Distinct families of cis-acting RNA replication epsilon elements from hepatitis B viruses

Augustine Chen and Chris M. Brown*
Biochemistry and Genetics Otago; University of Otago; Dunedin, New Zealand

**Keywords:** pgRNA, ε element, encapsidation, hepadnavirus, RNA secondary structure

Submitted: 09/28/11
Revised: 11/04/11
Accepted: 10/19/11
http://dx.doi.org/10.4161/rna.18649
*Correspondence to: Chris M. Brown; Email: chris.brown@otago.ac.nz

The hepadnavirus encapsidation signal, epsilon (ε), is an RNA structure located at the 5′ end of the viral pregenomic RNA. It is essential for viral replication and functions in polymerase protein binding and priming. This structure could also have potential regulatory roles in controlling the expression of viral replicative proteins. In addition to its structure, the primary sequence of this RNA element has crucial functional roles in the viral lifecycle. Although the ε element in hepadnaviruses share common critical functions, there are some significant differences in mammalian and avian hepadnaviruses, which include both sequence and structural variations.

Here we present several covariance models for ε elements from the Hepadnaviridae. The model building included experimentally determined data from previous studies using chemical probing and NMR analysis. These models have sufficient similarity to comprise a clan. The clan has in common a highly conserved overall structure consisting of a lower-stem, bulge, upper-stem and apical-loop.

The models differ in functionally critical regions—notably the two types of avian ε elements have a tetra-loop (UGUU) including a non-canonical UU base pair, while the hepatitis B virus (HBV) ε has a tri-loop (UGU). The avian ε elements have a less stable dynamic structure in the upper stem. Comparisons between these models and all other Rfam models, and searches of genomes, showed these structures are specific to the Hepadnaviridae. Two family models and the clan are available from the Rfam database.

**Hepatitis B Virus**

The human hepatitis B virus (HBV) is a major health problem worldwide with an estimated 370 million individuals chronically infected. Chronically infected patients have an increased risk of developing liver cirrhosis and liver cancer resulting in over a million deaths annually.1,2 HBV is a member of the Hepadnaviridae, a family of small hepatotropic DNA viruses. Hepadnaviruses are known to infect certain mammals (orthohepadnavirus) and birds (avihepadnavirus). These viruses have a unique replication lifecycle in that their partially double-stranded DNA genomes are replicated through an RNA intermediate, the pregenomic RNA (pgRNA).3 Hepadnaviruses are related to retroviruses in that they are both retrotranscribing viruses and share some general characteristics.

Current antiviral drugs such as interferon α and nucleoside analogs, while effective in some cases, have problems of limited efficacy and viral resistance after prolonged treatment.4,5 A better understanding between viral and host factors is therefore necessary to facilitate novel antiviral drugs and strategies. A key cis-acting RNA element that acts at several steps in the process is the ε encapsidation signal.

**The Structure and Location of ε Elements in Hepadnavirus RNAs**

The pgRNA also serves as the mRNA template for the translation of the replicative proteins, the core and polymerase (P) protein.6–8 The pgRNA is one of two greater than genome length mRNAs transcribed from the viral genomes. The
other mRNA being the precore RNA (pcRNA), from which the precore protein is translated.11,12 (Fig. 1)

The Functions of the ϵ Element in Reverse Transcription and Replication

In hepadnaviruses, the processes of reverse transcription and encapsidation of the pgRNA are facilitated by the ϵ encapsidation signal. The ϵ element spans a region of approximately 60 nucleotides and is located at both the 5’ and 3’ ends of the pgRNA and pcRNA. While both the pgRNA and pcRNA are translated, only the pgRNA is reverse transcribed and encapsidated. Efficient translation of the precore protein across the pcRNA ϵ element melts the RNA structure and permits pcRNA encapsidation.13-15 Furthermore, only the 5’ ϵ of the pgRNA has been shown to be essential for these processes, whereas the 3’ ϵ which have slightly different conformational structures is not used.16

During reverse transcription, the 5’ ϵ element recruits the P protein to the upper stem, then the TP domain of P initiates priming at the conserved bulged UUCA and the synthesis of the minus-strand DNA (Fig. 2). This process involves conformational changes in both the ϵ structure and bound P protein which open up the base pairing in the upper stem allowing reverse transcription from the bulge.17,19-21 These conformational changes and recruitment of P protein are also facilitated by cellular chaperones.17,19-21

Most of the encapsidation process for hepadnaviruses were determined from studies done on the avian Duck hepatitis B virus (DHBV) in vitro.15,18,20-22 In these studies, the sequence and structure at the upper stem, bulge and also sequences at the 5’ end DR1 acceptor site where complementary base-pairing to the ϵ element but due to the translation of the precore does not function in encapsidation.

There are significant variations between the different members of hepadnaviruses. These include notable primary sequence difference within the ϵ element between the avian and mammalian hepadnaviruses. There are also distinct differences in binding requirements for P protein at the upper stem which is less well base paired in most avian hepadnaviruses (except some DHBV, Fig. 2). In addition, the initiation of DNA synthesis successfully shown in the DHBV system in vitro has so far been unable to be shown for HBV, indicating significant differences in the elements.

This study aims to build covariance models of hepadnavirus ϵ elements. The ϵ element is well conserved in overall structure between the mammalian and avian hepadnaviruses, despite the viruses having significant genome divergence and differing in the presence or absence of other cis-acting elements. The avian and mammalian hepadnaviruses were determined from public DNA databases (see Materials and Methods). The sequences were chosen to

![Figure 1. A schematic representation of the greater than genome-length HBV pgRNA and pcRNA. Cs RNA elements, namely, ϵ, direct repeat 1 and direct repeat 2 (DR1 and DR2). The ϵ structure is present at both 5’ and 3’ termini of the pgRNA, but only the 5’ ϵ of the pgRNA is selectively recognized for packaging. It facilitates polymerase (P) binding as depicted by the Terminal Domain (TP) and Reverse Transcriptase (RT) domain. The TP domain initiates protein priming at the bulge of the 5’ ϵ and after initial priming translocates to the 3’ and DR1 acceptor site where complementary base-pairing to the ϵ element allows the RT to initiate minus strand DNA synthesis. The pcRNA is exactly the same as the pgRNA except for a longer 5’ leader, it encodes the precore ORF and also contains the ϵ element but due to the translation of the precore does not function in encapsidation.](www.landesbioscience.com)
represent the diversity of HBV genotypes (A-H) in a reference alignment used in Panjawarayan et al.28 Although the genotypes differ by over 8% sequence overall, the ε element is highly conserved due to its multiple functions. The secondary structure is conserved in all 32 members of the reference alignment, except for an A-G mismatch in the middle of the lower stem in all four genotype A viruses (orange in Fig. 3A). This mismatch is unexpected, but non-canonical A-G base pairs can be
Figure 3. Alignments of families of \( \varepsilon \) elements—HBV (A), DHBV (B) HHHBV (C) and a combined model AHBV (D). The SS line represents the consensus structure in dot-bracket notation, dots are unpaired, brackets are paired. In A-C, compensating base changes are depicted in green, base pairs incompatible with the consensus structure in orange. In the combined model D, blue shading represents compatibility with the structure line (SS_cons). Stem (Sm) loops and bulges are indicated. These Stockholm format files and models are available in the Supplementary Material.
accommodated with some distortion within an A-helix. However, this may indicate structural tolerance at this position. The closely related orthohepadnaviruses (ground squirrel and woodchuck hepatitis virus) ε elements have an inserted C after this point, also indicating tolerance (see Materials and Methods).

Some sequences show compensating base changes within the structure (green in Fig. 3A). These changes give independent support for the existence of these base pairs. Orthohepadnaviruses also have two of these compensating base changes, but no additional changes not seen in the HBV alignment. Notably one HBV genotype C (AB048704) has a compensating G-U closing pair adjacent to the apical tri-loop, providing additional covariance support for this pair previously observed in the NMR structure.28

A multiple alignment was assembled and manually refined by structure and sequence conservation to form a curated seed alignment (Fig. 3). Alignments of these four elements: HBVs, DHBVs, HHBV, and a combination of these two—Avian HBV ε (AHBV, left) were shown in Figure 3A–D and available in supplementary files. HBV ε (spolion (RF01407) and AHBV ε (spolion (RF01351) are also available through Rfam with corresponding Wikipedia entities.

Due to the significant sequence difference and function between the ε element of mammalian, heron and duck hepadnaviridae, alignments were initially done for each and separate covariance models built for each family (Fig. 3B and C). For DHBV there are several Chinese isolates for each family (scores 84, 271, 85 and 88). These four models were used for further analysis.

Comparison of the ε Models to Each Other to All Other Rfam Families

The four covariance models were compared with each, using CMCompare. In a comparison of related and unrelated Rfam models a score of over 20, or E < 1.0 were considered worthy of note. However about 7.4% of pairwise comparisons of Rfam models had scores over 20, and 6.3% over 28.26 The HHBV and DHBV model (AHBV, left) were shown to be compatible (blue shading Fig. 3B). These four models were used for further analysis.

Figure 3A

4 3 2 1

AHBV ε
DHBV ε
HHBV ε
DHCV ε

Figure 3A shows the upper stem extending into the loop, with less pairing in the upper stem (similar to HHBV, C) to form a new clan proposed here, so are not included as part of the clan.

Searching Sequence Databases for Similar Elements

These four covariance models (Fig. 3 and 4) were calibrated (using cmcalibrate) and used to search on both strands of all the curated RfamSeq viral genomes, the viral division of GenBank and RFamSeq10 using cmsearch (see Materials and Methods). Cmsearch generates a bit score report based on the match of the model to the sequence. It also provides an E value (which corresponds to the expected number of false positives in a database of this size). Hits with E values of < 0.1 are considered trustworthy.28

The HBVs model was built from sequences representing the diversity of the common HBV genotypes. It has significant matches to 6,910 sequences in the RFamSeq10 database, all of which are from mammalian HBVs. These matches represent the diversity of HBV genotypes (A-H). The search identified some additional mammalian HBV viruses, e.g., woodchuck HBV. Some apparently diverse matches are due to misclassifications in the EMBOJ taxonomy; one match is classed by EMBL and RFamSeq10 as

Figure 3B

4 3 2 1

AHBV ε
DHBV ε
HHBV ε
DHCV ε

Figure 3B shows the upper stem extending into the loop, with less pairing in the upper stem (similar to HHBV, C) to form a new clan proposed here, so are not included as part of the clan.
being a hepatitis A virus, but is clearly a HBV, one is classed as rock squirrel genome, but is a rock squirrel HBV. There is also a match to a synthetic human HBV containing construct. The next best match in RFamSeq10 is not significant, indicating that good matches to this element are not found in cellular genomes (e.g., the human host genome). Although portions of HBV can be integrated into the human genome, it is not unexpected that the reference human genome does not contain an s-like sequence. Similar results were obtained from the GenBank viral division and RefSeq viruses. DHBV, HHBV and the combined AHBV model all match the same six avian hepadnaviruses in RefSeq viruses (scores: 53–83, E ≤ 10^{-7}) with 54 hits in RFamSeq10, and 58 hits in the viral division of GenBank. Indeed, the combined avian model identifies the same set with better scores (score ≥ 40; E ≤ 5 × 10^{-7}). This combined model constitutes RFam model RF01313. The separate DHBV, HHBV models presented here are also available (see Supplemental Materials). Notably the combined model recognizes divergent viruses (e.g., Stork HBV sequences AJ251937).

The next best matches in viral genomes are marginal matches to long bacteriophage DNA genomes (DHBV, NC_015289, Score: 24, E = 0.43; HHBV, NC_012697, Score: 22, E = 0.92). The matched regions encode bacteriophage proteins and do not appear to be biologically significant.

There were no significant matches to other retro-transcribing viruses (best score: 5.0, E = 1.6). There were also no significant matches to HDV, this is not unexpected, as HDV is only dependent on HBV for envelopment and not encapsidation.

**Discussion**

We have shown that the RNA families of s replication elements proposed here comprise a clan with both RNA structural and functional similarities. The hepadnavirus s plays several key roles in the viral lifecycles and has a similar role in the avian hepadnavirus that the reference human genome does not contain an s-like sequence. Similar results were obtained from the GenBank viral division and RefSeq viruses.

DHBV, HHBV and the combined AHBV model all match the same six avian hepadnaviruses in RefSeq viruses (scores: 53–83, E ≤ 10^{-7}) with 54 hits in RFamSeq10, and 58 hits in the viral division of GenBank. Indeed, the combined avian model identifies the same set with better scores (score ≥ 40; E ≤ 5 × 10^{-7}). This combined model constitutes RFam model RF01313. The separate DHBV, HHBV models presented here are also available (see Supplemental Materials). Notably the combined model recognizes divergent viruses (e.g., Stork HBV sequences AJ251937).

The next best matches in viral genomes are marginal matches to long bacteriophage DNA genomes (DHBV, NC_015289, Score: 24, E = 0.43; HHBV, NC_012697, Score: 22, E = 0.92). The matched regions encode bacteriophage proteins and do not appear to be biologically significant.

There were no significant matches to other retro-transcribing viruses (best score: 5.0, E = 1.6). There were also no significant matches to HDV, this is not unexpected, as HDV is only dependent on HBV for envelopment and not encapsidation.

**Discussion**

We have shown that the RNA families of s replication elements proposed here comprise a clan with both RNA structural and functional similarities. The hepadnavirus s plays several key roles in the viral lifecycles and has a similar role in the avian hepadnavirus...

The models proposed here are specific to the hepadnaviruses. These models add to the basis for further research into the structural and functional probing and also NMR analyses on HBV, DHBV, and HHBV.

**Materials and Methods**

Sequences were extracted representing the diversity of mammalian and avian hepadnavirus genomes. The principal features of the structures in different functional states were extracted from the literature. For HBV there are over 6,000 sequences in public databases but many are identical in this region. The sequences chosen were from a previously published HBV reference set to represent common diversity and are similar to other published genotyping sets for HBV. The sequences of orthohepadnaviruses were also compared (NC_001484, NC_004107). They have an inserted C relative to HBV at position 5 (UGUUUCCA) and several compensating changes also found in HBV genomes.

For HHBV and DHBV, fewer sequences are available. There are other members of the avian hepadnaviruses but there is currently too little diversity in the data and insufficient experimental evidence to form an alignment from which to build separate models.

Alignments were done manually using AquaMacs in Ralee mode guided by structural probing and NMR studies (PDB:2OJ7, 2OJ8, 2K5Z, 2IXY, 2IXZ) and considering the modeling done of the lower stem by other groups. These alignments were determined by chemical and enzymatic probing and also NMR analyses on HBVs, DHBVs, and HHBVs.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.
Acknowledgments

Part of this work was supported by a NZ Health Research Council Grant (05/195) to Warren Tate, Elizabeth Poole and C.M.B. and a Human Frontier Science Program grant to Ian Macara, Anne Spang and C.M.B. (RGP0031_2009).

Supplemental Material

References

1. Chen D. Toward elimination and evaluation of hepatitis B virus infection. J Viral Hepat 2011; 18:568-75; PMID:21626272.
2. World Health Organisation. Hepatitis B. Available: www.who.int/csr/disease/hepatitis/bh.pdf. Accessed 2011.
3. Songer G, Messer WS. Hepatitis B virus biology. Microbiol Mol Biol Rev 2000; 64:1-57; PMID:17879476.
4. Trifunac HM. Antiviral therapy and resistance with hepatitis B virus infection. World J Gastroenterol 2007; 13:125-49; PMID:17320760.

Frontier Science Program grant to Ian Macara, Anne Spang and C.M.B. (RGP0031_2009). Thanks to Lauren Duba at Rfam for helpful suggestions on the models.

www.landesbioscience.com/journals/mbio/article/18649

Supplemental materials may be found here:

www.rnabiology/article/18649

© 2012 Landes Bioscience. Do not distribute.

136 RNA Biology Volume 9 Issue 2

1. Chen D. Toward elimination and evaluation of hepatitis B virus infection. J Viral Hepat 2011; 18:568-75; PMID:21626272.
2. World Health Organisation. Hepatitis B. Available: www.who.int/csr/disease/hepatitis/bh.pdf. Accessed 2011.
3. Songer G, Messer WS. Hepatitis B virus biology. Microbiol Mol Biol Rev 2000; 64:1-57; PMID:17879476.
4. Trifunac HM. Antiviral therapy and resistance with hepatitis B virus infection. World J Gastroenterol 2007; 13:125-49; PMID:17320760.

Frontier Science Program grant to Ian Macara, Anne Spang and C.M.B. (RGP0031_2009). Thanks to Lauren Duba at Rfam for helpful suggestions on the models.

www.landesbioscience.com/journals/mbio/article/18649

Supplemental materials may be found here:

www.rnabiology/article/18649

© 2012 Landes Bioscience. Do not distribute.