Nucleic Acids Research, 1995, Vol. 23, No. 19 3909–3915

Distinct requirements for primary sequence in the 5′- and 3′-part of a bulge in the hepatitis B virus RNA encapsidation signal revealed by a combined *in vivo* selection/*in vitro* amplification system

Andrea Rieger and Michael Nassal*

Zentrum für Molekulare Biologie, Universität Heidelberg, Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany

Received June 26, 1995; Revised and Accepted August 23, 1995

ABSTRACT

Hepatitis B virus (HBV) is a small DNA virus that replicates by reverse transcription of a terminally redundant RNA, the pregenome. Specific packaging of this transcript into viral capsids is mediated by interaction of the reverse transcriptase, P protein, with the 5′-proximal encapsidation signal ε. ε-function is correlated with the formation of a hairpin structure containing a bulge and a loop, each consisting of 6 nt. To analyse the importance of primary sequence in these regions, we have combined selection of encapsidation competent individuals from pools of randomized ε-sequences in transfected cells with *in vitro* amplification, thus bypassing the current experimental limitations of the HBV system. While no alterations of the authentic loop sequence were detectable, many different sequences were tolerated in the 3′-part of the bulge. However, at the two 5′-proximal bulge positions the wt sequence was strongly selected for, indicating that for RNA packaging close contacts between protein and the 5′- but not the 3′-part of the bulge are important. Such a bipartite organisation provides a structural basis for the recently demonstrated special role of the 3′-part of the bulge as template for the first nucleotides of (–)-strand DNA in HBV reverse transcription.

INTRODUCTION

Hepatitis B virus (HBV), the causative agent of B-type hepatitis in man, is the type member of the hepadnaviridae. These small enveloped DNA viruses, characterized by narrow host range and liver tropism, replicate through reverse transcription of an RNA intermediate inside the viral capsid (for reviews see refs 1,2). From their tiny genome of ~3 kb several terminally redundant genomic transcripts (Fig. 1) as well as three sets of subgenomic RNAs are produced which, owing to the presence of a single poly-adenylation signal, share a common 3′-end. All transcripts serve as mRNAs; one of the genomic RNAs, however, the pregenome, has unique characteristics: it encodes, bicistronically, the core protein forming the capsid shell and the reverse transcriptase (P-protein) and it is specifically packaged into capsids and there is the substrate for reverse transcription.

Pregenome packaging and initiation of capsid assembly are mediated by interaction of P protein with a 5′-proximal region on the RNA (3), the encapsidation signal ε (see Fig. 1A). In the presence of core and P protein, foreign RNAs (e.g. *lacZ*) carrying the ε-segment at their 5′-end can be specifically packaged into HBV capsids (Fig. 1B). This and the mutual dependence of RNA and P protein packaging (4,5) indicated that the interaction between P protein and ε is the decisive event in the encapsidation process. With the advent of a heterologous expression system for the P protein of a related avian virus (6), duck hepatitis B virus (DHBV), a direct RNA/protein association could recently be demonstrated (7). For HBV, unfortunately, no system for the production of isolated active P protein is as yet available.

We and others have instead used DNA transfection to study the effects of mutations in HBV ε on encapsidation (8–10) and shown, by correlating these data with secondary structure analysis of *in vitro* transcribed ε-RNAs, that activity depends on the formation of a bipartite secondary structure with a 6 nt bulge, 6 nt loop and a single unpaired U-residue (Fig. 1C). Deletions of these single-stranded regions abolished, or at least dramatically reduced encapsidation, suggesting that the overall structure of ε is important for P protein binding. Possible contributions of primary sequence, by contrast, are difficult to infer from the limited number of substitution variants analysed. For instance, some bulge mutants with up to 4 nt exchanges compared to the wt-bulge sequence were packaged, others were not (8). Hence, deriving a more detailed picture of the P protein/ε interaction would require analysis of many more mutants, an enormous effort in view of the large number of possible sequence combinations.

Hence, selection for P protein binding of individuals from pools of randomized ε-sequences would be a most appropriate approach, as demonstrated by the successful application of this technique to various interactions of nucleic acids with other molecules, both *in vitro* (11–13) and, more recently, *in vivo* (14,15). However, not only does the lack of a source for HBV P protein prevent employment of *in vitro* procedures but there are also no HBV-infectable cell lines that would allow for selection and multiplication of viable virus. DNA transfection into human hepatoma cell lines, on the other hand, provides access to at least small quantities of HBV capsids containing specific nucleic acid. We therefore established a combined system in which transfection is used to select, *in vivo*, encapsidation competent ε-variants from pools of partially ran-

* To whom correspondence should be addressed

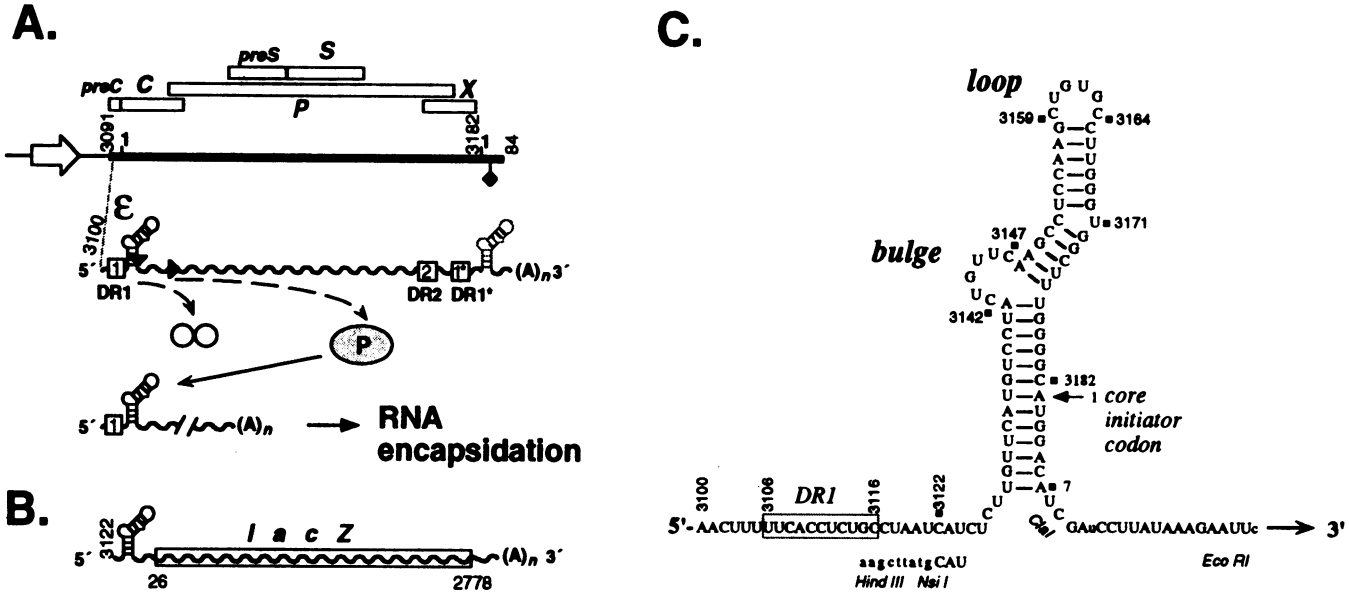

Figure 1. (**A**) Genome structure of HBV and ε/P protein mediated RNA pregenome encapsidation. The solid black line is a linear representation of the circular HBV genome (3182 bp). Open bars on the top show the four major ORFs. Numbers are nt positions according to the system of Pasek *et al.* (17). The diamond symbolizes the single poly-A signal. Transcription of the terminally redundant pregenome (wavy line) can also be driven by foreign promoters (arrowhead). Stem–loops symbolize the two copies of the ε-sequence; direct repeats are shown as boxes marked DR. Initiation sites for core and P protein translation are indicated by triangles. Interaction of P with 5′-ε initiates pregenome encapsidation and capsid assembly. (**B**) Principal structure of ε-*lacZ* reporter RNA. Fusion of an HBV fragment starting at position 3122 and encompassing the stem–loop structure mediates encapsidation of *lacZ* RNA. (**C**) Sequence and secondary structure at the 5′-end of the pregenome. Authentic RNA starts at position 3100 and encodes a *Sty*I site (CCUUGG; position 3164); the DNA ε-cassettes used here contain 5′-terminal *Hin*dIII/*Nsi*I sites (lower case sequence), and sites for *Cla*I (position 7) and *Eco*RI (position 22). CMV promoter driven transcription in the ε-*lacZ* constructs starts at position 3122 (center of *Nsi*I site).

domized ε-sequences; sufficient amounts for sequence analysis are obtained in a separate *in vitro* step in which the encapsidated RNAs are amplified by RT-PCR (Fig. 2).

This scheme was applied to pools of ε-*lacZ* RNAs with randomized loop or bulge sequences. While no alternatives to the wt-loop sequence were found, numerous bulge variants were packaged and hence able to interact with P protein. Sequence analysis revealed that the importance of primary sequence within the ε-bulge was very unevenly distributed: the wt sequence was strongly preferred at the two most 5′-proximal bulge positions while almost all nt were tolerated at the following positions. While surprising at first glance, this structural arrangement is in remarkable agreement with recent functional data showing that the nt whose nature is not important for packaging serve instead as template for a short DNA primer used to initiate first strand DNA synthesis in HBV (16).

MATERIALS AND METHODS

Chemicals and enzymes

Oligonucleotides were synthesized by standard phosphoramidite chemistry on an ABI 381B synthesizer. For partially randomized oligonucleotides, a freshly prepared equimolar mixture of amidites was used at the corresponding positions to ensure an approximately equal incorporation of the desired nt. Enzymes for standard molecular cloning experiments were obtained from New England Biolabs (Bad Schwalbach, Germany) or Boehringer (Mannheim, Germany). For RT–PCR, Superscript reverse transcriptase was purchased from Gibco-BRL (Eggenstein, Germany) and Pfu polymerase from Stratagene (Heidelberg, Germany). All enzymes were used as recommended by the manufacturer.

Plasmid constructs

Plasmid libraries encoding ε-*lacZ* fusion RNAs with randomized bulge or loop sequences were constructed using the previously described plasmid pCHG-3122 (8). In brief, the construct contains the CMV-IE promoter, followed by HBV sequence from position 3122 to 36 [numbering system of Pasek *et al.* (17)], that is the complete ε-stem–loop structure (position 3129–3127), flanked by restriction sites, the *E.coli lacZ* gene (positions 26–2778) and a poly-adenylation signal from SV40. To facilitate directional cloning, a 2.9 kb stuffer fragment from pBR322 was inserted between the *Nsi*I (overlapping with HBV position 3122) and *Sty*I (HBV position 3164) restriction sites; this allowed efficient separation of the desired double-digested vector fragment from linearized vector. For the intron-containing pCHG-3122Ad constructs, the adenovirus major late intron sequence (233 nt) present on plasmid pBS-Ad1 (ref. 18; kindly provided by Dr I. Mattaj, EMBL, Heidelberg) was amplified by PCR with primers creating flanking *Mst*II restriction sites and inserted into the *Mst*II site in the *lacZ* part (position 238) of pCHG-3122. The resulting sequence is CCTAGGG/gt-intron-ag/CTCGCGCCTAGGG (*Mst*II sites underlined, intron nt in lower case). For ε-bulge variants, the *Nsi*I–*Sty*I, for loop variants the *Hin*dIII (directly preceding *Nsi*I)–*Cla*I (HBV position 7) fragments were replaced by synthetic oligonucleotide duplexes with degenerate sequences at the corresponding positions (see Fig. 2). Aliquots of the ligation reactions were transformed into *E.coli* DH5α by electroporation. Plasmid DNA was isolated using the standard CsCl gradient procedure after growing ~15 000 colonies in liquid medium for 3–4 h. The presence of random sequence at the

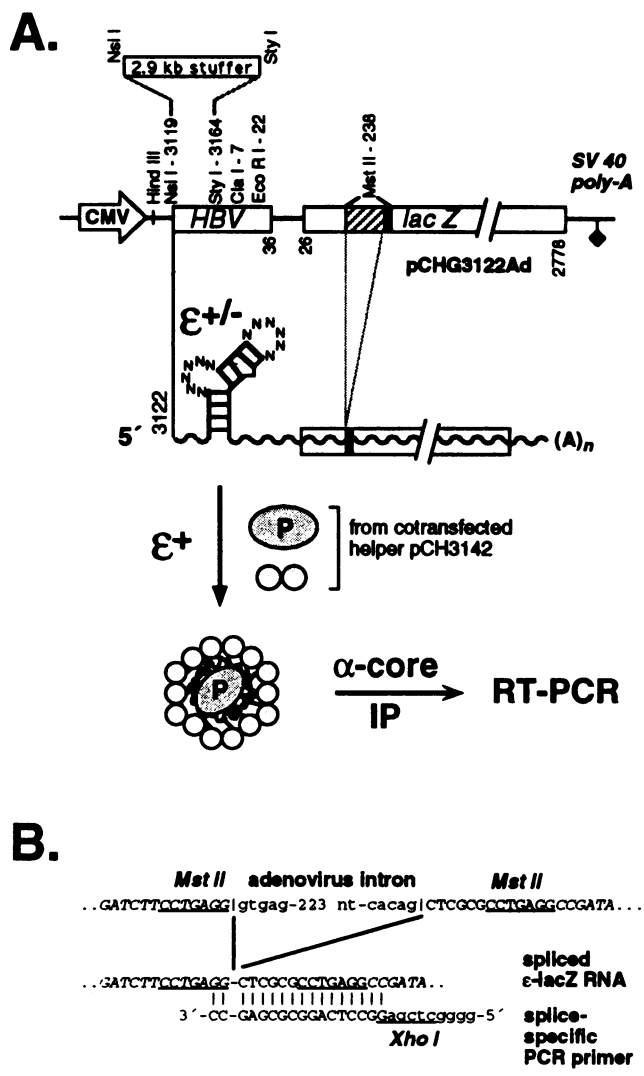

Figure 2. (**A**) Principle of *in vivo* selection/*in vitro* amplification system for encapsidation competent ε-variants. The relevant elements of the ε-*lacZ* expression plasmid pCHG3122Ad are: CMV-IE promoter (arrowhead), ε (box marked HBV) and *lacZ*, poly-A signal from SV40 and an adenovirus intron inserted into the *Mst*II site of *lacZ* (hatched box); the black box symbolizes a stretch of 13 nt remaining after intron splicing (see B for detail). For efficient library constructions, a 2.9 kb stuffer fragment was inserted between the *Nsi*I and *Sty*I sites, then replaced by synthetic, degenerate DNA duplexes. Transcription after transfection yields pools of ε-*lacZ* RNAs with randomized bulge or loop. Functional individuals (ε^+) are packaged by interaction with P and core protein provided by the helper construct pCH3142. Core particles are isolated by immunoprecipitiation. ε^+-RNAs are extracted and amplified *in vitro*. (**B**) Sequence at the exon–intron junctions in the ε-*lacZ* reporter constructs. Authentic *lacZ* sequence is shown in italics, vertical bars indicate the intron borders. Splicing removes the intron but leaves a stretch of additional 13 nt (black box in A) compared to *lacZ*. The downstream primer for RT–PCR amplification matches perfectly to the junction sequence and provides an *Xho*I site for cloning.

expected positions was confirmed by sequencing aliquots from the plasmid pools and DNA from several individual colonies.

Detection of intron-splicing by RNAse protection assay

The conditions for RNase protection assays were essentially as previously described (19). To detect splicing of the adenovirus intron from the ε-*lacZ* RNA, a 651 nt antisense RNA complementary over 580 nt the intron-containing *lacZ* part was prepared by *in vitro* transcription in the presence of [α-^{32}P]CTP (3000 Ci/mM). About 50 000 c.p.m. of gel-purified probe RNA were hybridized with RNA from transfected cells, either from cytoplasmic lysates, from immunoprecipitated core particles (8) or from the nuclei separated from the detergent lysate by low speed centrifugation. Protected fragments were separated on denaturing gels containing 6% polyacrylamide and visualized either by autoradiography on Kodak X-OMAT X-ray film or using a Phosphor-Imager system (Molecular Dynamics, Krefeld, Germany).

In vivo selection of encapsidation competent ε-variants

Ten micrograms of the pCHG3122Ad construct pools with randomized bulge or loop sequence and 5 µg of the encapsidation deficient helper construct pCH-3142 (3) per 10 cm plate were cotransfected into Huh7 cells using the standard calcium phosphate coprecipitation procedure. Three days post transfection, total RNA and encapsidated RNA from immunoprecipitated core particles was isolated by proteinase K treatment and phenol extraction as previously described (8,19).

In vitro amplification and characterization of encapsidation competent ε-variants

About one-third of the core RNA obtained from one 10 cm dish was reverse transcribed into cDNA using hexamer random primers and MuLV superscript reverse transcriptase. cDNA-amplification was performed using Pfu polymerase and one primer corresponding to HBV positions 3122–3139 with 5′-flanking *Hin*dIII and *Nsi*I sites, and a second primer complementary to the sequence GG/CTCGCGCCTGAGGCC created by splicing of the adenovirus intron (location of the former intron indicated by the slash) from the ε-*lacZ* constructs; the primer provided an additional 5′-proximal *Xho*I site for later cloning. For decontamination, the reactions were incubated with *Alu*I for 30 min, then UV-irradiated (Stratalinker 1800, Stratagene, Heidelberg, Germany) prior to addition of the cDNA template. Polymerase chain reaction products of the expected size were isolated from low melting agarose gels (FMC, Rockland, Maine) and cloned into pBSIISK(–) (Stratagene, Heidelberg, Germany) cut with *Hin*dIII and *Xho*I. Sequence analyses were performed using Sequenase (USB/Amersham, Braunschweig, Germany).

RESULTS

Principal design of the combined *in vivo*/*in vitro* system for selection of encapsidation competent HBV ε variants

The principle of the selection procedure is based on the transfection scheme we have previously used to study encapsidation of individual ε-variants (8), as outlined in Figure 2A. Pools of expression constructs for ε-*lacZ* fusion RNAs (here with randomized bulge or loop sequences) and an encapsidation-deficient HBV helper construct (pCH3142; ref. 3) providing core and P protein are cotransfected into Huh7 cells. Encapsidation competent ε-sequences are packaged into core particles which are

purified from the cell lysates by immunoprecipitation with anti-core antibodies. Nucleic acid isolated from the capsids is subjected to RT–PCR amplification and the products are cloned and sequenced.

Preparation of pools of partially randomized ε-*lacZ* expression constructs

As in transfection the amount of input DNA is limited, we restricted ourselves to analysing random pools of limited size. Both the loop and the bulge of ε consist of 6 nt; hence complete randomization of either subelement would nominally lead to 4^6, that is 4096, individuals in the pool. To ensure efficient cloning of synthetic oligonucleotide duplexes with random sequence at the desired positions, we first modified the previously described ε-*lacZ* construct pCHG3122 by insertion of a 2.9 kb stuffer fragment between the *NsiI* and *StyI* restriction sites bordering and inside of respectively, the ε-sequence (cf. Fig. 2A, top). In addition to inactivating the ε-signal in the parental construct, this allowed for efficient separation of doubly digested vector fragment from linearized vector. We routinely obtained, by electro-transformation of ~1/20 of the ligation reactions, 15 000–20 000 colonies. Preparative amounts of plasmid DNA were obtained by CsCl density centrifugation after the scraped-off colonies had been grown for limited time in liquid medium. The presence of random sequence at the expected positions was confirmed by sequencing aliquots from the plasmid preparations.

Precautions to prevent false-positives from amplification of input plasmid

Transfection by calcium phosphate coprecipitation under our standard conditions involves the use of microgram amounts (i.e. up to several pMol) of plasmid DNA per 10 cm cell culture dish. The yield of core particles from these cells, as measured by ELISA (20), is usually between 10 and 50 ng. Taking into account that core particles consist of probably 240 subunits of the 20 kDa core protein (21,22), the maximal amount of packaged RNA cannot exceed some 2–10 fMol. Hence even a small fraction of input plasmid contaminating the core particle preparation could seriously flaw the *in vitro* amplification step. Although our protocol involves treatment of the immunoprecipitated core particles with *S.aureus* nuclease, we sought to ensure the RNA rather than DNA origin of any RT–PCR product by an additional biological control. One such possibility is to take advantage of splicing as an RNA-specific process. We hence inserted the adenovirus intron sequence present on plasmid pBS-Ad1 (18) into the *lacZ* part of the ε-*lacZ* reporter constructs (Fig. 2A). By using appropriately located PCR primers, spliced RNA-derived products can be easily distinguished, by size, from nonspliced plasmid DNA derived products. As splicing leaves a unique region of 13 nt which is absent from the parental plasmid pCHG3122, this allows also for a discrimination against contaminations with intronless plasmids in use in the laboratory. Use of a PCR primer which matches perfectly only the newly created junction sequence further enhances specificity (Fig. 2B).

To monitor whether the adenoviral intron was properly removed from the ε-*lacZ* reporter RNAs, we transfected the wt construct pCHG3122Ad into Huh 7 cells and analysed by a splice-specific RNAse protection assay (Fig. 3C) the corresponding transcripts in equal aliquots of the cytoplasmic and the core particle fraction, and also in the nuclei from the detergent-lysed cells (Fig. 3A). The nuclear fraction (lane N) gave a relatively

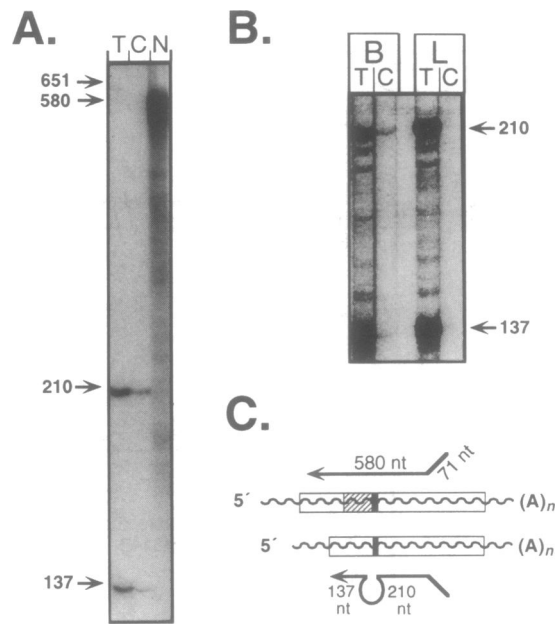

Figure 3. Splicing and encapsidation of ε-*lacZ* reporter RNAs. **(A)** Efficient splicing of an adenovirus intron from ε-*lacZ* RNA detected by RNase protection assay. Total RNA was isolated from cytoplasmic lysate (lane T), from immunoprecipitated cytoplasmic capsids (lane C) or from the pelleted nuclei (lane N) of cells transfected with pCHG-3122Ad. The two fragments of 137 and 210 nt in lanes C and T are characteristic for correctly spliced transcripts, the dominant 580 nt fragment in the nuclear fraction for unspliced RNA (see panel C). **(B)** Encapsidation of ε-*lacZ* RNAs with randomized bulge and loop. In this assay, cells were transfected with plasmid pools carrying random sequence in the ε-bulge (sample B) or loop (sample L). Note that core RNA from sample B but not L produced distinct signals of the expected size. **(C)** Schematic representation of RNAse protection assay. The *in vitro* transcribed antisense probe RNA of 651 nt is complementary over 580 nt to the unspliced transcript. Intron removal reduces complementarity to two fragments of 210 and 137 nt.

strong signal compatible with its origin from unspliced precursor RNA; by contrast, both the cytoplasmic total (lane T) and the core particle RNA (lane C) yielded the two signals expected for spliced RNA. We therefore used these intron containing plasmids to establish pools of partially randomized ε-*lacZ* sequences as described above.

Substantial primary sequence variability in the bulge but not the loop of encapsidation-competent ε-signals

For an approximate comparison of the fractions of encapsidation competent individuals in the ε-*lacZ* pools with randomized bulge and loop, we used the same RNAse protection assay as described above. As shown in Figure 3B (lanes T), both construct pools yielded comparable amounts of transcripts in the cytoplasmic total RNA fraction (which contains unpackaged RNA plus encapsidated transcripts present in core particles contained in this fraction). However, in the core fractions, weak but specific signals for encapsidated RNA could only be detected in the sample from the pool of bulge variants (compare lanes C of samples B and L in Fig. 3B). These observations were confirmed in independent experiments and they suggest that relatively many nt substitutions in the ε-bulge but only few in the loop are compatible with productive P protein binding. This conclusion is further supported by the experiments described below.

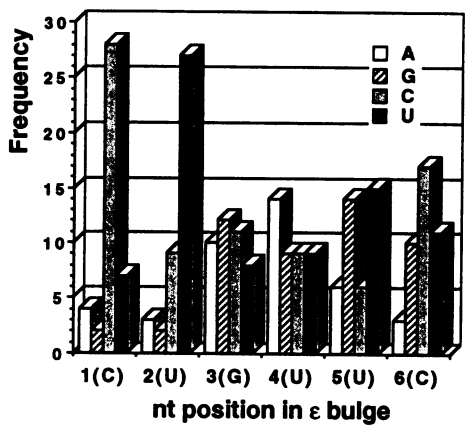

Figure 4. Importance of primary sequence for encapsidation at individual ε-bulge positions. Forty-one individual clones obtained by selection from the bulge pool were sequenced. On the x-axis, the six bulge positions are indicated in 5′- to 3′-direction; letters in parentheses show the wt-sequence. Each bar represents one of the four nt; the height of a bar corresponds to the sum of occurrences of a specific nt at a given position.

Primary sequence in the 5′- but not the 3′-proximal part of the ε-bulge is important for RNA encapsidation

For further characterization, core RNA from the pool of ε-bulge variants was subjected to RT–PCR amplification which yielded a product of the size expected for correctly spliced RNA. The *Hin*dIII and *Xho*I sites provided by the PCR primers were used for directional cloning into plasmid pBSII SK(–) and the relevant region of plasmid DNAs from 41 independent colonies was sequenced. The results are graphically summarized in Figure 4 which shows the frequency with which a specific nt was found at a specific bulge position.

The graph reveals a markedly uneven distribution: the majority of the isolates contained a C residue at position 1 (28/41) and U at position 2 (27/41), whereas no clear-cut preference for a specific nt was seen at the downstream bulge positions, perhaps with the exception of an apparent selection for C and against A at position 6. Strikingly, the sequence CU at positions 1 and 2 of the bulge corresponds to that of the wt and the second most frequent nt at these positions was the other pyrimidine. We conclude that for encapsidation and hence P protein binding, the specific nt at bulge positions 1 and 2 are particularly important, whereas those further downstream are not.

The ε-loop is probably a sequence-specific determinant for P protein binding

When the RT–PCR procedure was applied to the pool of loop variants, no specific products were obtained, in accord with the absence of a signal for encapsidated loop variants in the RNAse protection assay described above. Also, the positive result with the pool of bulge variants suggested that this negative outcome was not due to a principal technical problem. However, to further confirm this interpretation, we repeated the selection experiment with a small pool of only 4^2 variants in which loop positions 4 and 5 (originally UG) were randomized (data not shown). From these experiments only the wt sequence was isolated. Some variants were individually isolated and transferred into the context of a complete HBV genome to test for formation of replication competent capsids by the endogenous polymerase assay. In this reaction, encapsidated ('endogenous') P protein, if provided with dNTPs, reverse transcribes the RNA and further extends incomplete viral DNA to form relaxed circular and linear 3.2 kb DNA (e.g. ref. 8). Again no specific signals were obtained. Hence the most plausible explanation for the negative outcome of these experiments is that the majority of mutations within the ε-loop interfere with P protein binding, in other words that the ε-loop is to a large part a sequence-specific recognition element.

DISCUSSION

With an estimated 300 million chronic carriers, HBV is one of the most important viral pathogens, especially in view of the highly increased risk for irreversible liver damage that is associated with the carrier state and the lack of a generally applicable therapy (23). Despite substantial advances in understanding the molecular biology of the virus (1,24), the lack of appropriate *in vitro* and *in vivo* systems has so far impeded the application of selection-based analyses to HBV. Such procedures, initially mostly performed *in vitro* (11–13,25), have more recently been extended to *in vivo* studies, taking advantage of the self-amplification capability of viral systems, either using defined sequence alterations and monitoring virus evolution based on the inherently high error rate of RNA-dependent replicases (26) or by constructing virus libraries with limited regions of random sequence (14,15). In general, they confirmed the notion that sequence-specific contributions to RNA/protein interactions often reside in, or immediately adjacent to, single-stranded regions (for review see ref. 27). In this report we describe a combined *in vivo* selection/*in vitro* amplification procedure that exploits the only well established experimental system for HBV, that is DNA transfection, to further analyse the interaction between P protein and the RNA encapsidation signal ε. This interaction is central to pregenome packaging and capsid assembly, but, moreover, is essential for initiation of reverse transcription (16). The striking asymmetry in the importance of primary sequence in the 5′- and the 3′-part of the ε-bulge revealed by this technique is in full accord with these two different functions.

While theoretically straight-forward, a potentially serious problem for the *in vitro* amplification step of our approach is that the amount of input plasmid DNA exceeds by up to 1000-fold the maximum expectable amount of encapsidated RNA. This was accounted for by introducing an intron sequence into the ε-*lacZ* reporter constructs which was indeed efficiently spliced from its new context. Hence the products from *in vitro* amplification are certainly derived from RNA. A contamination with RNAs that are not properly encapsidated but nonspecifically associated with the core particle preparation can also not be rigorously excluded. However, several of the selected bulge variants were subjected to independent encapsidation tests, for instance, the endogenous polymerase assay (e.g. ref. 8), or formation of (–)-DNA (16). There is strong evidence that these processes occur only in the context of properly assembled core particles (20,28). Hence we are confident that our protocol faithfully selects for encapsidation competence.

Sequence-specific contributions to encapsidation competence in the single stranded bulge and loop of HBV ε

The most striking result with the pool of bulge variants was the pronounced tendency for maintaining the motif CU at the

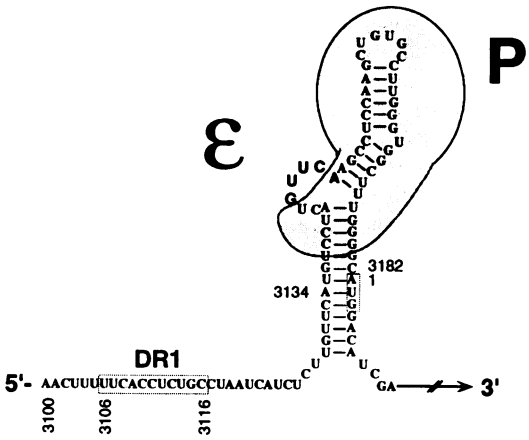

Figure 5. Model for ε/P protein interaction derived from the encapsidation phenotypes of variant ε-sequences. The 5′-terminal sequence of the pregenome is represented as in Figure 1C. The model assumes that nt within ε that can neither be deleted nor altered in sequence are in direct contact with P protein. The sequence between nt 3134 and 3182 is sufficient for encapsidation competence. The asymmetric shape of the protein symbolizes the unevenly distributed importance of primary sequence in the 5′- and 3′-part of the bulge. See the Discussion for details.

5′-proximal bulge positions which, remarkably, corresponds to the wt-sequence; the second most frequent nt at these positions was the other pyrimidine. The presence of virtually random sequence at the following positions provides an additional, internal control for the degeneracy of the original pool and hence the specificity of the selection procedure. Together, these data suggest that specific sequence matters predominantly at bulge positions 1 and 2. We propose that these nt are in close contact with P protein (see Fig. 5). Although no systematic study has been published that addresses the minimal bulge size, complete deletion of the bulge is incompatible with encapsidation (8,9). Hence it is probably the bulge structure as such that is required for productive interaction with P protein and the 4 nt whose specific nature does not matter for encapsidation have a more indirect structural role, for example by introducing a kink into the double-helical ε-structure (29). Our data also point to a certain selection for the wt C- and against an A-residue at position 6 of the bulge. We currently do not know the significance of this observation; however, an A-residue at position 6, according to computer prediction, would shift the bulge by 1 nt toward the 5′-end; this may be incompatible with P protein binding. The strong preference for the sequence CU, or at least two pyrimidines, at positions 1 and 2 of the bulge does not imply that all individuals with a different sequence are encapsidation incompetent; however, under competitive conditions (each transfected cell will have taken up many different DNA molecules, the transcripts of which compete for P protein) those conforming to this consensus obviously perform better.

While it is difficult to assess whether HBV sequences isolated from infected patients by PCR belong to replication competent or defective genomes, the high sequence conservation in natural isolates (30,31) of those bulge nt that tolerate an extreme variability in our assay is stunning. The reason for this apparent contradiction is discussed below.

The negative result with the ε-loop variants does not exclude that functional alternative loop sequences exist. Two of six previously described loop mutations, at position 1 (C>U) and 2 (U>A), were compatible with relatively efficient encapsidation (9). However, point mutations at the following positions (8,9,32), or moving the C-residue from position 4 to position 1 abolished or drastically reduced encapsidation. Probably, the fraction of encapsidation competent variants in our pool was too low (e.g. 16/4096 = 0.4% if all possible combinations were allowed at loop position 1 and 2) for the sensitivity of our current assay; by contrast, the corresponding fraction in the bulge pool is at least 10-fold higher (assuming that all nt are tolerated at bulge positions 3–6), and may approach 25% if the only constraint were to have pyrimidines at positions 1 and 2. Hence we are confident that very few sequence alterations are allowed in the ε-loop which therefore appears to be, to its larger part, a sequence specific determinant. Here, this view is strongly supported by the virtual absence, in natural isolates of HBV, of loop mutants (30,31).

A model for P protein/ε interaction: implications for HBV replication

Although direct binding of a hepadnaviral replication enzyme to RNA could only recently be demonstrated using *in vitro* translated DHBV P protein (7,33), genetic evidence strongly suggests that encapsidation competence is an appropriate measure of productive binding between HBV ε and P protein. Hence previous mutational data and those described above can be used to derive a tentative model for the interaction. The basic assumption for the model shown in Figure 5 is that those parts of ε that can neither be deleted nor grossly altered in sequence without loss of function are in close contact with P protein. While the lower half of the basal stem is dispensable (34), all of the apical structure is required for encapsidation and in this model is involved in protein binding. Likewise, the loop must be present, tolerates only very few changes in primary sequence and hence may be involved in direct contacts with P.

The nonsymmetrical shape of P in this model is intended to represent the uneven participation of the bulge nt in protein binding, reflecting the sequence-specific contribution of nt positions 1 and 2 but not the rest of the bulge region. In the absence of any high resolution structural data for both the RNA and P protein, such a model is certainly speculative. However, the general architecture implied by the model is at least conceivable: the total length of the double-stranded part of ε shown to support encapsidation encompasses 11 bp in the upper and 6–7 bp in the lower stem (34). If forming a continuous A-type helix, its length would be ~5 nm (assuming a helical pitch of ~3 nm/11 bp). The heterodimeric p66/p51 HIV-1 reverse transcriptase contains a nucleic acid binding cleft of ~10 nm length (35–37). Hence the entire upper part of ε may well be accomodated by the 90 kDa HBV P protein. Regarding the ε-bulge, our finding that almost any sequence is tolerated in its 3′-part argues against strong interactions of these residues with the protein. The two highly conserved 5′-bulge nt, however, may be involved in protein contacts. For instance, the recently determined high resolution structure of the MS2 operator/coat protein complex (38) confirmed direct contacts with the protein of most of the non-paired residues that were previously shown, by mutagenesis, to be essential for tight binding (39). However, alternative explanations can currently not be excluded. The invariably conserved A-7 residue in the MS2 operator serves a structural role without being directly bound to the protein (38). Also, the two conserved ε bulge nt may be part of a

higher order RNA structure; since deletions of the HBV specific 5′- and/or 3′-flanking sequences do not abolish ε function (34) we consider this less likely. We note, however, that the TAR stem–loop of HIV-1 provides binding sites not only for Tat protein but also cellular proteins (40,41). Our selection procedure would not discriminate between requirements within the ε-sequence for binding to P protein only or P protein plus additional cellular factors which might act as cofactors in encapsidation.

Regardless of these caveats, the model provides a structural rationale for the importance of the bulge in a second essential function of the ε/P protein interaction, that is in HBV replication; this also explains the seemingly contradictory results of selection in an infected patient and in our experimental system. Part of the sequence conservation in natural isolates is due to the coding function of ε for the precore protein that serves a precursor for the secretory core gene product HBeAg, although the antigen is not essential for the basic viral life-cycle (for review, see ref. 2). The decisive difference, however, is that the 3′-part of the ε-bulge serves as template for a short DNA primer used to initiate discontinuous (–)-DNA synthesis, as was recently suggested for duck hepatitis B virus (DHBV; refs 42,43). Using several bulge variants isolated by the selection procedure described above we could recently demonstrate that the first few nt of (–)-DNA are complementary to the 3′-half of the ε-bulge while the 5′-proximal nt are not copied (16). Primer transfer to a specific 3′-proximal site on the pregenome (DR1*; cf. Fig. 1) is in part dependent on sequence complementarity. Hence only viruses with matching bulge and DR1* sequences will be fully replication competent while our selection scheme asks only for encapsidation competence.

Though our system is relatively laborious compared to selection procedures performed entirely *in vitro* or *in vivo*, its intermediate level of complexity allows to study the sequence requirements for nucleic acid/protein interactions that go beyond simple binding but are less elaborate than those needed to establish a full infectious cycle. This will probably be advantageous also for other systems, but particularly if, like in HBV, many functions are overlappingly arranged on the same nt sequence. The ability to be reverse transcribed can also easily be built-in to the selection scheme by subjecting only capsid-borne DNA, after removal of RNA, to *in vitro* amplification. It will also be interesting to apply the system to the double-stranded regions of ε where previous analyses of individual mutants have not led to unambiguous conclusions about structure versus sequence-specific effects (9,10). A practical application will be the search for nucleic acid-based competitive inhibitors of the authentic P /ε interaction which would only have to support binding.

ACKNOWLEDGEMENTS

We thank H. Schaller for critical discussions and continous support and I. Mattaj for providing plasmid pBS-Ad1. This work was supported by the Deutsche Forschungsgemeinschaft through grants Na154/3-1 and SFB229-D3.

REFERENCES

1 Nassal, M. and Schaller, H. (1993) *Trends Microbiol.*, **1**, 221–226
2 Nassal, M. and Schaller, H. (1993) In Doerfler, W. and Boehm, P. (eds.) *Virus Strategies*. VCH, Weinheim, pp. 41–75
3 Junker-Niepmann, M., Bartenschlager, R. and Schaller, H. (1990) *EMBO J.*, **9**, 3389–3396
4 Bartenschlager, R., Junker-Niepmann, M. and Schaller, H. (1990) *J. Virol.*, **64**, 5324–5332
5 Hirsch, R.C., Lavine, J.E., Chang, L.J., Varmus, H.E. and Ganem, D. (1990) *Nature*, **344**, 552–555
6 Wang G.H. and Seeger, C. (1992) *Cell*, **71**, 663–670
7 Pollack, J.R. and Ganem, D. (1994) *J. Virol.*, **68**, 5579–5587
8 Knaus, T. and Nassal, M. (1993) *Nucleic Acids Res.*, **21**, 3967–3975
9 Pollack, J.R. and Ganem, D. (1993) *J. Virol.*, **67**, 3254–3263
10 Fallows, D. and Goff, S.P. (1995) *J. Virol.*, **69**, 3067–3073
11 Gold, L., Tuerk, C., Allen, P., Binkley, J., Brown, D., Green, L., MacDougal, S., Schneider, D., Tasset, D. and Eddy, S.R. (1993) In Gesteland, R.F. and Atkins, J.F. (eds) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 497–510
12 Szostak, J.W. and Ellington, A.D. (1993) In Gesteland, R.F. and Atkins, J.F. (eds) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 511–534
13 Burke, J.M. and Berzal-Herranz, A.(1993) *FASEB J.*, **7**, 106–112
14 Berkhout, B. and Klaver, B. (1993) *Nucleic Acids Res.*, **21**, 5020–5024
15 Klaver, B. and Berkhout, B. (1994) *EMBO J.*, **13**, 2650–2659
16 Nassal, M. and Rieger, A., submitted
17 Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., MacKay, P., Leadbetter, G. and Murray, K. (1979) *Nature*, **282**, 575–579
18 Konarska, M.M. and Sharp, P.A. (1987) *Cell*, **49**, 763–774
19 Nassal, M., Junker-Niepmann, M and Schaller, H. (1990) *Cell*, **63**, 1357–1363
20 Nassal, M. (1992) *J. Virol.*, **66**, 4107–4116
21 Crowther, R.A., Kiselev, N.A., Bsttcher, B., Berriman, J.A., Borisova, G.P., Ose, V. and Pumpens, P. (1994) *Cell*, **77**, 943–950
22 Kenney, J.M., von Bonsdorff, C.-H., Nassal, M. and Fuller, S.D. (1995) *Structure*, in press
23 Lok, A.S.F. (1994) *J. Viral Hepatitis*, **1**, 105–124
24 Nassal, M. (1995) In KrSusslich, H.-G. (ed.) *Retroviral Morphogenesis. Current Topics Microbiology Immunology*. Springer, Heidelberg, in press
25 Henderson, B.R., Menotti, E., Bonnard, C. and KYhn, L.C. (1994) *J. Biol. Chem.*, **269**, 17 481–17 489
26 Olsthoorn, R.C.L., Licis, N. and van Duin, J. (1994) *EMBO J.*, **13**, 2660–2668
27 Steitz, T. (1993) In Gesteland, R.F. and Atkins, J.F. (eds) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 219–238
28 Yu, M. and Summers, J. (1991) *J. Virol.*, **65**, 2511–2517
29 Riordan, F.A., Bhattacharyya, A., McAteer, S. and Lilley, D.M. (1992) *J. Mol. Biol.*, **226**, 305–310
30 Laskus, T., Rakela, J. and Persing, D.H. (1994) *Virology*, **200**, 809–812
31 Lok, A.S.F., Akarca, U. and Greene, S. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4077–4081
32 Tong, S.P, Vitvitsky, L. and Trepo, C. (1992) *Virology*, **191**, 237–245
33 Wang, G.H., Zoulim, F., Leber, E.H., Kitson, J. and Seeger, C. (1994) *J. Virol.*, **68**, 8437–8442
34 Knaus, T. and Nassal, M., unpublished data
35 Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A. (1992) *Science*, **256**, 1783–1790
36 Jacobo-Molina, A., Ding, J., Nanni, R., Clark, A. D., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. and Arnold, E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6320–6324
37 Rodgers, D.W., Gamblin, S.J., Harris, B.A., Ray, S., Culp, J.S., Hellmig, B., Woolf, D.J., Debouck, C. and Harrison, S.C. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 1222–1226
38 Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. and Liljas, L. (1994) *Nature*, **371**, 623–626
39 Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. (1991) *Prog. Nucleic Acids Res. Mol. Biol.*, **40**, 185–220
40 Baker, B., Muckenthaler, M., Vives, E., Blanchard, A., Braddock, M., Nacken, W., Kingsman, A.J. and Kingsman, S.M. (1994) *Nucleic Acids Res.*, **22**, 3365–3372
41 Gait, M.J. and Karn, J. (1993) *TIBS*, **18**, 255–259
42 Wang, G.H. and Seeger, C. (1993) *J. Virol.*, **67**, 6507–6512
43 Tavis, J.E. and Ganem, D. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4107–4111