Sequence- and Structure-Specific Determinants in the Interaction between the RNA Encapsulation Signal and Reverse Transcriptase of Avian Hepatitis B Viruses

JÜRGEN BECK AND MICHAEL NASSAL*

Zentrum für Molekulare Biologie, University of Heidelberg, D-69120 Heidelberg, Germany

Received 23 December 1996/Accepted 25 March 1997

Hepatitis B viruses (HBVs) replicate by reverse transcription of an RNA intermediate. Packaging of this RNA pregenome into nucleocapsids and replication initiation depend crucially on the interaction of the reverse transcriptase, P protein, with the cis-acting, 5′-end-proximal encapsidation signal ε. The overall secondary structure is similar in all of the hepadnaviral ε signals, with a lower and an upper stem, separated by a bulge, and an apical loop. However, while ε is almost perfectly conserved in all mammalian viruses, the ε signals of duck HBV (DHBV) and heron HBV (De and He, respectively) differ substantially in their upper stem regions, both in primary sequence and in secondary structure; nonetheless, He interacts productively with DHBV P protein, as shown by its ability to stimulate priming, i.e., the covalent attachment of a deoxynucleoside monophosphate to the protein. In this study, we extensively mutated the variable and the conserved positions in the upper stem of De and correlated the functional activities of the variant RNAs in a priming assay with secondary structure and physical P protein binding. These data revealed a proper overall structure, with the bulge and certain key residues, e.g., in the loop, being important constraints in protein binding. Many mutations at the evolutionarily variable positions complied with these criteria and yielded priming-competent RNAs. However, most mutants at the conserved positions outside the loop were defective in priming even though they had ε-like structures and bound to P protein; conversely, one point mutant in the loop with an apical structure different from those of De and He was priming competent. These results suggest that P protein binding can induce differently structured ε RNAs to adopt a new, common conformation, and they support an induced-fit model of the ε-P interaction in which both components undergo extensive structural alterations during formation of a priming-competent ribonucleoprotein complex.

The hepadnavirus family comprises several mammalian members, including the prototypic human hepatitis B virus (HBV), and two less closely related avian representatives, duck HBV (DHBV) and heron HBV (HHBV). All hepadnaviruses are small, enveloped DNA-containing viruses that replicate by reverse transcription (for reviews, see references 18 and 19) of an RNA intermediate, the pregenome, which also serves as mRNA for the capsid protein and the reverse transcriptase (P protein). Specific encapsidation of the RNA pregenome into nucleocapsids (10, 13, 21) and initiation of its reverse transcription (15, 20, 26, 27, 32) are mediated by the interaction of P protein, probably in concert with cellular chaperones (8), with the structured, cis-acting ε signal (Fig. 1A).

Human HBV ε adopts a characteristic bulge-and-loop secondary structure (13, 21). Its sequence and structure are highly conserved in all mammalian hepadnaviruses (16), and its 5′ end-proximal ε signal is sufficient for specific RNA encapsidation (10) and replication initiation (24). In the avian viruses, a second, downstream element (region 2; see reference 6) is required, but otherwise the ε signals of DHBV and HHBV (De and He, respectively) are clearly related to human HBV ε in function and overall structure (Fig. 1B); both contain a bulge (for sake of simplicity we use this term rather than the correct term, asymmetric internal loop; see reference 17) separating the lower and upper stems and an apical loop (4). In He, however, the region corresponding to the upper stem in De not only diverges in primary sequence from De at seven positions but also contains many fewer Watson-Crick base pairs. Nonetheless, He interacts productively (33), and with only about twofold lower efficiency than De (4), with DHBV P protein. While this promiscuity suggests a rather relaxed specificity, human HBV ε is completely discriminated against by DHBV P protein (22), indicating the existence of specific constraints for productive binding. To elucidate the sequence- and structure-specific determinants underlying the interaction between P protein and ε, we analyzed a series of avian ε mutants in which we simultaneously replaced several nucleotides (nt) in the apical region, either 6 of the 7 residues that are variable between De and He (v positions) or all 3 residues in this region that are conserved between DHBV and HHBV (k positions); in addition, we altered two positions in the De loop sequence that is completely conserved in He. In most cases, the new nucleotides did not correspond to the authentic De or the He sequence (Fig. 1B and Table 1).

First, we tested productive interaction with DHBV P protein of the mutant RNAs using a trans-priming assay that measures the ε-dependent covalent transfer to P protein of a deoxynucleoside monophosphate (dNMP) from the corresponding deoxynucleoside triphosphate (dNTP); this assay revealed a variety of activities whose levels ranged from zero to essentially wild-type (wt) levels. We then correlated functional activity with the predicted and, for several representative variants, experimentally determined secondary structures. Finally, we used a direct assay to distinguish physical and productive binding to DHBV P protein. Our experiments revealed a surprising flexibility in primary sequence in particular at the variable positions. The inactive variants, on the other hand, al-
that are assumed to have direct contacts with the protein be-

TABLE 1. Structural and functional properties of D variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>Mutated positions</th>
<th>ΔΔG°* (kcal/mol)</th>
<th>Experimental two-dimensional structure</th>
<th>P protein interactiona</th>
</tr>
</thead>
<tbody>
<tr>
<td>V series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De</td>
<td>AUCUGU</td>
<td>0</td>
<td>wt De</td>
<td>+++ + + + +</td>
</tr>
<tr>
<td>He</td>
<td>CCUCUCU</td>
<td>1.2</td>
<td>wt He</td>
<td>+++ + + + +</td>
</tr>
<tr>
<td>V1</td>
<td>GAGCAA</td>
<td>&lt;0</td>
<td>He-like</td>
<td>+++ + + + +</td>
</tr>
<tr>
<td>V3</td>
<td>UUACAC</td>
<td>0.9</td>
<td>He-like</td>
<td>+++ + + + +</td>
</tr>
<tr>
<td>V7</td>
<td>UAGCAA</td>
<td>&lt;0</td>
<td>He-like</td>
<td>+++ + + + +</td>
</tr>
<tr>
<td>V28</td>
<td>GGACAA</td>
<td>2.0</td>
<td>+ + + ND</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td>GAAGAC</td>
<td>8.8</td>
<td>+ + + ND</td>
<td></td>
</tr>
<tr>
<td>V5</td>
<td>UAAGAC</td>
<td>7.9</td>
<td>non-De and non-He</td>
<td>+ + + ND</td>
</tr>
<tr>
<td>V24</td>
<td>GAAGAA</td>
<td>6.2</td>
<td></td>
<td>+ + + ND</td>
</tr>
</tbody>
</table>

| K series |                 |                 |                                      |                        |
| wt De or He | UGU |                 |                                      |                        |
| K2      | AAC              | 2.4             | + + ND                               |                         |
| K9      | CAC              | 2.4             | He-like                              | − + + +                |
| K10     | CCA              | 1.6             | He-like                              | − + + +                |
| K11     | GAC              | 2.6             | + + ND                               |                         |
| K12     | GUU              | 1.6             | + + + ND                             |                         |
| K22     | CCG              | 0.3             | + + + ND                             |                         |
| K23     | CAA              | 1.6             | + + + ND                             |                         |
| K26     | CUC              | 2.4             | + + + ND                             |                         |

| L mutants |                 |                 |                                      |                        |
| wt       |                  |                 |                                      |                        |
| L4      | CU               | 0               | wt De                                | − − − −                |
| L5      | UA               | 2.5             | non-De and non-He                    | +++ + + + +            |
| L4,5    | CA               | 1.4             | + + + ND                             |                         |

a For structural context, see Fig. 1B. The positions mutated in the V series were v1 to v6, those mutated in the K series were k1 to k3, and those mutated in the L mutants were l4 and l5.

b ΔΔG = ΔG − ΔGnon, i.e., the difference in free energy (ΔΔG) between the most stable ε-like (ΔGε) and non ε-like (ΔGnon ε) structures as calculated by M-FOLD.

† Priming activity at 1 μM mRNA RNA was normalized to wt De (100%) and the De-independent priming signal (0%). Binding efficiency at 80 nM ε RNA was normalized to wt De (100%) and P-protein-independent background binding (0%). + + +, 50 to 100%; ++, 20 to 50%; +, 10 to 20%; ±, 1 to 10%; −, <1%; ND, not determined.

V series

<table>
<thead>
<tr>
<th>Variant</th>
<th>Mutated positions</th>
<th>ΔΔG°* (kcal/mol)</th>
<th>Experimental two-dimensional structure</th>
<th>P protein interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>De</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MATERIALS AND METHODS

Chemicals, enzymes, and cells. Reagents for agarose gel electrophoresis were purchased from Serva (Heidelberg, Germany), and polyclonal antibody gels were cast from precast acrylamide solutions (AppliChem, Darmstadt, Germany). Enzymes for molecular cloning experiments were obtained from Boehringer (Mannheim, Germany) or New England Biolabs (Bad Schwalbach, Germany). For enzymatic structure probing, RNase A and T1 from Boehringer were used. In vitro translations were performed with rabbit reticulocyte lysate from Promega (Heidelberg, Germany). Radiolabeled protein labeling were obtained from Amersham (Braunschweig, Germany). All plasmids were transformed into Escherichia coli Top 10 cells (Invitrogen, NV Leek, The Netherlands) and grown

caused their substitution impedes protein binding. The conserved residues outside the loop apparently have a special role, as several mutants bound to P failed to support priming. However, we also identified a point variant with a secondary structure whose apical region differs from those of both De and He but still mediates priming. Together, these data show that a complex but tractable hierarchy of structure and sequence-specific determinants is involved in productive ε-P protein interaction; they also suggest structural flexibility within the RNA as an important criterion for protein binding.

FIG. 1. (A) Functional roles of ε-P interaction. The line symbolizes a linear version of the circular DHBV genome, the diamond represents the polyadenylation signal, and the open rectangles above the line represent the three major ORFs. Numbers are nucleotide positions. The wavy lines represent the termination signal, and the open rectangles above the line represent the three major termination. The models shown are those in best accord with direct structural data (4). Positions varying between Hε and Hε are shown as boxes. Region 2 is an as yet not well-defined second element required for pregenome encapsidation. Binding of P protein to 5’-P triggers addition of ε-P protein in-

allowed us to determine several general requirements for productive binding: one is the formation of an ε-like overall structure and another is the presence of certain key residues that are assumed to have direct contacts with the protein be-
Design of mutant De sequences. Six of the seven positions where He differs from De are clustered at the base of the Dε loop (designated v1 to v6 [Fig. 1B]), while the loop sequence itself and three residues at its boundaries (designated k1 to k3) are identical. The term loop sequence is here defined as the apical 7 nucleotides that were predicted to be unpaired in the originally proposed De structure model (10), although direct data are more compatible with the slightly different arrangement shown in Fig. 1 (4). Assuming that sequence conservation and divergence in the authentic viral RNA elements may reflect different structural and functional contributions from the corresponding positions, we chose to separately mutate the k and the v positions.

To this end, we introduced into the previously described Dε plasmid (3) synthetic degenerate DNA cassettes in which either the v or k residues were replaced by, mostly, non-De or non-He nucleotides. From randomly picked colonies we isolated seven different V and nine K variants. Their sequences are listed in Table 1, together with those of the loop variants L4, L5, and L4,5, in which, in the context of wt De, the U residues at loop positions 4 and 5, either alone or in combination, were exchanged for C or A, respectively. L4 and L4,5 have previously been shown to have strongly reduced activities in the priming assay (2).

Functional analysis of De and He variants by in vitro priming. DHBV P protein, in vitro translated in rabbit reticulocyte lysate, exhibits enzymatic activity (31) as measured in a priming assay (Fig. 2A). A functional De sequence, either on the RNA used for P protein in vitro translation (which contains region 2 of the encapsidation signal as part of the P ORF) or added as a separate RNA in trans, mediates the covalent transfer to the protein of an appropriate dNMP from the corresponding [32P]dNTP (34). This reaction mimicks the first step of reverse transcription of the RNA pregenome and provides a sensitive test for the functional integrity of variant De sequences; a negative result does not, however, discriminate between a lack of physical versus productive binding (see below). In our study, the variant ε sequences were in vitro transcribed, isolated, and subjected to the priming assay at a standard trans-RNA concentration of 1 μM (for wt De RNA, P protein labeling reaches a maximum at this concentration; see references 4 and 33). Usage of equal aliquots from a single preparation of P protein served to exclude experimental variation in the levels of efficiency of in vitro translation.

The results obtained with the individual variants are shown in Fig. 2B. Most evident were the extensive variation in the signal intensities produced by the V variants, ranging from almost background to wt levels (compare, e.g., Fig. 2B, lanes V3 and V5), and the general strong reduction in signal intensities observed for the K variants (e.g., lanes K2 to K26). Of the loop mutants, L4 and L4,5 were essentially inactive while L5 produced a relatively strong signal (L lanes). Semiquantitative values were obtained by measuring the signal intensities with the PhosphorImager (Table 1). Hence, on one hand, a surprising large number of nucleotide exchanges were tolerated without functional consequences, in particular at the v positions, while other combinations, even of a few foreign nucleo-
otides as in the K variants or a single substitution as in variant L4, led to reduction if not complete loss of function. With the following experiments, we sought to link these functional differences with the secondary structures of the individual RNAs.

High sequence flexibility at the variable positions in the context of an ε-like overall structure. First, we used the M-FOLD program (38) to predict the secondary structures of the variant ε RNAs, although the validity of this theoretical approach is limited in some respects; for instance, the program slightly favors non-ε-like secondary structures for De and He over the experimentally confirmed bulge-and-loop arrangement (4). For wt De and He, the calculated differences in free energy between ΔG values of the most stable ε-like and non-ε-like structures (termed ΔΔG values and calculated for ε RNAs of 76 nt in length) are rather small (0 kcal/mol for De and 1.2 kcal/mol for He). Many variants exhibited similarly low ΔΔG values ranging from <0 to 2.6 kcal/mol, which would not allow for unambiguous statements about their structures. However, for the V variants we noted a striking correlation between inactivity in the priming assay and a large positive ΔΔG value (>6.0 kcal/mol) (Table 1). Such a large unfavorable ΔΔG plausibly reflects the possibility that the corresponding RNAs adopt a stable non-ε-like structure that is incompatible with P protein binding. To test this notion, we selected a representative pair of V mutants with drastically different activities, namely V3 and V5, for direct enzymatic secondary structure analysis. Both variants contain the same non-De and non-He nucleotides at five of the six mutated v positions and differ only at v4. Mutant V3 carries a C residue, displays near wt priming activity, and has a small ΔΔG value (0.9 kcal/mol). V5, with a G residue at the corresponding position, is essentially inactive and exhibits a large ΔΔG (7.9 kcal/mol). This pair of mutants was therefore particularly attractive for distinguishing between structural and sequence-specific contributions to productive P protein binding.

While we have not completely determined the structures of these RNAs, we took advantage of the fact that the existence of an ε-like structure can be easily monitored by its sensitivity to digestion with RNase A at U2574 in the De bulge. The absence of this signal, in turn, indicates the absence of the ε-like bulge; similarly, G2589 is a hallmark residue in the wt De loop (Fig. 1; these residues are also explicitly marked in Fig. 4 to 7). For simpler comparison, we use the DHBV numbering system also to indicate the corresponding residues in He (authentic HHBV numbers may be easily obtained by adding 6 to the given DHBV position number).

As the presence of more than one conformer can lead to superposition of cleavage signals from different structures, we first analyzed De, He, and variants V3 and V5 as well as variant L5 (see below) by nondenaturing gel electrophoresis. Because of the known 3′ end heterogeneity of in vitro transcripts, RNAs of homogeneous lengths were isolated by electrophoresis on denaturing gels and renatured and their mobilities were compared under denaturing and native conditions (Fig. 3). Expectedly, all transcripts migrated identically in the presence of urea, while distinct mobility differences were revealed on the nondenaturing gel. All variants tested in this assay produced only one major band, suggesting the presence of one major conformer. Identical mobility patterns were obtained regardless of whether renaturation of the transcripts was performed in TE buffer or in the TMK buffer (not shown) used in the subsequent probing experiments. These were essentially performed as previously described for wt De and He (4).

The results for the active variant V3 and the inactive variant V5 are shown in Fig. 4A. Most evident was the very small number of accessible nucleotides in V5, all in a contiguous region partially overlapping the De loop; in particular, the bulge-specific signal at U2574 was virtually absent. V3, by
contrast, produced a series of signals, both at U2574 and in the apical region. Many of these coincided with nucleotides also accessible in He (Fig. 4A, He lanes). These data indicate a largely base-paired, non-ε-like arrangement for V5 that is fully compatible with the calculated most stable model (Fig. 4B), and that arrangement is further substantiated by the significantly increased mobility of V5 RNA on the nondenaturing gel (compare lanes V3 and V5 in Fig. 3). The active variant V3, by contrast, shares many structural features with He, especially the bulge and a relatively open apical region. We conclude that a stable non-ε-like structure interferes with productive binding to P, that even a single nucleotide exchange can bring about this type of drastic structural change, and that formation of such improper structures is predictable by a large ΔΔG value as defined above; this latter assumption is substantiated by the fact that all V variants with ΔΔG-values of >6.0 kcal/mol (Table 1) were inactive. Hence, the only major restriction for sequence divergence at the v positions seems to be that new nucleotides must not, in the given sequence context, induce formation of stable non-ε-like structures.

Sequence-specific constraints to priming competence at the conserved k positions despite an ε-like overall structure. All of the nine K variants displayed drastically reduced activities in the trans-priming assay (Fig. 2), suggesting that at least one of the k positions contributes in a sequence-specific way to productive P binding. In contrast to those of the V variants, the calculated ΔΔG values were small and gave no clear indications for stable non-ε-like structures. We therefore directly analyzed the secondary structure of the best-performing variant, K12, which had about 15% of the activity of wt De, and that of K9, which was essentially inactive; their primary sequences differ at k positions 1, 2, and 3 (K12, GUA; K9, CAC). The nuclease sensitivity patterns of both variants (Fig. 5A) were similar to that of wt He; the bulge indicator at nt U2574 was clearly detected, and the series of further products indicated a similarly nonpaired apical region. One apparent difference was the relatively strong signal, in K12, at position U2597, which was weak in wt He and virtually absent from K9. How this relates to the partial priming proficiency of K12 is currently not clear. However, while subtle structural differences cannot be monitored by enzymatic probing, the data provide direct evidence that it is not overall structure, as with V5, that reduces their priming competence but rather the nature of the replaced nucleotides. Surprisingly, however, the lowered or abolished priming competence of the K mutants did not correlate with an inability to bind to P protein (see below).

Drastically different consequences of mutations at individual positions in the De loop sequence. All of the variants described above contained multiple mutations in comparison to De and He. The loop sequence (positions 2587 to 2593 in De and positions 2593 to 2599 in He) is completely conserved between the two viruses, and several previously published data had suggested that the loops in De (22, 28) and human HBV ε (21, 23) contain important sequence-specific determinants for P binding. The loop sequence (positions 2587 to 2593 in De and positions 2593 to 2599 in He) is completely conserved between the two viruses, and several previously published data had suggested that the loops in De (22, 28) and human HBV ε (21, 23) contain important sequence-specific determinants for P binding. The three loop variants we analyzed differ from wt De only at one or two positions (in L4, U2590 is replaced by C; in L5, U2591 is replaced by A; and L4,5 contains both mutations). In accord with previous data (2), L4 and L4,5 exhibited drastically reduced priming activities, but L5, at a concentration of 1 μM, gave a signal of about two-thirds of that of wt De (Fig. 2B, L lanes).

Surprisingly, the experimental structure analysis of the variants (Fig. 6) revealed that the inactive mutant L4 produced a nuclease sensitivity pattern that was absolutely indistinguishable from that of wt De (Fig. 6A; compare De and L4 lanes); a similar result was obtained with L4,5 (L4,5 lanes). Hence, despite having a De-like secondary structure, these mutants are unable to productively interact with P protein; the single nucleotide exchange in L4 acts on the sequence-specific rather than on the overall structural level. The active variant L5 showed clear signals around U2574, which, though they were weaker than in wt De, indicated that the bulge nucleotides were exposed; however, a whole series of new fragments was
generated in a region corresponding to the upper-right half-
stem of De (Fig. 6A, L5 lanes) and the strongest loop-specific
signal was shifted from G2589 to G2586. This nuclease pattern
does not accurately correspond to any of the stable conforma-
tions predicted by M-FOLD. To test whether L5 RNA adopts
two or more similarly stable conformations producing overlap-
ing patterns, we also subjected this variant to nondenaturing
gel electrophoresis; as only a single major band was observed
between physical and productive binding that leads to priming.
The assay relies on immobilizing an N-terminally His-tagged
DHBV P protein to a Ni²⁺-containing matrix; binding of
labeled ε RNA to P protein is monitored by scintillation count-
ing and reanalysis of the bound RNA by denaturing PAGE.
RNA immobilization to the matrix is P protein dependent and
specific, as De but not human HBV ε RNA is retained by
DHBV P protein (3).

In one set of experiments, we compared the amounts of
various internally labeled ε RNAs which, after incubation at an
average RNA concentration of about 80 nM with in vitro-
translated P protein, were immobilized to the solid phase (Fig.
7A). A semiquantitative evaluation of these data is shown in
Table 1. The priming-incompetent variants V5 and L4 did not
significantly bind to P protein, confirming our conclusion that
an improper overall structure (V5) or mutation of a key resi-
due (such as U2590 in L4) prevents binding. However, not only
the priming-competent wt RNAs and variants V3 and L5 but
also K12, K9, and, to a lesser extent, K2 (not shown) bound to
P protein, despite their severely reduced or even completely
negative priming phenotype.

These results were further corroborated in priming compe-
tition experiments. Here, a normal trans-priming reaction with
200 nM wt De RNA was supplemented with a 10-fold molar
excess of several K mutants and the nonbinding L4 variant. In
this format, binding-competent variants were expected to com-
pete with wt De RNA and hence to reduce the priming signal.
In accord with their inactivity at a concentration of 80 nM (Fig.
2B), the tested K variants gave signals only slightly above
background when they were present alone at a 2 µM concen-
tration (Fig. 7B, lanes without wt De). However, all K variants,
but not L4, substantially diminished wt-De-mediated labeling
of P protein (Fig. 7B, lanes with wt De) and therefore acted as
competitive inhibitors. Hence, many K variants are able to
interact with P protein despite their severe priming defects.
This clearly attests to the existence of requirements for priming
that go beyond mere physical binding.

**DISCUSSION**

The ε sequence is highly conserved in all mammalian hepa-
titis B viruses. Not unexpectedly, therefore, the ε signals and P
proteins of human HBV and woodchuck HBV (WHV) can
functionally substitute for each other (37). In this evolutionary
light, the substantial sequence and structure divergence in the
apical part of the avian hepadnavirus ε signals is per se re-
markable; even more surprising is the similarly efficient inter-
action of De and He with DHBV P protein (4, 33).

The combined structural and functional analysis of the De
variants reported in this paper allows for several generaliza-
tions that explain this apparent promiscuity: (i) in accord with
their divergence in De and He, the v positions tolerate many
mutations and hence are not essential for sequence-specific
contacts with protein; (ii) even here, however, single nucleo-
tide exchanges can be deleterious if they induce stable, non-
bulge-containing structures; (iii) such nonproductive structures

**FIG. 5.** Secondary structures of variants K9 and K12. (A) Enzymatic probing.
Probing data were obtained as described for variants V3 and V5 in the legend to
Fig. 4. (B) Secondary structure models. Cleavage positions derived from the
results of experiments shown in panel A are marked by arrows (see the legend to
Fig. 4B for clarification); the structures shown are those that fit best the exper-
imental data. Authentic residues at the k positions are indicated inside the loop;
the asterisks indicate cleavage sites containing the preferred RNase A target
pyrimidine A.
can be predicted by a large unfavorable ΔΔG value; (iv) certain key residues act as sequence-specific determinants, as demonstrated by the loop variant L4, in which replacement of a single U prevents protein binding despite preservation of the wt De secondary structure; (v) physical binding is insufficient for functional activity as illustrated by many k-position mutants; and (vi) not only the natural He RNA but also an artificial De variant, L5, with a substantially differing apical structure can productively interact with DHBV P protein. Based on these results, we propose a dynamic model for the ε-P interaction that accounts for the phenotypes of all analyzed RNAs.

Sequence flexibility and constraints in the apical part of avian hepadnavirus encapsidation signals. The initial functional screen by the trans-priming assay revealed drastically different activities for the V variants, while most mutations at the conserved k positions and a single mutation at one position (L4), but not at another position (L5), markedly reduced or abolished productive binding (Table 1).

The straightforward conclusion from the activity of many V variants with non-De and non-He nucleotides at five of six positions is that many different sequences fulfill all requirements for productive interaction with P protein. Without consideration of the role of the De or He sequence as part of the pre-C ORF, this result indicates that there is little selective pressure on conserving the primary sequence at the v positions; this conclusion plausibly explains the divergence between De and He. The v positions also show the highest variability when the ε signals of all DHBV sequences deposited in the GenBank database are compared (not shown). In addition, these data argue strongly against sequence-specific contributions of the v positions to P protein binding, since active RNAs containing each of the four possible nucleotides were found. Nonetheless, some V mutants were essentially inactive; our data indicate that this defect is caused by the formation of stable, non-ε-like structures. The inactive variant V5 showed very few nuclease-sensitive sites, and the bulge-specific RNase A cleavage products were virtually absent. The existence of the largely base-paired structure lacking the bulge derived from these experiments is further supported by the high mobility of V5 RNA on native polyacrylamide gels (Fig. 3); bulges, as present in wt De and He RNAs and in the active variant V3, have been demonstrated to induce kinks that slow down migration (5).

The proposed V5 structure also conforms to the most stable structure predicted by the M-FOLD program. While it is difficult to judge the relevance of alternative predicted structures
with similar calculated free energies, as with wt Dε and Hε (4), the differences in stability (Table 1) between ε-like and non-ε-like structures for V5 and all other inactive V variants were drastically higher (ΔΔG = 6.2 to 8.8 kcal/mol) than for the active variants (ΔΔG = ±2.0 kcal/mol); also, the absolute ΔG values for these improper structures were much larger than for wt Dε and the active variants (ΔG = −16 to −18 kcal/mol versus −10 to −13 kcal/mol). Hence, a large positive ΔΔG value is most likely a valid criterion to predict the inactivity of ε mutants. The low intrinsic stability of the ε-like arrangement in the avian signals is probably the reason for the dramatic influence that sometimes even single nucleotide substitutions had on overall RNA structure. For human HBV ε, by contrast, the experimentally confirmed bulge-and-loop arrangement is also calculated to be much more stable than any non-ε-like structure (ΔΔG = −6.0 kcal/mol for a 57-nt ε RNA).

For the K variants, as well as the loop mutants, the ΔΔG criterion was not applicable, as all calculated stability differences were low (less than 2.6 kcal/mol); this suggested in turn that grossly aberrant overall structures were not responsible for their greatly reduced or abolished activities. Of the nine k-position mutants tested in the priming assay, only K12 produced a reasonable signal (about 15% of that of wt Dε). Enzymatic probing of variants K12, K9, and K10 (not shown) produced signals compatible with all RNAs having ε-like features, in particular, the typical bulge. In the apical region, relatively many signals were observed, mostly at the same positions as in Hε. Currently, we cannot rule out the possibility that minor structural alterations compared with Dε or Hε contribute to the reduced priming activities of the K variants; also, we have not explicitly tested whether these RNAs are present in one or more different conformations. However, our data suggest that the K variants have similar overall ε-like structures and that their priming defects relate to sequence rather than structure. This argues for an involvement of one or more of the k positions in sequence-specific interactions with the protein that do not, however, act on the level of physical but of productive binding (see below). The functional defects of the K variants are mirrored by the evolutionary conservation of the k positions in Dε and Hε and also in the other known DHBV sequences.

A clear-cut example for the importance of RNA sequence in P protein recognition was provided by variant L4, in which the replacement of U2590 by C, in accord with previous observations (2; mutant SLM13 in reference 29), drastically reduced functional activity but did not induce any detectable structural differences from wt Dε. This striking dependence of priming (and binding to the protein [see below]) on the chemical nature of the base in the context of an unaltered overall structure suggests U2590 as a direct contact point between RNA and protein.

In this light, the substantial priming activity of mutant L5 was unexpected, as in this mutant U is replaced by the chemically less related A. With the caveat that the observed nucleo-

4978 BECK AND NASSAL J. VIROL.

Downloaded from http://jvi.asm.org on April 23, 2021 by guest

se products arose from more than one RNA conformer present in the test solution (although we detected only one major band on native polyacrylamide gels), the presence of the nucleo-

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.

4978 BECK AND NASSAL J. VIROL.
FIG. 8. Models for the interaction between avian hepadnavirus ϵ signals and DHBV P protein. (A) Lock-and-key model. P protein has a binding pocket whose shape determines whether an RNA can bind or not; binding is prevented by an improper overall RNA structure (e.g., that of variant V5) or the absence of one or more key residues, even in the context of a correct overall structure (variant L5; ×). This simple model, however, does not distinguish between physical and productive binding, and it does not explain productive binding by RNAs with differently structured apical regions. (B) Induced-fit model. As in panel A, overall structure and key residues are important, but in addition, structural flexibility in the apical ϵ part allows for a conformational change in the RNA (and, possibly, the protein). All binding-competent RNAs meet this criterion, but priming competence requires additional structural alterations in the RNA and the protein (see reference 29; the structural alterations are symbolized by the different shape of the P protein); in binding-competent but priming-deficient RNAs, such as several K variants, the ability to undergo such secondary conformational changes appears to be defective.

REFERENCES