pregenomic RNA encapsidation analysis of eleven missense and nonsense polymerase mutants of human hepatitis B virus. J. Virol. **65:**3617–3624.
24. **Skinner, M. A., V. R. Racaniello, G. Dunn, J. Cooper, P. D. Minor, and J. W. Almond.** 1989. New model for the secondary structure of the 5′ non-coding RNA of poliovirus is supported by biochemical and genetic data that also show that RNA secondary structure is important in neurovirulence. J. Mol. Biol. **207:**379–392.
25. **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.
26. **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.
27. **Valenzuela, P., M. Quiroga, J. Zaldivar, P. Gray, and W. J. Rutter.** 1980. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. ICN-UCLA Symp. Mol. Cell. Biol. **18:**57–70.
28. **Will, H., W. Reiser, T. Weimer, E. Pfaff, M. Buscher, R. Sprengel, R. Cattaneo, and H. Schaller.** 1987. Replication strategy of human hepatitis B virus. J. Virol. **61:**904–911.