Thermodynamics and NMR studies on *Duck*, *Heron* and *Human* HBV encapsidation signals

Frederic C. Girard, Otmar M. Ottink, Kirsten A.M. Ampt, Marco Tessari and Sybren S. Wijmenga*

Department of Physical Chemistry/Biophysical Chemistry, Radboud University, Toernooiveld 1 6525 ED Nijmegen, The Netherlands

ABSTRACT

Hepatitis B virus (HBV) replication is initiated by binding of its reverse transcriptase (P) to the apical stem-loop (*AL*) and primer loop (*PL*) of epsilon, a highly conserved RNA element at the 5′-end of the RNA pregenome. Mutation studies on *duck/heron* and *human in vitro* systems have shown similarities but also differences between their P–epsilon interaction. Here, NMR and UV thermodynamic data on *AL* (and *PL*) from these three species are presented. The stabilities of the *duck* and *heron AL*s were found to be similar, and much lower than that of *human*. NMR data show that this low stability stems from an 11-nt internal bulge destabilizing the stem of *heron AL*. In *duck*, although structured at low temperature, this region also forms a weak point as its imino resonances broaden to disappearance between 30 and 35°C well below the overall *AL* melting temperature. Surprisingly, the *duck-* and *heron AL*s were both found to be capped by a stable well-structured UGUU tetraloop. All *avian AL*s are expected to adhere to this because of their conserved sequence. *Duck PL* is stable and structured and, in view of sequence similarities, the same is expected for *heron -* and *human PL*.

INTRODUCTION

The Hepatitis B virus (HBV) is the most common cause of liver infection in the world (1,2). More than 300 million people worldwide are estimated to be chronically infected by HBV (2) and chronic HBV infection carriers have a great risk to develop severe liver diseases, including cirrhosis and liver cancer, resulting in a million deaths annually (1). No treatment for the efficient elimination of HBV in infected patients exists as yet. HBV is a member of the *Hepadnaviridae* family, consisting of hepatotropic DNA-viruses, which also includes related animal viruses such as *duck* HBV (DHBV) and *heron* HBV (HHBV). HBV has a small (3.2 kb), relaxed circular, partially double-stranded DNA genome. It replicates its DNA genome by reverse transcription of an RNA intermediate, the pregenomic RNA (pgRNA), for reviews see (3–7). The pgRNA also serves as mRNA for the capsid (or core) and viral polymerase (P protein). The P protein contains the evolutionarily conserved reverse transcriptase (RT) domain, a middle spacer region, a C-terminal RNase H (RH) domain and a unique template domain (TP) at its N-terminus, which acts as a primer for reverse transcription (Figure 1A). Replication is initiated by binding of the P protein to the viral encapsidation signal (7), a 60-nt stem-loop, called epsilon, located at the 5′-end of the pgRNA (Figure 1A) (7–9). Binding of P to ε triggers two main events: recruitment of core proteins to form the viral capsid and synthesis of a 3–4-nt DNA primer, covalently attached to a tyrosine residue in the TP domain, using the ε primer loop as a template (Figure 1A). The resulting complex subsequently translocates to a 3′-proximal ε RNA element of the pgRNA where full-length DNA synthesis is primed using the 4-nt DNA primer (7–11).

Detailed knowledge on the crucial P–ε interaction has been derived from biochemical studies on a cell-free and chaperone-dependent reconstituted system based on *duck* HBV (8,10,12–16). The *in vitro* system shows both P–ε binding and priming. Using truncated P protein constructs, it was demonstrated that P–ε interaction requires sequences from both RT and TP protein domains (17). On the RNA side, it was established that the loop at the apex of apical stem-loop of DHBV-ε is essential for binding and primer synthesis (Figure 1A) (10). Recent SELEX experiments using this system further defined the structure and sequence elements in the apical stem-loop of DHBV crucial for binding and/or priming (16). For instance, the middle of the stem underlying the loop should be weakly- or not base paired at all. Based on these biochemical studies, Nassal cs. proposed that replication initiation is a two-step process in which initial physical RNA binding (and recognition) is followed by a structural rearrangement leading to a priming- competent complex
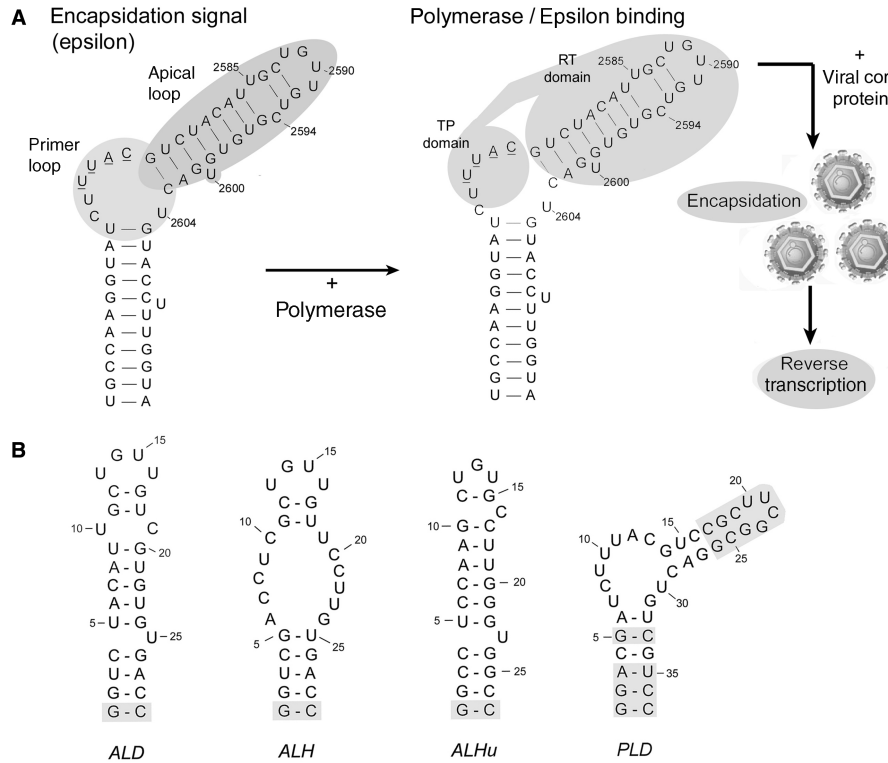
Figure 1. HBV encapsidation signal. (**A**) Schematic of the replication initiation, showing the P–ε interaction, which leads to encapsidation of the pgRNA and subsequent priming. (**B**) RNA constructs studied by NMR and UV melting corresponding to apical stem-loop of *duck* HBV (*ALD*), *heron* HBV (*ALH*), *human* HBV (*ALHu*) and primer loop of *duck* HBV (*PLD*). To aid in assignment of ALD and PLD, two additional shorter RNAs were expressed (**S6**). The gray-boxed residues represent structurally silent mutations from wild-type, introduced at the 5′- and 3′-ends to aid for expression or to cap and stabilize the upper stem (*PLD*). For ease of comparison the sequence numbering of Hu *et al.* (16) is also indicated, e.g. U_{15} of *ALD* (and *ALH*) is then numbered U_{2590}.

(10,16). Interestingly, in this *in vitro* system, the P protein binds to both *duck* ε, with a well-defined upper stem structure, and *heron* ε, characterized by a lack of base pairing in its upper stem, but it does not bind to *human* ε.

Most recently, a cell-free and chaperone-dependent *in vitro* reconstitution system was also developed for *human* HBV (15,18). It shows P–ε binding, but in contrast to the DHBV system, not priming. Similarly to DHBV, in *human* HBV, sequences from both the RT and TP domains are required for binding of P to ε. Surprisingly, the ε-apical loop is not needed for P–ε binding, in contrast to DHBV where it is essential. The ε-apical loop is, however, required for encapsidation. Moreover, the structural features, i.e. requirements for base pairing in the stem part of the apical stem-loop, differ from those in DHBV. In *human* HBV, the upper part of the stem of the apical stem-loop needs to be base paired and the bulged out U (Figure 1B) is essential for binding, while the corresponding bulged U in DHBV is dispensable.

Although the structural basis and sequence requirements for P–ε binding and priming are emerging, several questions remain and a full understanding of the molecular basis for the specific interactions between P and ε awaits high-resolution structural and thermodynamic data. The importance of high-resolution structural data is underlined by the NMR studies of the *human* HBV ε apical stem-loop, which showed that its conserved apical loop folds into a pseudo-triloop, whereas secondary structure programs predicted a hexaloop (19,20). Hairpin loops with the potential to form pseudo-triloops are found in many RNA sequences (21,22); their involvement in protein interactions and common appearance suggest that they may be an important protein-binding motif (19,22). This highly conserved structure motif is thus likely to be important for the interaction with P and/or capsid protein. Moreover, the *duck* P-protein can carry out binding and primer synthesis using *duck* ε as well as *heron* ε, but not *human* ε. This is remarkable considering that *duck* and *heron* ε are predicted to have completely different structural organization and thus stabilities (Figure 1B). Probing data on free *heron* ε suggest a fully open apical stem-loop, in contrast to *duck* ε, which appears to have a folded stem and loop region (10). This again raises the question concerning the exact structure or sequence determinants that trigger the ε–P binding interaction.

In this article, we present and compare structural (NMR) and thermodynamic data of the apical stem-loop and primer loop of ε RNAs from *duck*, *heron* and *human*. These data more precisely define the structure and stability of these RNA elements crucial for ε–P interaction and may aid in further understanding its molecular basis.

RESULTS

The ε RNA element contains two main structural features (Figure 1A) common to all HBV strains: (i) an apical

stem-loop (*AL*), recognized by the P protein to which it binds, triggering the reverse transcription process; (ii) an internal asymmetric loop, here referred to as primer-loop (*PL*), consisting of seven unpaired residues from which four residues are used by the polymerase as a template to synthesize a covalently bound DNA primer (5′-GTAA-3′ for *avian*, 5′-TGAA-3′ for *human*). Several of these RNA sequences were designed and expressed (see the Materials and methods section), based on the sequence of the wild-type strains reported for the *duck*-, *heron*- and *human* HBV. For sake of clarity, they are here referred to as *ALD* (ε *Apical Loop of Duck*), *ALH* (ε *Apical Loop of Heron*), *ALHu* (ε *Apical Loop of Human*) and *PLD* (ε *Primer Loop of Duck*) (Figure 1B).

NMR-based structure investigations

Duck HBV epsilon apical stem-loop (ALD). Resonance assignment was performed according to standard procedures (23,24). Of the 20 observed imino resonances (Figure 2A), 16 could be unambiguously assigned by combining data from [^{1}H,^{1}H]-NOESY and [^{1}H,^{15}N]-HSQC spectra in H_2O (Figure 2B and C, and S1) and extended to non-exchangeable resonances using [^{1}H,^{1}H]-NOESY spectra and [^{1}H,^{13}C]-HSQC in D_2O (Materials and Methods and Supplementary Data S1). The secondary structure as derived from NMR data is displayed in Figure 2D.

The lower stem (residues 1–4/26–29) conforms to an A-type helix as evidenced by the typical sequential and cross-strand NOE contacts (e.g. A_6-H2 ↔ G_{24}-H1′, A_{27}-H2 ↔ C_4-H1′), and by chemical shift values close to standard helix values (S2) (25). The central stem (residues 5–9/20–24) also forms an A-helix, again evidenced by the typical sequential and cross-strand NOE contacts of exchangeable and non-exchangeable protons (e.g. A_8-H2 ↔ G_{22}-H1′/G_{22}-H8/G_{22}-NH1, A_8-H1′ ↔ G_{22}-NH1, A_6-H2 ↔ G_{24}-H8, A_6-H1′ ↔ G_{24}-NH1) and standard A-helix chemical shifts. The lower and central stem experience stacking interaction as evidenced by the sequential NOE contacts (G_{26}-NH1 ↔ U_5-NH3 and between non-exchangeable protons of residues 4 and 5). Residue U_{25} is bulged out of the helix, because U_{25}-H5 proton resonates close to 6.0 p.p.m., experiencing little or no ring current (S2) and sequential NOE contacts between G_{24}/U_{25} and U_{25}/G_{26} are absent. Finally, a short A-helical upper stem (residues 11–12/17–18) is found as typical NOE contacts and A-helix chemical shifts are observed here. The residues of potential non-canonical base pair U_{10}:C_{19} shows sequential NOE contacts to both U_9:G_{20} and G_{11}:U_{18} characteristic for helical stacking. Furthermore, the H1′/H5/H6 chemical shifts of U_{10} and C_{19} are nearly A-helical, demonstrating the presence of ring current and thus, stacking interactions (S2). Therefore, both U_{10} and C_{19} must be inserted into the helix. Although one of two unassigned imino resonances could be the U_{10} imino, it exchanges too fast to allow the observation of NOEs. The NMR data for the structured tetraloop and its calculated structure are presented in the relevant section below.

COSY spectra show that most residues in the *ALD* adopt C3′-endo sugar puckering [$J_{H1'-H2'} < 3$ Hz, (23)] characteristic for an A-helix, only the sugar moieties of loop residues G_{14}, U_{15} and U_{16} adopt an C2′-endo conformation [$J_{H1'-H2'} > 6$ Hz, (23)]. Finally, the narrow spread around A-helix-like values of ^{31}P chemical shifts (data not shown), indicate that (on average) the phosphate backbone conforms to regular A-helix-like conformation (23).

Heron HBV epsilon apical stem-loop (ALH). In the ^{1}H-NMR imino spectra of *ALH*, ∼10 imino resonances are observed, of which 8 could be assigned (Figure 3, top). Of these, four could be assigned to the top stem-loop G_{11}CUGUUGU$_{18}$ via chemical shift comparison with the *ALD* spectra (Figure 3, bottom). The presence of imino resonances at the regular (Watson–Crick) chemical shifts shows the existence of a structured G_{11}CUGUUGU$_{18}$ stem-loop, analogous to what is observed for *ALD*. We do observe relative broadening compared to *ALD*, probably caused by increased solvent exchange, indicating a somewhat lower stability. The other four were sequentially assigned to the lower 4-bp stem based on NOEs. The large internal loop must be unstructured, as no evidence for base pairing was found within this region. The unassigned 2 broad resonances at ∼11 p.p.m. (indicated by asterisks) have chemical shifts typical of slightly protected iminos, not involved in base pairing. In conclusion, *ALH* conforms to the secondary structure displayed in Figure 1B.

Avian epsilon apical stem-loop: structure of the UGUU motif. Structure calculations were carried out on the highly conserved G_{11}CUGUUGU$_{18}$ *avian* top loop sequence, using the experimental NMR data collected for *ALD* (Materials and methods section; Structural Statistics in S7). After initial calculation using only classical NMR restraints, the resulting structures were further refined using chemical shifts as additional restraints (CS refinement). A well-defined tetraloop structure was found with RMSDs to the mean of 1.73 and 0.74 Å, including either all residues (11–18) or excluding flexible residues 14 and 15, respectively. The back-calculated chemical shifts agree generally well with the observed chemical shifts.

The resulting structure models indicate the formation of a U_{13}:U_{16} base pair stacked onto the C_{12}:G_{17} closing base pair, as well as stacking of G_{14} onto U_{13} on the major groove side, while U_{15} does not stack (Figure 2E, left, center). The stacking of the pertinent residues is confirmed by their back-calculated and experimental $\triangle\delta$ values close to 0 (Figure 2E, right). U_{15} does not stack at all, so that $\triangle\delta > 0$. This lack of stacking of U_{15} also leads to U_{16} $\triangle\delta > 0$, because aromatic chemical shifts are mostly affected by the residue on the 5′ side (25). Residues 11–13 and 16–18 converge well (Figure 2E, left). The convergence for G_{14} is somewhat less (Figure 2E, left), suggesting a certain degree of mobility. No convergence is reached for U_{15}, suggesting a large degree of flexibility. For this residue the experimental $\triangle\delta$ fall just outside the error bar of the back-calculated $\triangle\delta$ (Figure 2E, right).
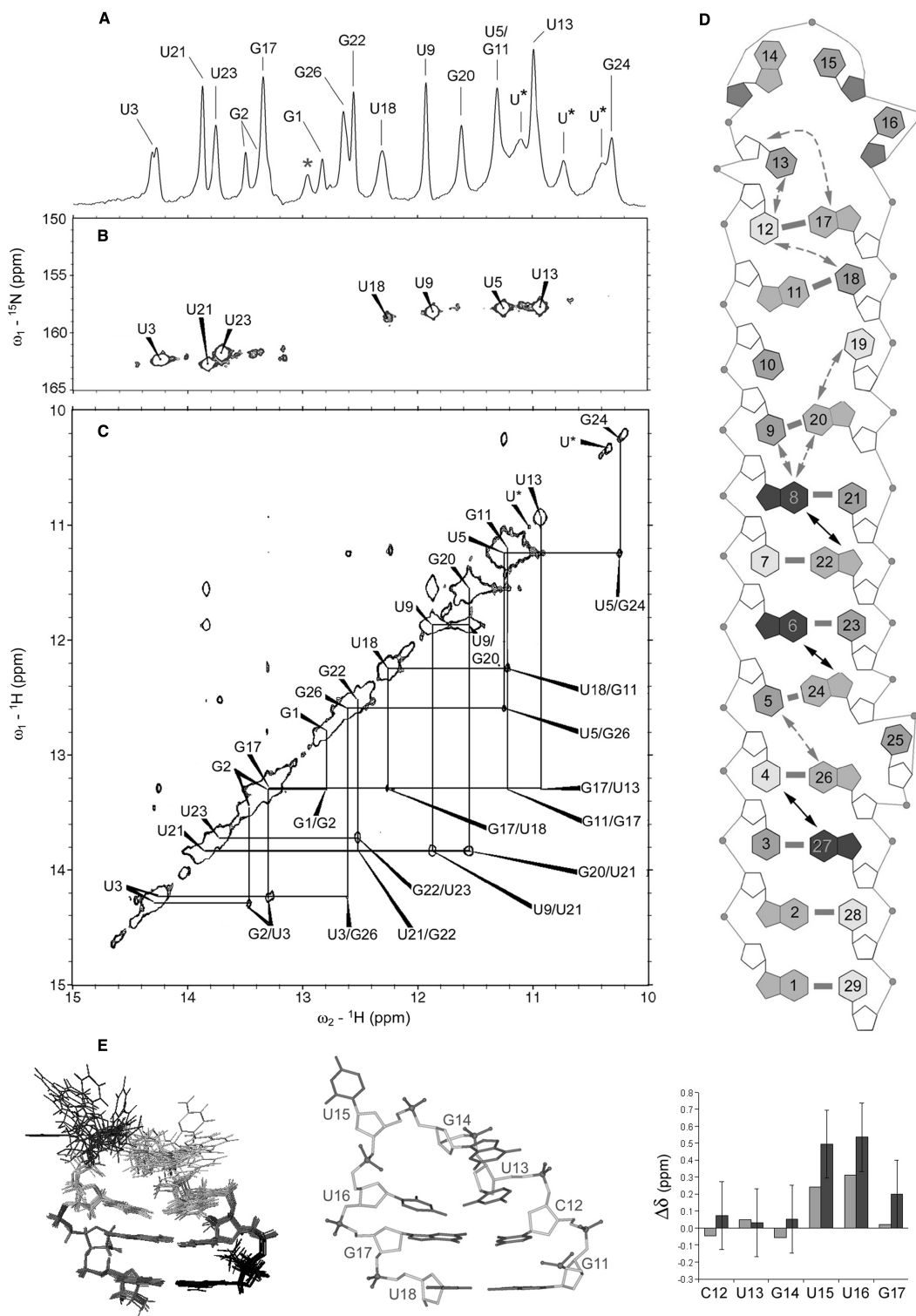
Figure 2. Solution-NMR investigation of the *duck* apical stem-loop structure (*ALD*). (**A**) ^{1}H-imino spectra (5°C, H_2O); assigned peaks are labeled by residue numbers, unassigned peaks are labeled with *. (**B**) [^{1}H,^{15}N]-HSQC of the selectively ^{15}N-U-labeled *ALD* (5°C, H_2O); (**C**) Imino region of *ALD* [^{1}H,^{1}H]-NOESY spectrum (15°C, H_2O); (**D**) Secondary structure model derived from NMR data: thick gray bars, base-pair; dark double arrows, NOEs involving non-exchangeable protons; dashed gray double arrows, NOEs involving exchangeable protons; pentagons, sugar rings (open, N-pucker, filled, S-pucker). (**E**) Structure model of UGUU tetraloop based on NMR classical restraints and after chemical shift refinement: Left: overlay of 10 lowest-energy structures; Middle: structure details, major groove view; Right: comparison of experimental $\Delta\delta$ (light gray) and back-calculated $\Delta\delta$ (dark gray). $\Delta\delta$ equals the mean of ($\delta_{calc\{H1', H5, H6, H8\}} - \delta_{helix\{H1', H5, H6, H8\}}$) or ($\delta_{exp\{H1', H5, H6, H8\}} - \delta_{helix\{H1', H5, H6, H8\}}$); the chemical shifts, δ_{calc}, were averaged over the 10 lowest-energy structures. Residues for which $\Delta\delta \approx 0$, experience ring current and thus stacking as in an A-helix; $\Delta\delta > 0$ and <0 indicate reduced and increased ring current and thus stacking, respectively, as compared to an A-helix environment. The error bars on the back-calculated $\Delta\delta$s correspond to $\pm 1.5\sigma$ of back-calculated and experimental chemical shifts of loop residues in a large data base of pdb-deposited RNA structures (25).
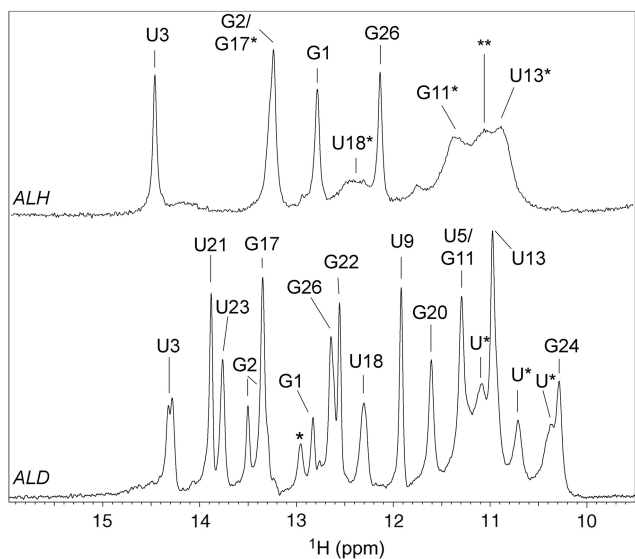
Figure 3. Comparison of the ^{1}H-imino NMR spectra (5°C, H_2O) of *heron-* (top) and *duck* (bottom) apical stem-loop. Assigned peaks are labeled by residue numbers and unassigned peaks by asterisks. Residue numbers in the upper part of *ALH* are superscripted with an asterisk for clarity.

This can be explained by occasional stacking of U_{15} not fully accounted for in the ensemble of calculated structures.

The U_{13}:U_{16} base pair in the UGUU tetraloop is slightly buckled and of the 'cis-wobble' type (26,27). It is comparable after initial calculation to that found in 1NA2, and in the final structure after CS refinement to that in 1MZP, [1NAZ and 1MZP accession codes in RCSB databank (28,29)]. The 'cis-wobble' conformation contains H-bonds between U_{13}-O2:U_{16}-NH3 and U_{13}-NH3:U_{16}-O4. Efficient H-bonding of U_{13}-NH3: U_{16}-O4 in the U_{13}:U_{16} pair together with stacking of G_{14} can account for the protection of the U_{13}-NH3 (see above, Figure 2A). In the derived structure model, the U_{16}-NH3, although H-bonded, is more solvent exposed due to the high flexibility and lack of stacking of U_{15}. This may explain why the U_{16}-NH3 resonance is not detected in the [^{1}H,^{1}H]-NOESY.

In conclusion, NMR experimental data show that the top part of the *avian* epsilon apical stem-loop folds into a well-defined UGUU tetraloop motif: characterized by a non-canonical closing U_{13}:U_{16} base-pair (type cis-wobble) onto which G_{14} stacks, while U_{15} is flexible, highly solvent exposed and easily accessible to intermolecular interactions. The initial and refined structures are deposited in the pdb-bank with accession codes, 2OJ7 and 2OJ8, respectively.

Human HBV epsilon apical stem-loop (ALHu). The (3D) structure of the *ALHu* stem-loop (Figure 1B) has been previously investigated with NMR by Flodell *et al.* (19,20,30). It consists of a stem formed by 10 bp with one bulged-out U and capped with a stable pseudo tri-loop also called a 'lone pair UGU tri-loop' (21).

Structural Investigation of the duck HBV primer loop (PLD). Combining data from [^{1}H,^{1}H]-NOESY (Figure 4C) and [^{1}H,^{15}N]-HSQC (Figure 4B) spectra allowed for unambiguous assignment of 13 of the 15 imino resonances observed in the 1D NMR spectrum (Figure 4A, and S3). From these data, we derived the secondary structure model of *PLD* (Figure 4D). Base pairs are formed in the *lower stem* up to A_6:U_{32}. The U_7:G_{31} wobble base pair at the top of the lower stem, closing the lower part of the primer loop, was not observed in the [^{1}H,^{1}H]-NOESY. It is thus open or in fast exchange with an open form. In the *upper stem* the G_{14}:C_{29} base pair is formed, thereby locking the primer loop in its upper region. The remaining unassigned imino signals in the 1D NMR spectrum, at ~11.8 and 11.2 p.p.m., can thus either be from U_7:G_{31} or by exclusion be assigned to Us from the primer loop region.

The primer loop region is well folded as follows from the extensive network of long-range NOE contacts: U_9-H6 ↔ G_{31}-H1′, U_{10}-H6 ↔ G_{31}-H1′, U_{10}-H6 ↔ A_{12}-H1′, U_{11}-H6 ↔ U_{30}-H1′ and A_{12}-H2 ↔ C_{29}-H1′ (Figure 4D). Furthermore, most primer loop residues experience A-helix-like stacking interactions, as evidenced by ring currents close to that in an A-helix; only A_{12} and C_{13} show reduced ring currents and thus reduced stacking (S4).

Thermodynamics

Human and avian epsilon HBV apical loops display significantly different thermodynamics. The thermodynamic properties of *ALD*, *ALH* and *ALHu* RNA structures were investigated by UV melting and analyzed following the practical recommendations of Puglisi and Tinoco (31,32) as described in more detail in the Materials and Methods section. The raw experimental melting curves are presented in Figure S5. The first derivatives of the fitted experimental curves are shown in Figure 5, their melting temperatures and associated thermodynamic parameters are compiled in Table 1. All thermal denaturation curves were analyzed assuming an all-or-none (or two-state) model (31,32) to derive the thermodynamic parameters. This approximation is appropriate for melting of non-interacting short helices as investigated here (31,33). The absence of intermolecular interactions was assessed for each individual RNA sequence by recording melting curves at high and low RNA concentrations (31). Because the sequences are all single stranded, the melting is mono-molecular and must be independent of concentration. In addition, the effect of salt (monovalent, Na^+ as well as divalent, Mg^{2+}) was investigated. As no significant effect was observed when using Mg^{2+}, we do not present these data, except in the case of *ALHu* below. The analysis was completed by comparison of experimentally derived values of free energy changes with those computed from NMR-derived secondary structure and using the thermodynamic parameter tables published by Serra and Turner (34) (Table 1).

ALD sequence is less stable than expected on the basis of the NMR model. Thermal denaturation of *ALD* in 1.0 M NaCl occurs via a single transition with a melting
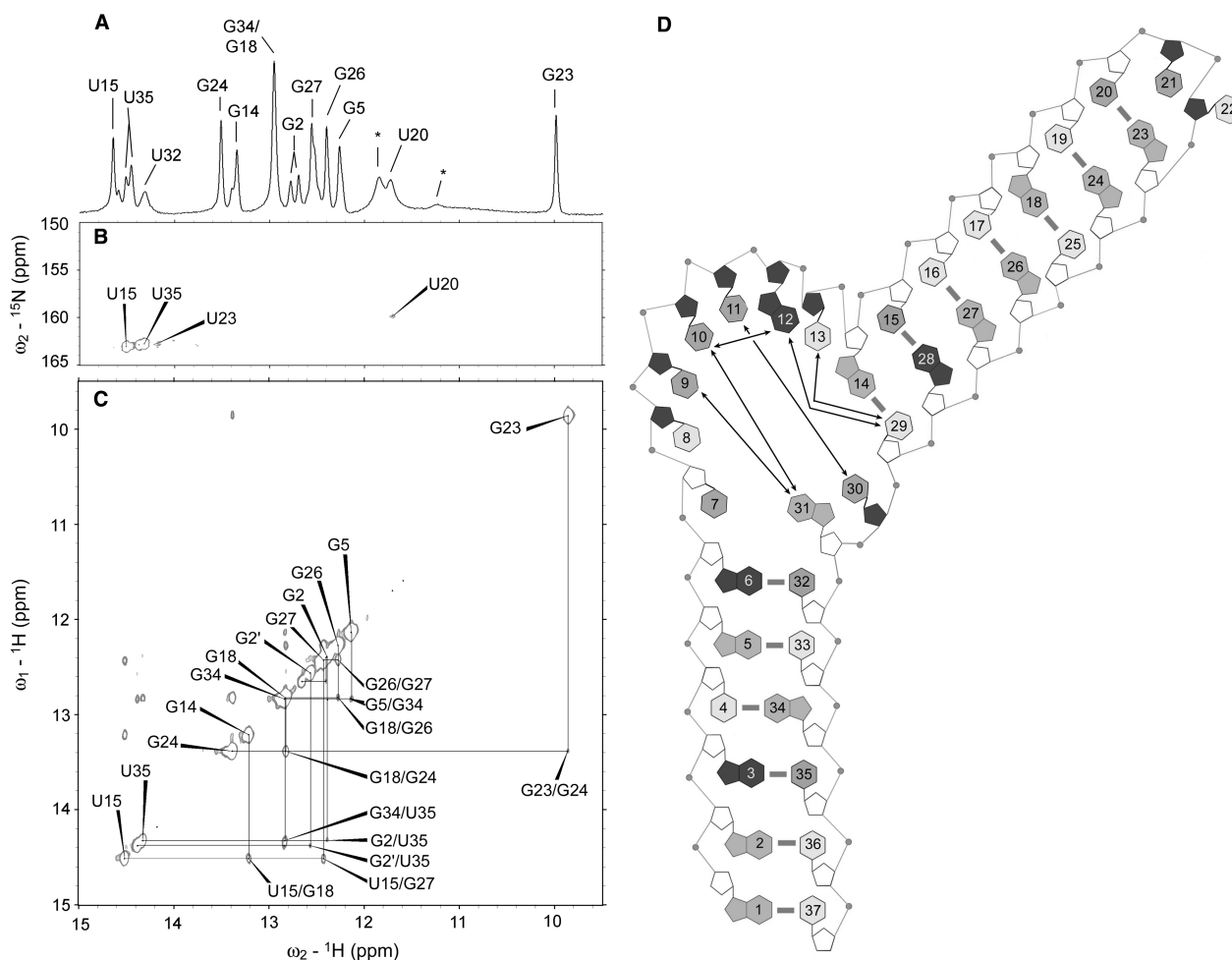
Figure 4. The *duck* primer loop secondary structure (*PLD*) from NMR data. (**A**) [1]H-imino spectra (5°C, H_2O); (**B**) [^{1}H,^{15}N]-HSQC of the selectively ^{15}N-U labeled *PLD* (5°C, H_2O); (**C**) Imino region of *PLD* [^{1}H,^{1}H]-NOESY spectrum (15°C, H_2O); (**D**) Secondary structure model derived from experimental NMR data. The same coding is used as in Figure 2.

temperature (T_m) of 64.9°C (Figure 5, Table 1, S5). The transition is mono-molecular because RNA concentration does not affect T_m (Table 1).

The standard Gibbs free energy change (ΔG°_{exp}) of the transition is $-5.3 \, kcal \cdot mol^{-1}$ as follows from a van't Hoff plot (Materials and Methods section). However, both full-length (ALD_n) and the $n-1$ abortive (ALD_{n-1}) products were combined into one sample to compensate for the low *in vitro* expression. The absence of C_{29} in the ALD_{n-1} sequence increases ΔG° with $2.1 \, kcal \cdot mol^{-1}$ as follows from Turner's tables (34). The relative concentrations of ALD_n and ALD_{n-1} were estimated to be equal, based on denaturing PAGE and NMR analysis. The presence of this 50% ALD_{n-1} requires a correction of $-1.05 \, kcal \cdot mol^{-1}$, leading to a final ΔG°_{exp} of $-6.3 \, kcal \cdot mol^{-1}$ for full-length *ALD*.

Using Turner's tables and the NMR-derived secondary structure model, we computed a predicted ΔG°_{calc} of $-9.3 \, kcal \cdot mol^{-1}$. Thus *ALD* is predicted to be $2.0-3.0 \, kcal \cdot mol^{-1}$ more stable than what is experimentally observed. This is a significant difference as ΔG°_{calc} can typically be predicted with an error margin of 10% (34).

The $2.0-3.0 \, kcal \cdot mol^{-1}$ difference in stability corresponds to a loss of 2–3 Watson–Crick bp, or alternatively to the presence of an additional 2-nt bulge in the sequence (34). Hence we must conclude that one or more unidentified elements within the ALD structure are less stable than expected; our NMR melting studies (see below) demonstrate that it is the middle part of the ALD stem that is relatively unstable.

ALH sequence: thermodynamics fully supports the NMR model. Thermal denaturation of *ALH* occurs via a broad transition with a T_m of 63.1°C in standard 1.0 M NaCl conditions (Figure 5). The absence of an RNA concentration effect on the T_m indicates that the transition is mono-molecular (Table 1). We derived in standard conditions a ΔG° of $-2.5 \, kcal \cdot mol^{-1}$. This value matches perfectly with the one computed using Turner's tables and the NMR-derived secondary structure model ($-2.4 \, kcal \cdot mol^{-1}$, Table 1).

Strikingly, even though *ALH* is clearly less structured than *ALD*, their experimentally derived melting temperatures are in the same range: 63.1°C for *ALH* and 64.9°C

for *ALD* (Table 1). They nevertheless display a significant difference in free energy change (~ 4 kcal·mol^{-1}). The width of the transition's first derivative (Figure 5) is much larger for *ALH* than for *ALD*, indicating a smaller ΔH° for *ALH* (35). This smaller ΔH° must be compensated by a smaller ΔS°, because $\Delta G^\circ = 0$ at the melting temperature and both sequences have approximately the same T_m values. Both sequences have the same length and thus a similar entropy in the denatured form. Hence the initial entropy of *ALH* must be larger than that of *ALD*. This is consistent with the NMR secondary structure models, which show that *ALH* has a large floppy internal loop and *ALD* does not have (see above).

ALHu sequence: highly thermodynamically stable. The thermal denaturation of *ALHu* shows a narrow transition with a T_m of 79.5°C (with 0.15 M NaCl). The absence of an RNA concentration effect on the T_m indicates that the transition is mono-molecular.
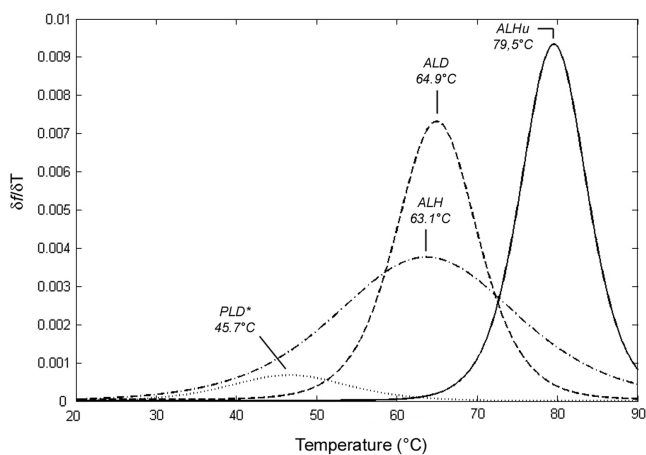

Figure 5. UV melting transition recorded for *ALD*, *ALH*, *ALHu*, and *PLD*. On the figure are displayed the first derivatives of the fitted experimental melting curves as described in the Materials and Methods section; original experimental data are presented in S5.

The $\Delta G^\circ_\mathrm{exp}$ is -11.8 kcal·mol^{-1} (0.15 M NaCl, Table 1). Thermodynamic parameters from a melting curve under standard conditions could not be determined (i.e. 1.0 M NaCl), as the sequence was so stable that the base line in the high-temperature domain could not be defined with reasonable confidence. The $\Delta G^\circ_\mathrm{exp}$ is in good agreement with $\Delta G^\circ_\mathrm{calc}$, which is based on the NMR-derived secondary structure (-12.2 kcal·mol^{-1}).

NMR melting analysis shows *ALD* structural instability

To identify the residues involved in the thermal melting, the process was followed by solution NMR. In a nucleic acid base pair, the iminos (from G and U residues) are hydrogen bonded and protected from exchange with water in this closed state. Exchange can only take place in the open state (36). Raising the temperature shifts the equilibrium towards the open state. Proton exchange between the imino proton and the solvent, broadens, but does not shift, the imino proton resonance. In contrast, the alteration between closed and open states produces a shift, but usually no appreciable broadening (37). Usually the line broadening is much more pronounced than the change in chemical shift. Therefore, broadening of the imino resonances (and small changes in their chemical shifts) directly proves a shift towards the open state (36).

The ^{1}H-imino NMR spectra of *ALD* were followed from 5 to 40°C in steps of 5°C (Materials and Methods section, Figure 6A). At 40°C, i.e. below the T_m observed by UV melting (see above), the imino resonances assigned to G_2-NH1, U_3-NH3 and G_{26}-NH1 are still intense and narrow. Also, the peaks assigned to G_{11}-NH1, G_{17}-NH1 and U_{18}-NH3, although broadened, are still observable, even at 40°C. This indicates that both the top loop and the basal stem regions are relatively stable. On the other hand, the U_{21}-NH3, G_{22}-NH1 and U_{23}-NH3 resonances, in the middle region already broaden and vanish between 30 and 35°C. Thus, these central base pairs open before the top loop and basal stem region. In conclusion, a relatively unstable motif is located in the middle region of the

Table 1. Thermodynamic parameters of the apical loop sequences derived for different Na$^+$- and RNA concentrations

Sample	[RNA] µmol·l^{-1}	[Na$^+$] mol·l^{-1}	T_m °C	ΔH° kcal·mol^{-1}	ΔS° cal·mol^{-1}·K^{-1}	ΔG° kcal·mol^{-1}
ALD	2.4	1.0	64.9	−64.2	−190.0	−5.3/−6.3[a]
			67.7*	−95.4*	−279.9*	−8.6*
						−9.3**
	2.6	0.15	56.8	−63.7	−193.2	−3.8
	0.9	0.15	57.6	−63.0	−190.6	−3.9
ALH	1.7	1.0	63.1	−31.5	−93.6	−2.5
			63.5*	−53.4*	−158.6*	−4.2*
						−2.4**
	1.8	0.15	57.9	−37.0	−111.7	−2.3
	0.6	0.15	57.7	−27.2	−81.8	−1.9
ALHu	n.d.	1.0	n.d.	n.d.	n.d.	n.d.
			80.2*	−96.6*	−273.4*	−11.8*
						−12.2**
	2.2	0.15	79.5	−92.8	−263.2	−11.1
	0.9	0.15	80.0	−96.6	−273.5	−11.8

The values followed by * and ** are computed values derived with *Mfold* (42) and with tables by Serra and Turner (34), respectively. [a]Value corrected for dangling end-effect in *ALD* (see text). For *ALH*, the predicted ΔG° computed with *Mfold* is found to be slightly lower ($\Delta G^\circ - 4.2$ kcal·mol^{-1}) than using Serra and Turner's tables, because *Mfold* considers the internal loop region in the structure less destabilizing than the experimental tables by Serra and Turner.
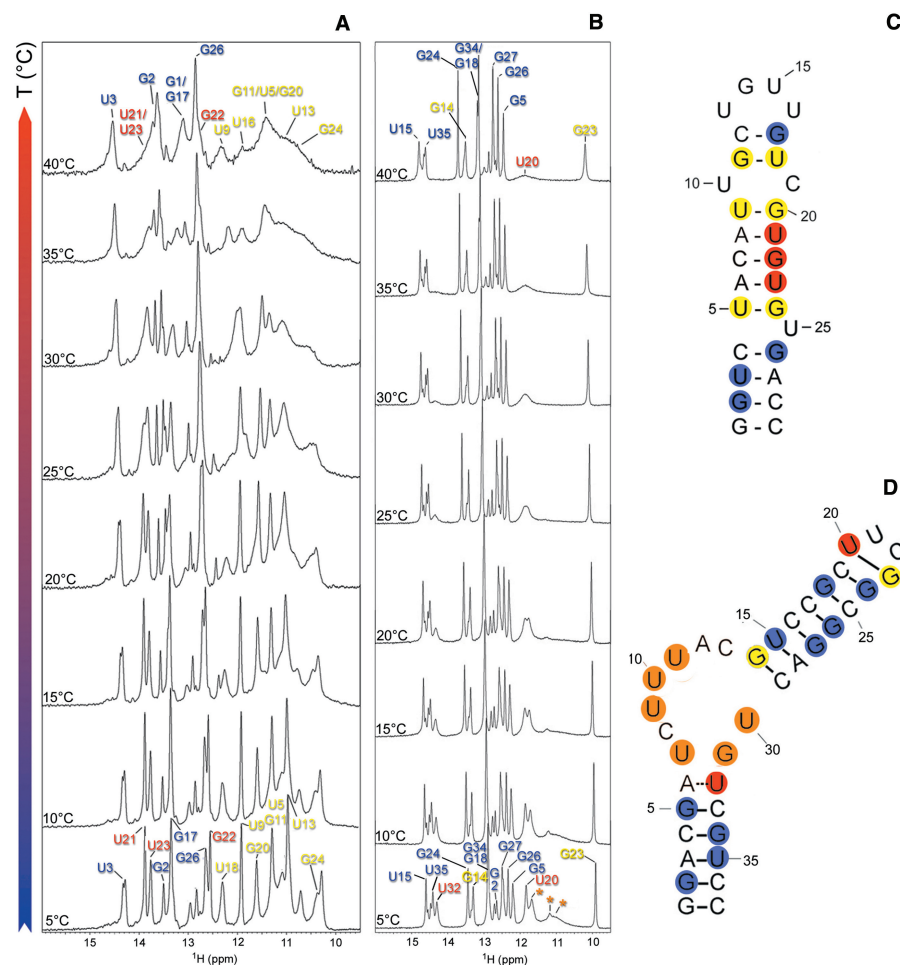
Figure 6. NMR melting study. (**A**) *ALD* ^{1}H-imino spectra; (**B**) *PLD* ^{1}H-imino spectra; (**C**) Structural weakness of *ALD* is highlighted using the following color coding: *blue*, stable base pair, imino peak visible at 40°C; *yellow*, base pair affected by enhanced exchange at 40°C, imino broadened; *red*, base pair highly affected by exchange at 40°C, imino peak vanished. The central region (red) 'melts' between 30 and 35°C. (**D**) Structural weakness of *PLD* is highlighted using the same color coding as in C, except for: *orange*, residues from the primer loop region, non-specifically assigned, affected by exchange in the low temperature range. Imino resonances in A and B are indicated by residue type and number with color code corresponding to that used in C and D.

sequence, lowering the apparent overall stability as derived from UV melting.

Thermodynamic properties of *PLD*

The thermal denaturation of *PLD* is characterized by two well-resolved transitions. The first (T_m 45.7°C, $\Delta G°$ of -1.38 kcal·mol^{-1}, Table 1) gives rise to an increase of only $\sim$1% in UV absorbance (hyperchromicity) (Figure 5). For comparison, the complete unfolding of a double-stranded RNA typically causes an increase in overall UV absorbance of 15–20% (31). The first transition is thus related to a structural element where only partial unpacking of nucleic bases occurs, e.g. the primer loop. The second transition is of much larger amplitude, $\sim$15% hyperchromicity, and centered at $\sim$85°C (S5). Although it is well separated from the first, its high T_m does not allow us to obtain a well-defined base line in the high temperature domain. Consequently, its T_m and associated thermodynamic parameters cannot be

accurately derived. The upper stem of *PLD* is capped with a highly stable UUCG tetraloop [$T_m > 70$°C, (38)] to facilitate NMR assignment (Figure 1B). Because of its high T_m and its relatively large hyperchromicity, the second transition most likely includes as RNA elements, the UUCG loop and the stems.

To unambiguously identify the residues involved in the first and second melting transitions, the melting process was followed by ^{1}H imino NMR (Figure 6B). At 40°C, all broad resonances between 11.7 and 10.5 p.p.m. (orange in Figure 6B) as well as the U_{32} ^{1}H-imino resonance at 14.3 p.p.m. (red in Figure 6B) have completely vanished. All other ^{1}H-imino resonances remain sharp and intense. The ^{1}H-imino of U_{20}, engaged in the closing UG base pair of the UUCG tetraloop, is also broadened, but remains detectable (red in Figure 6B). In summary, we observe that the resonances that vanish at 40°C are within/or close to the epsilon primer loop. The first transition can thus unambiguously be assigned to a conformational rearrangement affecting the primer loop residues. The sharpness

of the remaining imino resonances demonstrates the existence of well-formed upper and lower stems above the transition (45.7°C).

DISCUSSION

Although knowledge of the molecular basis for P–ε binding and priming is emerging from biochemical probing methods (10,15,18,39), several questions remain concerning the exact structure or sequence determinants of ε that trigger the P–ε interaction. A full understanding of the molecular basis for P–ε binding and priming awaits high-resolution structural and thermodynamic data. Our NMR results and thermodynamic data define and compare the structure and stability characteristics of the *AL* (and *PL*) segments of ε RNAs from *human*, *duck* and *heron* HBV. They are discussed below in terms of structure motifs important for P–ε interaction.

The apical stem-loop of epsilon in avian HBV is capped by a well-structured UGUU tetraloop

We first discuss the structure of the top of the *avian* ε apical stem-loop with sequence: $GCUGU_{2590}UGU$ [for ease of comparison definition and numbering of Hu *et al.* Figure 1A (16)]. Our NMR data and structure calculations show that in the *duck* ε apical stem-loop the above sequence folds into a well-defined tetraloop structure (Figure 2E). It contains a non-canonical U_{2588}:U_{2591} base pair (type cis-wobble) stacked onto the closing base pair C_{2587}:G_{2592}, while residue G_{2589} is in turn stacked onto U_{2588} on the major groove side. Most interesting is the flexible U_{2590}, which is solvent exposed and thus available for intermolecular interaction. The UGUU tetraloop is new as far as we are aware; for instance, no UGUU tetraloop is observed in the ribosome.

Although the structure calculations were carried out on the *duck* sequence, our NMR data also show the presence of G_{2592} and G_{2586} imino resonances in the *heron* apical stem-loop and with similar chemical shifts as in *duck* (Figure 3). This unequivocally demonstrates the formation of base pairs C_{2587}:G_{2592} and G_{2586}:U_{2593} with similar conformation in *duck* and *heron* strains. Hence, the sequence UGUU, closed by the C_{2587}:G_{2592}/G_{2586}:U_{2593} base pairs, must take on this UGUU tetraloop fold (albeit possibly with some dynamics in the loop itself) not only in *duck* but also in *heron*. The formation of a tetraloop in *heron* was unexpected because the middle part of the apical stem has no complementary base pairs, predicting it to be non- or loosely folded.

The following conclusions can now be drawn: (a) Given the high conservation of the $GCUGU_{2590}UGU$ sequence among all *avian* HBV strains (16), one can reasonably conclude that the prototypic *avian* ε signal contains a well-defined tetraloop structure with two underlying base pairs at the top of the apical stem-loop (Figures 2E and 7). In terms of P–ε interaction, we can further conclude and propose the following: (b) Given the observation that in all SELEX-generated binders one or two closing base pairs are seen (10,16), one can reasonably conclude that efficient *avian* P–ε binding requires formation of a

tetraloop closed by one or two underlying base pairs. (c) Given the observation of a tetraloop sequence in all SELEX-generated binders (16), we propose that this tetraloop fold is required and acts as a recognition determinant for the RT-domain of protein P, although alternatively interaction(s) with chaperone proteins cannot be excluded. (d) Given that mutation $U_{2590}C$ causes complete loss of binding, without likely affecting the tetraloop structure, we propose that the solvent-exposed U_{2590} interacts with the RT-domain of protein P upon initial binding (Figure 7).

An unstable region is present in the middle of all avian epsilon apical stem-loops: this recurrent functional motif is required for P-binding

The second motif involves the middle part of the apical stem. From their pool of SELEX-generated RNA, Hu *et al.* (16) found that mutants, which bind P with high affinity, present poor base pairing within the middle stem region. This finding is substantiated by mutagenesis experiments, which introduced more efficient base pairing
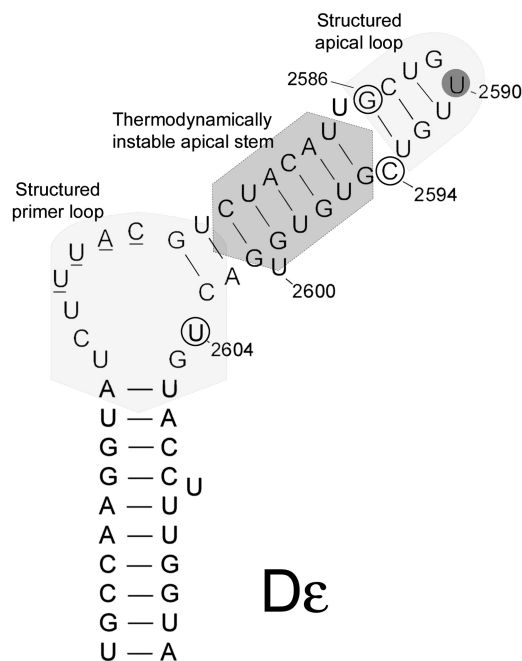

Figure 7. Schematic of the functional prototypic *avian* HBV ε signal. It summarizes our structural and thermodynamic data and some pertinent mutation/probing and sequence data from Nassal cs.(16), who proposed as the functional prototypic ε signal an open structure of the *AL* as in Hε. Combined with our data this model can now be refined. The functional prototypic *avian* ε signal appears to be: an *AL*, unstructured or weakly structured in the middle stem (gray) and capped with a structured UGUU tetraloop closed by two underlying base pairs; a structured pimer loop (T_m 47°C, light gray) closed at its tip by a stable C:G base pair, whereas the U:G closing base pair at its bottom is absent or not stably formed. U_{2590} (dark gray circle) in the conserved UGUU tetraloop is a key recognition element in initial binding (see text). Binding is followed by structural rearrangements to obtain a priming competent complex (10,16). This should involve residues G_{2586}, C_{2594} and U_{2604} (open circle), as they are not P accessible upon initial binding and have been proposed (16) to be part of a network of anchor residues, enabling precise positioning of P over the initiation site (see text).

in the *duck* apical stem. According to the authors, this absence of base pairing, and thus instability or flexibility, is a general characteristic of all *avian* strains except the *duck* strain. The existence of relatively extensive base pairing in the *duck* sequence appeared to be an exception to the rule among *avian* species (16). However, our NMR data show that, although the *ALD* structure is well defined and structured at low temperature (<ca. 30°C), it does contain a weak middle stem region, which undergoes a 'melting' transition between 30 and 35°C, i.e. the imino resonances of this region broaden to disappearance in this temperature range in contrast to those of the upper and lower stems (Figure 6A and C). Thus, our data show that the ε-apical stem-loop of DHBV is not an exception but has characteristics similar to that of all *avian* strains (Figure 7).

Initial binding is followed by structural changes, which lead to a priming-competent complex

A priming-competent complex requires initial binding to be followed by structural rearrangements (10,16). The DHBV ε–P complex was investigated by chemical probing in an isolated state obtained just after a 2/4-nt primer was synthesized (10). In this state, the stem of the ε-apical stem-loop is melted and interacts with the RT-domain of the P-complex. Due to this interaction, many residues in the apical stem are protected, but its middle region, identified by NMR melting as a structurally weak point, contains a central $GU_{2598}G$ hotspot, more reactive to Pb^{2+} attack, showing it is solvent exposed (10) (Figure 7). The SELEX data (16) show that three residues, G_{2586}, C_{2594} and U_{2604}, are of special interest as their mutation abrogates priming but not binding and they, except for G_{2586}, are part of the P footprint (Figure 7). This suggests that these residues represent anchoring points directing or relaying, after initial binding, the necessary structural rearrangements leading up to the priming-competent complex as proposed by Hu *et al.* (16). Our structural data on *AL* and *PL* show that these residues are not accessible to P interactions upon initial binding. Initial binding must therefore be followed by structural rearrangements to obtain a priming-competent complex. Interestingly, these residues are located at the border of or just outside the more stable stem regions. These rearrangements could be triggered by higher dynamics of these residues, which would populate the low-populated states where these residues are P accessible. We are currently investigating their dynamics.

The free energy ($\Delta G°$) cost for complete opening of the primer loop and apical stem-loop is ~13, ~8 and ~4 kcal·mol^{-1} for the *human*, *duck* and *heron* sequences, respectively (Table 1). These values define an upper bound on the free energy required for priming considering that partial conformational rearrangement(s) may as well suffice. The ~8 kcal·mol^{-1} for *duck* appears to be the maximum that can be accommodated by the *duck* P protein complex, since further stabilization of the middle part of the apical stem abrogates binding. Why *human* ε RNA is not binding competent in the DHBV *in vitro* system could be due to its high stability. The study by Hu

and Boyer (18) on a recently developed *in vitro human* HBV system appears to indicate that in *human* HBV, the apical loop is not needed for binding of ε to P, but instead for encapsidation. This interaction with core proteins may in *human* ε RNA be needed to enforce the structural rearrangements required for priming.

CONCLUSIONS

In this article, the structures and stabilities of the apical stem-loops (*AL*) of *human*, *duck* and *heron* HBV epsilon RNA elements have been investigated. UV melting shows that the *avian AL*, *duck* and *heron*, are much less stable than *human AL*. NMR data show that this low stability stems from the unstructured 11-nt internal bulge in the middle of the stem of *heron AL*. In *duck*, this region, although structured at low temperature, also forms a structural weak point, as evidenced by broadening to disappearance of their imino resonances between 30 and 35°C, well below the overall melting temperature of ca. 60°C. The *duck and heron ALs* are both capped by a stable UGUU tetraloop closed by two underlying base pairs. All *avian AL* are expected to adhere to this because of their sequence conservation. The seven-residue primer-loop of ε *duck* HBV is structured, having a melting temperature of 45.7°C and a Gibbs free energy change equal to −1.4 kcal·mol^{-1}. In view of sequence similarities the same is expected for *heron*- and *human PL*. The present structure and thermodynamic data combined with mutation data thus refine the model for the functional prototypical *avian* ε signal (Figure 7).

MATERIALS AND METHODS

RNA sample preparation

RNA sequences *ALD*, *ALH*, *ALHu* and *PLD* (Figure 1B) were synthesized by standard *in vitro* transcription using a single-stranded DNA template and T7 RNA polymerase (40). Both full-length (ALD_n) and the $n-1$ abortive (ALD_{n-1}) products were combined into one sample to compensate for the low *in vitro* expression. A selectively U-labeled *ALD* sample was also prepared following the same procedure, using $^{13}C/^{15}N/^2H_{1', 3', 4', 5', 5''}$-UTP. It was prepared according to Cromsigt *et al.* (41). In addition, segments of the *AL* as well as a *PLD-U30* mutant were synthesized to support and confirm NMR resonance assignments (S6). *PLD* and *PLD-U30* were each expressed in 3 different forms, for instance non-isotopically labeled, $^{13}C/^{15}N/^2H_{1', 3', 4', 5', 5''}$-Uracil and $^2H_{1', 3', 4', 5', 5''}$-Uracil labeled.

UV melting experiments

RNA samples were re-suspended in the appropriate T_m buffer (50 mM Tris-HCl pH 7, 150 mM NaCl). Prior to the T_m experiment, samples were snap-cooled after 1 min denaturation at 90°C. For each sample (*ALD*, *ALH*, *ALHu* and *PLD*), experiments were carried out at high and low RNA concentrations (Table 1). The thermal denaturation/renaturation curves were recorded at 260 nm over a temperature range of 20–90°C using a fixed rate of

$0.5°C \cdot min^{-1}$ on a Varian Cary 300-Bio UV-Vis spectrometer. The melting curves were analyzed following closely the recommendations described by Puglisi and Tinoco (31). Briefly, the experimental melting curves were baseline corrected by fitting a linear function to its pre-transition part and subsequently subtracting the slope over the whole temperature range. The baseline-corrected experimental curves were normalized and subsequently fitted in MATLAB with a general sigmoidal function as shown in Equation (1), which includes baseline (x_4) and offset (x_5),

$$y = \frac{x_1}{1 + e^{-x_2(x-x_3)}} + x_4 x + x_5 \qquad 1$$

The normalization constant (x_1) should now be 1, while baseline (x_4) and offset (x_5) terms should be zero. If not, the pre-transition part should be adjusted and the baseline correction and normalization iterated. In this way a careful baseline correction at low and high temperature of the experimental melting curves was achieved. Melting temperatures were derived from the maximum of the fitted curves' first derivative. Thermodynamics parameters ($\Delta H°$, $\Delta S°$ and $\Delta G°$) were subsequently derived from the van't Hoff plot of the normalized and baseline-corrected melting curves, again all according to Puglisi and Tinoco's recommendations (31). Thermodynamic parameters of each RNA sequence were also computed using either Turner's tables (34) and the NMR-derived secondary structure and/or secondary structure prediction using Zuker's algorithm (42,43).

NMR spectroscopy

All NMR samples were prepared in the appropriate NMR buffer (10 mM Na-Phosphate pH 6.7, 0.1 mM EDTA), to a final concentration of 1.2, 1.0, 0.4, 0.8, 0.8 and 0.8 mM for *ALD*, $^{13}C/^{15}N/^2H$-U labeled *ALD*, *ALH*, *PLD*, $^{13}C/^{15}N/^2H$-U labeled *PLD* and ^2H-U labeled *PLD*, respectively. Spectra in water (95:5, $H_2O:D_2O$) were recorded at 5 and 15°C on a 500 MHz Varian Inova spectrometer. Water suppression was achieved with a Watergate (44) sequence combined with water flip-back pulse (WFB) based on a SINC-type-shaped pulse. The 2D [^1H,^1H]-NOESY spectra in water, used for resonance assignment (23), were recorded with 100 and 200 ms mixing time, collecting 2048×700 complex points for a spectral width of 14 kHz in both dimensions. For the NMR melting experiments, 1D-WFB/watergate ^1H spectra were recorded every 5°C from 5 to 40°C, collecting 128 scans and using 2000 complex points for a spectral width of 14 kHz. For the U-labeled *ADL* sequence, [^1H,^{15}N]-HMQC-WFB was recorded on a 600 MHz Varian Inova spectrometer, collecting 2048×400 points for a spectral width of 15 and 2.5 kHz, respectively. In D_2O, all spectra were recorded on an 800 MHz Varian Inova spectrometer. A set of 2D [^1H,^1H]-NOESY experiments was carried out at 15 and 25°C using 30, 120, 300 and 500 ms mixing time. For each spectrum, 2048×800 complex points were acquired over a spectral width of 8 kHz in both direct and indirect dimensions. [^1H,^1H]-DQF-COSY experiments, acquired in a constant time fashion using a refocusing delay of 40 ms, were recorded with the same

respective spectral width and resolution. Sugar puckering modes were derived from observable/non-observable H1′-H2′ coherences in the COSY spectra. [^1H,^{13}C]-HSQC and ^{13}C-decoupled [^1H,^1H]-NOESY spectra of the U-labeled *ALD* sample were recorded for assignment purposes (23,24,41). Finally, 1D ^{31}P and [^1H,^{31}P]-HMBC spectra were recorded to derive information on phosphate backbone conformation (23). The spectra were processed with NMRpipe (45) prior to resonance assignment using Sparky software (46).

Structure calculations

Structure calculations on $G_{11}CUGUUGU_{18}$ of *ALD* were carried out with X-PLOR 3.851 using a torsion angle dynamics protocol (47,48). Starting from an extended structure, 800 structures were generated using classical NMR restraints (Table S7). The 10 lowest-energy structures (Table S7) were selected for further refinement employing chemical shifts as additional restraints as described in the legend of Table S7, generating 10 refined structures out of each selected structure. From the resulting 100 refined structures, the 10 with lowest energy were selected and their ^1H chemical shifts analyzed using Nuchemics (25,49). The calculated ^1H shifts were averaged among the ensemble to account for motional averaging (50) and subsequently for each nucleotide $\triangle\delta$ values were computed as described in the legend of Figure 2E.

SUPPLEMENTARY DATA

Supplementary data is available at NAR Online.

REFERENCES

1. Buendia,M.A. (1992) Hepatitis B viruses and hepatocellular carcinoma. *Adv. Cancer Res.*, **59**, 167–226.
2. Lee,W.M. (1997) Medical progress – hepatitis B virus infection. *New Engl. J. Med.*, **337**, 1733–1745.
3. Ganem,D. and Schneider,R. (2001) *Fields Virology*. Lippincott, Williams & Wilkins, Philadelphia, PA, USA.
4. Nassal,M. (2000) *DNA Virus Replication*. Oxford University Press, Oxford.
5. Nassal,M. (1999) Hepatitis B virus replication: novel roles for virus-host interactions. *Intervirology*, 100–116.
6. Ganem,D. and Varmus,H.E. (1987) The molecular-biology of the hepatitis-B viruses. *Annu. Rev. Biochem.*, **56**, 651–693.

7. Nassal,M. and Schaller,H. (1996) Hepatitis B virus replication – an update. *J. Viral Hepat.*, **3**, 217–226.

8. Beck,J., Bartos,H. and Nassal,M. (1997) Experimental confirmation of a hepatitis B virus (HBV) epsilon-like bulge-and-loop structure in avian HBV RNA encapsidation signals. *Virology*, **227**, 500–504.

9. Tavis,J.E., Perri,S. and Ganem,D. (1994) Hepadnavirus reverse transcription initiates within the stem-loop of the RNA packaging signal and employs a novel strand transfer. *J. Virol.*, **68**, 3536–3543.

10. Beck,J. and Nassal,M. (1998) Formation of a functional hepatitis B virus replication initiation complex involves a major structural alteration in the RNA template. *Mol. Cell. Biol.*, **18**, 6265–6272.

11. Wang,G.H. and Seeger,C. (1993) Novel mechanism for reverse transcription in hepatitis B viruses. *J. Virol.*, **67**, 6507–6512.

12. Beck,J. and Nassal,M. (1997) Sequence- and structure-specific determinants in the interaction between the RNA encapsidation signal and reverse transcriptase of avian hepatitis B viruses. *J. Virol.*, **71**, 4971–4980.

13. Beck,J. and Nassal,M. (2001) Reconstitution of a functional duck hepatitis B virus replication initiation complex from separate reverse transcriptase domains expressed in Escherichia coli. *J. Virol.*, **75**, 7410–7419.

14. Beck,J. and Nassal,M. (2003) Efficient hsp90-independent *in vitro* activation by Hsc70 and Hsp40 of duck hepatitis B virus reverse transcriptase, an assumed Hsp90 client protein. *J. Biol. Chem.*, **278**, 36128–36138.

15. Hu,J.M. and Anselmo,D. (2000) *In vitro* reconstitution of a functional duck hepatitis B virus reverse transcriptase: posttranslational activation by Hsp90. *J. Virol.*, **74**, 11447–11455.

16. Hu,K.H., Beck,J. and Nassal,M. (2004) SELEX-derived aptamers of the duck hepatitis B virus RNA encapsidation signal distinguish critical and non-critical residues for productive initiation of reverse transcription. *Nucleic Acids Res.*, **32**, 4377–4389.

17. Seeger,C., Leber,E.H., Wiens,L.K. and Hu,J.M. (1996) Mutagenesis of a hepatitis B virus reverse transcriptase yields temperature-sensitive virus. *Virology*, **222**, 430–439.

18. Hu,J., Flores,D., Toft,D., Wang,X. and Nguyen,D. (2004) Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. *J. Virol.*, **78**, 13122–13131.

19. Flodell,S., Petersen,M., Girard,F., Zdunek,J., Kidd-Ljunggren,K., Schleucher,J. and Wijmenga,S. (2006) Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res.*, **34**, 4449–4457.

20. Flodell,S., Schleucher,J., Cromsigt,J., Ippel,H., Kidd-Ljunggren,K. and Wijmenga,S. (2002) The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.*, **30**, 4803–4811.

21. Lee,J.C., Cannone,J.J. and Gutell,R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.*, **325**, 65–83.

22. Haasnoot,P.C.J., Olsthoorn,R.C.L. and Bol,J.F. (2002) The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop C. *RNA*, **8**, 110–122.

23. Wijmenga,S.S. and van Buuren,B.N.M. (1998) The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**, 287–387.

24. Cromsigt,J., van Buuren,B., Schleucher,J. and Wijmenga,S. (2001) Resonance assignment and structure determination for RNA. *Methods Enzymol.*, **338**, 371–399.

25. Cromsigt,J.A.M.T.C., Hilbers,C.W. and Wijmenga,S.S. (2001) Prediction of proton chemical shifts in RNA – their use in structure refinement and validation. *J. Biomol., NMR*, **21**, 11–29.

26. Leontis,N.B. and Westhof,E. (1998) Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.*, **31**, 399–455.

27. Lemieux,S. and Major,F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.*, **30**, 4250–4263.

28. Theimer,C.A., Finger,L.D., Trantirek,L. and Feigon,J. (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc. Natl. Acad. Sci. USA*, **100**, 449–454.

29. Nikulin,A., Eliseikina,I., Tishchenko,S., Nevskaya,N., Davydova,N., Platonova,O., Piendl,W., Selmer,M., Liljas,A. *et al.* (2003) Structure of the L1 protuberance in the ribosome. *Nat. Struct. Biol.*, **10**, 104–108.

30. Flodell,S., Cromsigt,J., Schleucher,J., Kidd-Ljunggren,K. and Wijmenga,S. (2002) Structure elucidation of the hepatitis B virus encapsidation signal by NMR on selectively labeled RNAs. *J. Biomol. Struct. Dyn.*, **19**, 627–636.

31. Puglisi,J.D. and Tinoco,I. (1989) Absorbency melting curves of RNA. *Methods.Enzymol.*, **180**, 304–325.

32. Applequist,J. and Damle,V. (1965) Thermodynamics of helix-coil equilibrium in oligoadenylic acid from hypochromicity studies. *J. Am. Chem. Soc.*, **87**, 1450–1455.

33. Breslauer,K.J., Sturtevant,J.M. and Tinoco,I. (1975) Calorimetric and spectroscopic investigation of helix-to-coil transition of a ribo-oligonucleotide – Ra7u7. *J. Mol. Biol.*, **99**, 549–555.

34. Serra,M.J. and Turner,D.H. (1995) Predicting thermodynamic properties of RNA. *Energ. Biol. Macromol.*, **259**, 242–261.

35. Marky,L.A., Kallenbach,N.R., Mcdonough,K.A., Seeman,N.C. and Breslauer,K.J. (1987) The melting behavior of a DNA junction structure – a calorimetric and spectroscopic study. *Biopolymers*, **26**, 1621–1634.

36. Gueron,M., Kochoyan,M. and Leroy,J.L. (1987) A single-mode of DNA base-pair opening drives imino proton-exchange. *Nature*, **328**, 89–92.

37. Nonin,S., Leroy,J.L. and Gueron,M. (1995) Terminal base-pairs of oligodeoxynucleotides – imino proton-exchange and fraying. *Biochemistry*, **34**, 10652–10659.

38. Williams,D.J. and Hall,K.B. (2000) Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J. Mol. Biol.*, **297**, 1045–1061.

39. Hu,J.M. and Boyer,M. (2006) Hepatitis B virus reverse transcriptase and epsilon RNA sequences required for specific interaction *in vitro*. *J. Virol.*, **80**, 2141–2150.

40. Milligan,J.F. and Uhlenbeck,O.C. (1989) Determination of RNA protein contacts using thiophosphate substitutions. *Biochemistry*, **28**, 2849–2855.

41. Cromsigt,J.A.M.T.C., Schleucher,J., Kidd-Ljunggren,K. and Wijmenga,S.S. (2000) Synthesis of specifically deuterated nucleotides for NMR studies on RNA. *J. Biomol. Struct. Dyn.*, 211–219.

42. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

43. Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48–52.

44. Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2**, 661–665.

45. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) Nmrpipe – a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR*, **6**, 277–293.

46. Kneller,D.G. and Kuntz,I.D. (1993) Ucsf Sparky – an NMR display, annotation and assignment tool. *J. Cell. Biochem.*, 254–254.

47. Stein,E.G., Rice,L.M. and Brunger,A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.*, **124**, 154–164.

48. Wu,B., Girard,F., van Buuren,B., Schleucher,J., Tessari,M. and Wijmenga,S. (2004) Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold. *Nucleic Acids Res.*, **32**, 3228–3239.

49. Wijmenga,S.S., Kruithof,M. and Hilbers,C.W. (1997) Analysis of H-1 chemical shifts in DNA: assessment of the reliability of H-1 chemical shift calculations for use in structure refinement. *J. Biomol. NMR*, **10**, 337–350.

50. Huthoff,H., Girard,F., Wijmenga,S.S. and Berkhout,B. (2004) Evidence for a base triple in the free HIV-1 TAR RNA. *RNA*, **10**, 412–423.