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

Received January 15, 2007; Revised and Accepted February 19, 2007

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 ALs 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 ALs were both found to be capped by a stable well-structured UGUU tetraloop. All avian ALs 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 carries 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 (3.2kb), 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 et al. 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

*To whom correspondence should be addressed. Tel: +31 24 3653384/2678; Fax: +31 24 3652112; Email: s.wijmenga@nmr.ru.nl

© 2007 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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.

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. U15 of ALD (and ALH) is then numbered U2590.

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 [1H,1H]-NOESY and [1H,13N]-HSQC spectra in H2O (Figure 2B and C, and S1) and extended to non-exchangeable resonances using [1H,1H]-NOESY spectra and [1H,13C]-HSQC in D2O (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. A6–H2 ↔ G24–H1’, A27–H2 ↔ C4–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. A8–H2 ↔ G23–H1’/G22–H8/G22–NH1, A8–H1’ ↔ G23–NH1, A6–H2 ↔ G24–H8, A6–H1’ ↔ G24–NH1) and standard A-helix chemical shifts. The lower and central stem experience stacking interaction as evidenced by the sequential NOE contacts (G23–NH1 ↔ U15–NH3 and between non-exchangeable protons of residues 4 and 5). Residue U25 is bulged out of the helix, because U25–H5 proton resonates close to 6.0 p.p.m., experiencing little or no ring current (S2) and sequential NOE contacts between G24/U25 and U25/G26 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 U10:C14 shows sequential NOE contacts to both U9–G20 and G11–U18 characteristic for helical stacking. Furthermore, the H1’/H5/H6 chemical shifts of U10 and C14 are nearly A-helical, demonstrating the presence of ring current and thus, stacking interactions (S2). Therefore, both U10 and C14 must be inserted into the helix. Although one of two unassigned imino resonances could be the U10 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 [ΔH1'-H12 < 3 Hz, (23)] characteristic for an A-helix, only the sugar moieties of loop residues G14, U15 and U16 adopt a C2'-endo conformation [ΔH1'-H12 > 6 Hz, (23)]. Finally, the narrow spread around A-helix-like values of 31P 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 1H-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 G11–CUGUUGU18 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 G11–CUGUUGU18 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 G11–CUGUUGU18 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 U13–U16 base pair stacked onto the C12–G17 closing base pair, as well as stacking of G14 onto U13 on the major groove side, while U15 does not stack (Figure 2E, left, center). The stacking of the pertinent residues is confirmed by their back-calculated and experimental Δδ values close to 0 (Figure 2E, right). U13 does not stack at all, so that Δδ > 0. This lack of stacking of U15 also leads to U16 Δδ > 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 G14 is somewhat less (Figure 2E, left), suggesting a certain degree of mobility. No convergence is reached for U15, suggesting a large degree of flexibility. For this residue the experimental Δδ fall just outside the error bar of the back-calculated Δδ (Figure 2E, right).
Figure 2. Solution-NMR investigation of the *duck* apical stem-loop structure (*ALD*). (A) $^1$H-imino spectra (5°C, H$_2$O); 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$_2$O); (C) Imino region of *ALD* $[^1$H,$^3$H]-NOESY spectrum (15°C, H$_2$O); (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 $^{13}$C and back-calculated $^{13}$C equals the mean of ($^{13}$calc$^{1H_0}$, H$_5$, H$_6$, H$_8$) or ($^{13}$exp$^{1H_0}$, H$_5$, H$_6$, H$_8$) and back-calculated $\Delta \delta$ (dark gray). $\Delta \delta$ equals the mean of ($^{13}$calc$^{1H_0}$, H$_5$, H$_6$, H$_8$) - $^{13}$calc$^{1H_0}$, H$_5$, H$_6$, H$_8$). The chemical shifts, $^{13}$calc, were averaged over the 10 lowest-energy structures. Residues for which $\Delta \delta > 0$, 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).
This can be explained by occasional stacking of U₁₅ not fully accounted for in the ensemble of calculated structures.

The U₁₃:U₁₆ 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 U₁₃-O₂:U₁₆-NH₃ and U₁₅-NH₃:U₁₆-O₄. Efficient H-bonding of U₁₃-NH₃: U₁₆-O₄ in the U₁₅:U₁₆ pair together with stacking of G₁₄ can account for the protection of the U₁₃-NH₃ (see above, Figure 2A). In the derived structure model, the U₁₅-NH₃, although H-bonded, is more solvent exposed due to the high flexibility and lack of stacking of U₁₅. This may explain why the U₁₅-NH₃ resonance is not detected in the [¹H,¹H]-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 U₁₃:U₁₆ base-pair (type cis-wobble) onto which G₁₄ caps, while U₁₅ 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 by 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-triloop also called a ‘lone pair UGU tri-loop’ (21).

Structural Investigation of the duck HBV primer loop (PLD). Combining data from [¹H,¹H]-NOESY (Figure 4C) and [¹H,¹⁵N]-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 A₆:U₃₂. The U₇:G₃₁ wobble base pair at the top of the lower stem, closing the lower part of the primer loop, was not observed in the [¹H,¹H]-NOESY. It is thus open or in fast exchange with an open form. In the upper stem the G₁₄:C₂₉ 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 U₇:G₃₁ 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: U₅-H₆ ⇔ G₃₁-H¹’, U₁₀-H₆ ⇔ G₃₁-H¹’, U₁₀-H₆ ⇔ A₁₂-H¹’, U₁₁-H₆ ⇔ U₃₀-H¹’ and A₁₂-H₂ ⇔ C₂₉-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 A₁₂ and C₁₃ 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, Mg²⁺) was investigated. As no significant effect was observed when using Mg²⁺, 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
The transition is mono-molecular because RNA concentration does not affect $T_m$ (Table 1). The standard Gibbs free energy change ($\Delta G^\circ_{\text{exp}}$) of the transition is $5.3$ kcal/mol 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 C29 in the $ALD_{n-1}$ sequence increases $\Delta G^\circ$ with $2.1$ kcal/mol 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 $-0.05$ kcal/mol, leading to a final $\Delta G^\circ_{\text{exp}}$ of $-5.3$ kcal/mol for full-length $ALD$.

Using Turner’s tables and the NMR-derived secondary structure model, we computed a predicted $\Delta G^\circ_{\text{calc}}$ of $-2.5$ kcal/mol. Thus $ALD$ is predicted to be $2.0–3.0$ kcal/mol more stable than what is experimentally observed. This is a significant difference as $\Delta G^\circ_{\text{calc}}$ can typically be predicted with an error margin of 10% (34).

The 2.0–3.0 kcal/mol 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 $T_m$ of $63.1\, ^\circ\text{C}$ 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 $\Delta G^\circ$ of $-2.5$ kcal/mol. This value matches perfectly with the one computed using Turner’s tables and the NMR-derived secondary structure model ($-2.4$ kcal/mol, Table 1).

Strikingly, even though $ALH$ is clearly less structured than $ALD$, their experimentally derived melting temperatures are in the same range: $63.1\, ^\circ\text{C}$ for $ALH$ and $64.9\, ^\circ\text{C}$.
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 ΔG° = 0 at the melting temperature and both sequences have approximately the same Tm 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 Tm of 79.5°C (with 0.15 M NaCl). The absence of an RNA concentration effect on the Tm indicates that the transition is mono-molecular.

![Figure 5](image.png)

**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.

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

<table>
<thead>
<tr>
<th>Sample</th>
<th>[RNA] μmol·l⁻¹</th>
<th>[Na⁺] μmol·l⁻¹</th>
<th>Tm°C</th>
<th>ΔH° kcal·mol⁻¹</th>
<th>ΔS° cal·mol⁻¹·K⁻¹</th>
<th>ΔG° kcal·mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALD</td>
<td>2.4</td>
<td>1.0</td>
<td>64.9</td>
<td>−64.2</td>
<td>−190.0</td>
<td>−5.3/−6.3*</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>0.15</td>
<td>56.8</td>
<td>−63.7</td>
<td>−193.2</td>
<td>−3.8</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.15</td>
<td>57.6</td>
<td>−63.0</td>
<td>−190.6</td>
<td>−2.5</td>
</tr>
<tr>
<td>ALH</td>
<td>1.7</td>
<td>1.0</td>
<td>63.1</td>
<td>−53.5*</td>
<td>−185.6*</td>
<td>−4.2*</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>0.15</td>
<td>57.9</td>
<td>−57.0</td>
<td>−111.7</td>
<td>−2.3</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.15</td>
<td>57.7</td>
<td>−27.2</td>
<td>−81.8</td>
<td>−1.9</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>0.15</td>
<td>79.5</td>
<td>−92.8</td>
<td>−263.2</td>
<td>−11.1</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.15</td>
<td>80.0</td>
<td>−96.6</td>
<td>−273.5</td>
<td>−11.8</td>
</tr>
</tbody>
</table>

The values followed by * and ** are computed values derived with Mfold (42) and with tables by Serra and Turner (34), respectively. *Value corrected for dangling end-effect in ALD (see text). For ALH, the predicted ΔG° computed with Mfold is found to be slightly lower (ΔG° = −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.

### 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 1H-imino NMR spectra of ALD were followed from 5 to 40°C in steps of 5°C (Materials and Methods section). At 40°C, i.e. below the Tm observed by UV melting (see above), the imino resonances assigned to G22-NH1, U32-NH3 and G35-C=H1 are still intense and narrow. Also, the peaks assigned to G11-NH1, G17-NH1 and U18-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 U21-NH3, G22-NH1 and U23-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
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 \((T_m = 45.7^\circ C, \Delta G^\circ = -1.38 \text{ kcal-mol}^{-1})\), 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 \(\sim 85^\circ C\) (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 UUGC tetraloop \([T_m > 70^\circ C, (38)]\) 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 UUGC loop and the stems.

To unambiguously identify the residues involved in the first and second melting transitions, the melting process was followed by \(^1\text{H} \text{ imino NMR (Figure 6B)}. At 40^\circ C, all broad resonances between 11.7 and 10.5 p.p.m. (orange in Figure 6B) as well as the \(U_{20} \ ^1\text{H-imino resonance at 14.3 p.p.m. (red in Figure 6B)}\) have completely vanished. All other \(^1\text{H-imino resonances remain sharp and intense. The} \(^1\text{H-imino of } U_{20} \text{ engaged in the closing UG base pair of the UUGC tetraloop, is also broadened, but remains detectable (red in Figure 6B). In summary, we observe that the resonances that vanish at } 40^\circ C \text{ are within/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: GCUGU2590UGU [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 U2588:U2591 base pair (type cis-wobble) stacked onto the closing base pair C2587:G2592, while residue G2589 is in turn stacked onto U2588 on the major groove side. Most interesting is the flexible U2590, 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 G2592 and G2586 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 C2587:G2592 and G2586:U2593 with similar conformation in duck and heron strains. Hence, the sequence UGUU, closed by the C2587:G2592/G2586:U2593 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 GCUGU2590UGU 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 U2590C causes complete loss of binding, without likely affecting the tetraloop structure, we propose that the solvent-exposed U2590 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...
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 (\(\approx30^\circ\text{C}\)), 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 \(\varepsilon\)-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 \(\varepsilon\)-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 \(\varepsilon\)-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\(_{598}\)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 \(\approx13\), \(\approx8\) and \(\approx4\) 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 \(\approx8\) 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 \(\varepsilon\) 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 \(\varepsilon\) to P, but instead for encapsidation. This interaction with core proteins may in human \(\varepsilon\) 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 *\(\varepsilon\)* *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* \(\varepsilon\) 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\(^2\)H\(_1\), \(3\), \(4\), \(5\), \(s\)=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 as control to support and confirm NMR resonance assignments (S6). PLD and PLD-U30 were each expressed in 3 different forms, for instance non-isotopically labeled, \(^{15}\)C/\(^{15}\)N\(^2\)H\(_1\), \(3\), \(4\), \(5\), \(s\)=Uracil and \(^{13}\)C\(^{15}\)N\(^2\)H\(_1\), \(3\), \(4\), \(5\), \(s\)=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·min⁻¹ on a Varian Cary 300-Bio UV-Vis spectrophotometer. 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 (x₄) and offset (x₅),

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

The normalization constant (x₁) should now be 1, while baseline (x₄) and offset (x₅) 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 (ΔH°, ΔS° and ΔG°) were subsequently derived from the van’t Hoff plot of the normalized and baseline-corrected melting curves, again 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, ¹³C/¹³N/²H-U labeled ALD, ALH, PLD, ¹³C/¹³N/²H-U labeled PLD and ²H-U labeled PLD, respectively. Spectra in water (95:5, H₂O:D₂O) 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 x 700 complex points for a spectral width of 14 kHz in both dimensions. For the NMR melting experiments, 1D-WFB/watergate ³H 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,¹³N]-HMOC-WFB was recorded on a 600 MHz Varian Inova spectrometer, collecting 2048 x 400 points for a spectral width of 15 and 2.5 kHz, respectively. In D₂O, 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 x 800 complex points were acquired over a spectral width of 8 kHz in both direct and indirect dimensions. [¹H,¹H]-DFQ-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,¹³C]-HSQC and ¹³C-decoupled [¹H,¹H]-NOESY spectra of the U-labeled ALD sample were recorded for assignment purposes (23,44,41). Finally, 1D ³¹P and [¹H,³¹P]-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 G₁₁CUGUUUGU₁₈ 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 Δδ values were computed as described in the legend of Figure 2E.

SUPPLEMENTARY DATA

Supplementary data is available at NAR Online.

ACKNOWLEDGEMENTS

Valuable discussions with Prof. Michael Nassal (Freiburg, Germany) on different aspects of the HBV P-e recognition are gratefully acknowledged. This work was supported by grants from the Dutch Science Foundation (SW), the EU (FP6 STREP, FSG-VRNA, SW), the Swedish Natural Science Research Council (SW), Biotechnology Fund Umea University (SW) and the Foundation for Strategic Research Sweden/I&V (SW). Funding to pay the Open Access publication charge was provided by EU FP6, FSG-V-RNA.

Conflict of interest statement. None declared.

REFERENCES