Published in final edited form as:

Virus Res. 2020 January 15; 276: 197825. doi:10.1016/j.virusres.2019.197825.

Tracing the evolutionary history of hepadnaviruses in terms of e antigen and middle envelope protein expression or processing

Qianru Wang^{a,1}, Yanli Qin^{b,1}, Jing Zhang^a, Lucy Jia^c, Shuwen Fu^a, Yongxiang Wang^a, Jisu Li^c, Shuping Tong^{a,c}

^aDepartment of Pathobiology, Key Laboratory of Medical Molecular Virology, School of Basic Medical Sciences, Fudan University, Shanghai, China;

^bDepartment of Infectious Diseases, Huashan Hospital, Fudan University, Shanghai, China;

^cLiver Research Center, Rhode Island Hospital and Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA.

Abstract

Hepatitis B virus (HBV) is the prototype of hepadnaviruses, which can be subgrouped into orthohepadnaviruses infecting mammals, avihepadnaviruses of birds, metahepadnaviruses of fish, and herpetohepadnaviruses of amphibians and reptiles. The middle (M) envelope protein and e antigen are new additions in the evolution of hepadnaviruses. They are alternative translation products of the transcripts for small (S) envelope and core proteins, respectively. For HBV, e antigen is converted from precore/core protein by removal of N-terminal signal peptide followed by furin-mediated cleavage of the basic C-terminus. This study compared old and newly discovered hepadnaviruses for their envelope protein and e antigen expression or processing. The S protein of bat hepatitis B virus (BHBV) and two metahepadnaviruses is probably myristoylated, in addition to two avihepadnaviruses. While most orthohepadnaviruses express a functional M protein with N-linked glycosylation near the amino-terminus, most metahepadnaviruses and herpetohepadnaviruses probably do not. These viruses and one orthohepadnavirus, the shrew hepatitis B virus, lack an open precore region required for e antigen expression. Potential furin cleavage sites (RXXR sequence) can be found in e antigen precursors of orthohepadnaviruses and avihepadnaviruses. Despite much larger precore/core proteins of avihepadnaviruses and their limited sequence homology with those of orthohepadnaviruses, their proximal RXXR motif can be aligned with a distal RXXR motif for orthohepadnaviruses. Thus, furin or another basic endopeptidase is probably the shared enzyme for hepadnaviral e antigen maturation. A precore-derived cysteine residue is involved in forming intramolecular disulfide bond of HBV e antigen to prevent particle formation, and such a cysteine residue is conserved for both orthohepadnaviruses and avihepadnaviruses. All orthohepadnaviruses have an X gene, while all avihepadnaviruses can

Correspondence to: Shuping Tong, Liver Research Center, 55 Claverick Street, Providence, RI 02903, USA. shuping_tong_md@brown.edu.

¹These authors contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

express the e antigen. M protein expression appears to be the most recent event in the evolution of hepadnaviruses.

Keywords

Hepatitis B virus; hepadnavirus; evolution; e antigen; myristoylation; middle envelope protein

INTRODUCTION

Human hepatitis B virus (HBV) is the prototype of a family of hepatotropic DNA viruses, or hepadnaviridae (Seeger and Mason, 2015). Its 3.2-kb covalently closed circular genome harbors the core, polymerase (P), envelope, and X genes, with the 2.4-kb P gene overlapping partially with core and X genes but completely with the envelope gene (Fig. 1A). Both the core and surface (envelope) genes have 5' extensions to generate extra protein forms with distinct functions. Thus while core protein is product of core gene alone, translation initiated from the precore ATG codon generates fused precore/core protein. It is subject to proteolytic cleavage in the secretory pathway for conversion to hepatitis B e antigen (HBeAg), a serological marker of HBV infection (Fig. 2B) (Magnius and Espmark, 1972). The preS region is subdivided into preS1 and preS2 to enable the expression of large (L) and middle (M) envelope proteins in addition to the abundant small (S) envelope protein (Fig. 1B & 2A). M protein has extra preS2 domain than the S protein, while L protein has extra preS1 domain relative to M protein. Altogether, a total of seven proteins are expressed from the four genes closely spaced inside a small circular genome (Fig. 1B). The core protein is the building block of core particle, inside which the multi-functional P protein catalyzes genome replication. HBeAg is not required for genome replication, but rather implicated in the establishment of persistent HBV infection by induction of immune tolerance (Milich et al., 1990; Milich and Liang, 2003). The S protein is the major envelope protein on both the 42-nm virions (enveloped core particles) and smaller (22-nm) subviral particles (SVPs; empty envelope protein particles), with the latter exceeding the former by a factor of 100– 100,000 fold (Ganem and Prince, 2004). The L protein is selectively enriched on virions and 22-nm filaments (aggregated subviral particles). The M protein promotes virion secretion but is not essential for virion or SVP formation/release or virion infectivity, and the late stage of chronic HBV infection may select for point mutations of the preS2 ATG codon to specifically silence M protein expression (Fig. 3) (Fernholz et al., 1993; Garcia et al., 2009; Ni et al., 2010). Similarly, a nonsense mutation in the precore region (G1896A) is often selected at the HBeAg-negative phase of chronic infection to ablate HBeAg expression (Carman et al., 1989; Okamoto et al., 1990; Tong et al., 1990).

Animal hepadnaviruses were first discovered about 40 years ago in woodchucks, ground squirrels, and Pekin ducks (Marion et al., 1980; Mason et al., 1980; Summers et al., 1978). Duck hepatitis B virus (DHBV) was instrumental in characterizing the molecular biology and lifecycle of hepadnaviruses such as genome replication by reverse transcription from an RNA precursor (Summers and Mason, 1982). The 3.3-kb genomes of woodchuck hepatitis virus (WHV) and ground squirrel hepatitis virus (GSHV) also harbor 4 genes to encode 7 proteins, while the 3.0-kb DHBV genome lacks X gene and does not have an internal ATG

codon to express the M protein. Thus, it represents a simplified version of hepadnavirus. Subsequently genomes of heron, snow goose, crane, stork, Ross's goose, sheldgoose, and parrot hepatitis B viruses were cloned and sequenced (Chang et al., 1999; Guo et al., 2005; Piasecki et al., 2012; Pult et al., 2001; Sprengel et al., 1988). Their genetic organization is quite similar to that of DHBV, although some might have an X-like gene. Besides HBV, primate hepadnaviruses have also been discovered in woolly monkeys and capuchin monkeys (de Carvalho Dominguez Souza et al., 2018; Lanford et al., 1998). Hepadnaviral genomes were also isolated from domestic cats (Aghazadeh et al., 2018; Lanave et al., 2019), frogs, cichlids, skinks, lizards, and different types of fish (Dill et al., 2016; Hahn et al., 2015; Lauber et al., 2017). Most impressively, a large number of bat hepatitis B virus (BHBV) genomes have been sequenced around the world from horseshoe bat, roundleaf bat, tent-making bat, long-fingered bat, etc (Drexler et al., 2013; He et al., 2013; He et al., 2015; Wang et al., 2017; Nie et al., 2018). Very recently, shrew hepatitis B virus (SHBV) was identified from China, Europe, and Africa (Nie et al., 2019; Rasche et al., 2019). Phylogenetically, hepadnavirus can be grouped into orthohepadnavirus infecting mammals, avihepadnavirus of birds, herpetohepadnavirus of amphibians and reptiles, metahepadnavirus and parahepadnavirus infecting fish (Lauber et al., 2017).

The P protein of a hepadnavirus consists of three functional domains: terminal protein (TP), reverse transcriptase (RT), and RNase H (RH), with the TP and RT domains separated by a nonessential spacer of variable length (Jones and Hu, 2013). The coding sequence for the spacer corresponds to the preS region of the envelope gene. In 2017, sequence analysis revealed a family of non-enveloped fish viruses related to hepadnaviruses (Lauber et al., 2017). These so-called "nackdnaviruses" have a genome size of 2.7–3.1kb, with the 5' end of the P gene overlapping with the 3' end of core gene just like hepadnaviruses. However, the preS region is absent and the S region has no coding capacity. Furthermore, nackdnaviruses lack the X gene and do not have an open precore region for e antigen production. Thus, they are most likely immediate ancestors of simplest hepadnaviruses such as herpetohepadnaviruses and metahepadnaviruses (Fig. 3). All hepadnaviruses require simultaneous expression of L and S proteins to ensure formation and release of enveloped virus particles, and initiation of infection. M protein, X protein, and e antigen apparently evolved later, although whether that occurred in a sequential manner has not been established. Here we analyzed newly sequenced members of hepadnaviruses for their potential to express the M protein and e antigen. We made an educated guess by resorting to experimental data obtained with HBV, DHBV, and WHV. We also paid attention to myristoylation of the envelope proteins.

MATERIALS AND METHODS

Genomic sequences of hepadnaviruses were mostly retrieved from GenBank. Sequences for nackdnaviruses, some herpetohepadnaviruses and metahepadnaviruses were obtained from supplementary file of a previous publication (Lauber et al., 2017). The Blast program was used to align DNA or protein sequence. SignalIP-5.0 was used to predict signal peptide cleavage site. The DNA and protein sequences of different HBV genotypes shown in the figures are the relatively common ones.

RESULTS AND DISCUSSION

1. Two members of herpetohepadnaviruses cannot produce myristoylated L protein.

The HBV L protein is myristoylated at N-terminal glycine residue following removal of the initiating methionine, which is not required for virion formation but essential for infectivity (Bruss et al., 1996; Gripon et al., 1995). Presence of the myristoylation signal (N-terminal MG dipeptide) in L protein is conserved among initially cloned HBV, WHV, GSHV, and DHBV, and also in nearly all the subsequently discovered hepadnaviruses. That in a way helps to define the N-terminus of L protein, especially for BHBV isolates having a variable length of the preS1 region and for some avihepadnaviruses having closely spaced in-frame methionine codons at the 5' preS1 region. Similar to HBV, mutating that glycine codon in the DHBV genome maintained virion morphogenesis but abolished viral infectivity (Macrae et al., 1991). For HBV, myristoylated preS1 peptide 2–48 of genotype D could bind efficiently to sodium taurocholate cotransporting polypeptide (NTCP), the functional HBV receptor (Yan et al., 2012). Human NTCP could also serve as a receptor for some BHBV isolates, and bat NTCP might mediate natural BHBV infection (Drexler et al., 2013). Thus for primate hepadnaviruses and probably BHBV, L protein myristoylation is required for efficient NTCP interaction. However, there is no evidence to suggest NTCP as the receptor for WHV or DHBV, nor does it serve as the receptor for shrew hepatitis B virus (Rasche et al., 2019). Instead two other host proteins were previously identified as binding partners for DHBV L protein and candidate DHBV receptors (reviewed in Li and Tong, 2015), although their interaction with L protein was not myristoylation-dependent. Thus how L protein myristoylation contributes to the infectivity of most non-human hepadnaviruses remains to be clarified. Surprisingly, skink hepatitis B virus and spiny lizard hepatitis B virus do not have MG dipeptide at the N-terminus of L protein (Fig. S1). Both are members of herpetohepadnaviruses. The other member, Tibetan frog hepatitis B virus, has the N-terminal MG sequence (Fig. S2B).

2. Some hepadnaviruses, especially BHBV, probably produce myristoylated S protein.

Although not pointed out in recent publications (Drexler et al., 2013; He et al., 2013; He et al., 2015; Nie et al., 2018; Wang et al., 2017), nearly all the BHBV isolates have MG dipeptide at the beginning of the S domain (Fig. 4), thus raising the possibility of myristoylated S protein. Such a myristoylation signal can also be found in the S protein of domestic cat hepadnavirus, some metahepadnaviruses such as Astatotilapia metahepadnavirus and white sucker hepadnavirus (Fig. 3). As previously reported (Pult et al., 2001), such a signal is also present in S protein of heron hepatitis B virus and stork hepatitis B virus (Fig. 3). All the three stork hepatitis B virus clones with initiating methionine in the S region and all the 9 heron hepatitis B virus clones available in GenBank have the MG dipeptide sequence at the beginning. Herons and storks are evolutionarily closely related. Whether the S protein from hepadnaviruses with N-terminal MG dipeptide is indeed myristoylated, and if that has functional consequences, remain unsettled.

3. Most herpetohepadnaviruses and metahepadnaviruses probably do not express M protein.

Both M and S proteins of HBV are translated from the 2.1-kb RNA with heterogeneous 5' end. Thus, its relatively short preS2 domain (55aa) is probably needed to ensure localization of the preS2 AUG within the 5' end of the longest 2.1-kb RNA for translation initiation. We assume that other hepadnaviruses also rely on the S protein transcript to express M protein. Indeed avihepadnaviruses do not express M protein, and their S gene ATG codon is separated from the upstream in-frame ATG codon in the preS region by 99 – 122 codons (Table 1). The preS2 domain is 60aa for WHV and GSHV capable of M protein expression, 52aa for domestic cat hepadnavirus, and 45 – 50aa for BHBV (Table 1; Fig. 5). For HBV, M protein is glycosylated at N4 in the preS2 domain via the NXS/T motif, although its functional significance has not been examined. N-terminal glycosylation of M protein (if it is indeed expressed) is conserved in most orthohepadnaviruses including woolly monkey hepatitis B virus, Capuchin monkey hepatitis B virus, WHV, and GSHV. As previously noticed (Drexler et al., 2013), most BHBV isolates also have a glycosylation site at the N-terminal preS2 domain (position 2 or 3). However, the glycosylation site is absent in some BHBV isolates (such as JX941466 and KY962686), and both isolates of domestic cat hepadnavirus (Fig. 5; Table 1). The cat virus probably does not express M protein due to poor content of the Kozak sequence (CAAATGT; Table 1). Among herpetohepadnaviruses and metahepadnaviruses, white sucker hepadnavirus, spiny lizard hepatitis B virus, and skink hepatitis B virus might have a preS2 domain of 64aa, 69aa, and 74aa, respectively, although the glycosylation site(s) is not close to the N-terminus (Fig. 5; Table 1). The latter two viruses have a poor Kozak sequence to make M protein expression less likely (Table 1). For Tibetan frog hepatitis B virus, Astatotilapia metahepadnavirus, and tetra metahepadnavirus, the next in-frame ATG codon is 163, 107, and 96 codons upstream of S gene to make M protein expression from the S transcript unlikely (Table 1; Fig. S2B–D). Bluegill hepatitis B virus cannot express M protein due to absence of internal methionine codon within the preS region (Fig. S2A). Overall, most BHBV isolates very likely can express M protein with similar function as that of HBV, whereas among herpetohepadnaviruses and metahepadnaviruses only white sucker hepadnavirus might express an M-like protein.

4. Biological significances of e antigen expression and biosynthetic pathway of HBeAg production.

The e antigen is N- and C-terminally cleaved form of precore/core protein (Fig. 2B). Therefore, a nonsense or frameshift mutation introduced to the precore region can abolish e antigen expression for both DHBV and WHV, but will not impair core protein expression or genome replication in cell culture, and viral infectivity in vivo (Chang et al., 1987; Chen et al., 1992; Schlicht et al., 1987). Nevertheless, a WHV mutant unable to express the e antigen has reduced chronicity rate of 17% when inoculated to neonatal woodchucks in contrast to the parental clone (70%) (Cote et al., 2000). From our own experience, a naturally occurring precore mutant was replication competent in cell culture, but induced transient infection in a chimpanzee (Tong et al., 1991a; Tong et al., 1991b). Therefore, e antigen expression promotes persistent infection, at least for orthohepadnaviruses. The HBV precore region and core gene have coding capacities for 29aa and 183aa, respectively. Translation initiated from

precore ATG will generate a fused precore/core protein (p25). The N-terminal 19 residues serve as the signal peptide to target this protein to the secretory pathway followed by its removal by signal peptidase (Garcia et al., 1988; Standring et al., 1988). The intermediate, p22, has 10 extra residues at the N-terminus than core protein. Further removal of the C-terminal arginine-rich sequence generates p17, the mature HBeAg found in patient sera (Fig. 2B). The second cleavage makes HBeAg unable to bind RNA or DNA, a feature of the core protein required for its accommodation of genome replication. While core protein can dimerize via Cys61-Cys61 inter-molecular disulfide bond to form core particle, this is prevented in HBeAg due to intramolecular disulfide bond between Cys61 and a cysteine (Cys-7) from the C-terminal precore peptide (Nassal and Rieger, 1993; Wasenauer et al., 1993).

The C-terminus of HBeAg purified from patient sera contained TTVV sequence corresponding to residues 146–149 in core protein (Takahashi et al., 1983). A furin inhibitor could block HBeAg secretion from a genotype D clone in cell culture (Messageot et al., 2003). Furin is a proprotein convertase (PC) in the secretory pathway with a minimum recognition sequence of RXXR. Four such motifs can be identified in the precore/core protein of most (non-A) HBV genotypes downstream of the TTVV sequence: ¹⁵¹RRGR¹⁵⁴ (site 1), ¹⁵⁴RSPR¹⁵⁷ (site 2), ¹⁶⁴RRRR¹⁶⁷ (site 3), and ¹⁷²RRRR¹⁷⁵ (site 4) (Fig. 6B). Site-directed mutagenesis revealed secreted HBeAg as cleavage product after ¹⁵¹RRGR¹⁵⁴, or site 1 (Messageot et al., 2003). Low amount of cleavage product at R167 was observed inside cells but not culture supernatant, but it became secreted as HBeAg if R154 site was mutated. Mutating R167 did not prevent cleavage at R154 suggesting no need for sequential processing from the C-terminus. We found serum HBeAg migrated as a single band for genotypes B and D, whereas two large size forms were more prominent for genotype A (Ito et al., 2009). Genotype A has a 6-nt insertion in the core gene. The DR dipeptide inserted converts ¹⁵¹RRGRSPR¹⁵⁷ sequence into ¹⁵¹RRDRGRSPR¹⁵⁹, which not only separates the first and second sites, but also converts the first site from RRGR into the less favorable RRDR (Fig. 6B). Through site-directed mutagenesis and transfection experiments we demonstrated that the two dominant HBeAg species for genotype A correspond to cleavage after site 2 (¹⁵⁶RSPR¹⁵⁹) and site 3 (¹⁶⁶RRRR¹⁶⁹), respectively (Ito et al., 2009). The genotypic difference in HBeAg maturation could explain why a single size form of HBeAg was identified from Japanese patients (Takahashi et al., 1983) and transfection experiments using genotype D (Messageot et al., 2003), whereas in vitro translation or transfection with genotype A clones produced multiple size forms (Salfeld et al., 1989; Standring et al., 1988).

5. All hepadnaviruses probably use a precore-derived cysteine residue to keep e antigen non-particulate.

Both avihepadnaviruses and orthohepadnaviruses are expected to express the e antigen. Despite some early investigation (Schlicht, 1991; Schlicht et al., 1987), the N- and C-terminal cleavage sites in the precore/core protein to generate the DHBV e antigen have not been precisely mapped. Most orthohepadnaviruses have a precore peptide of 28–30 residues with high sequence conservation (Fig. 6A). According to SignalP-5.0 program, cleavage by the signal peptidase will leave the C-terminal 10aa of the precore peptide on mature e antigen (Fig. 6A). This is consistent with experimental work for HBV (Garcia et al., 1988;

Standing et al., 1988). Only tent-making bat hepatitis B virus (TBHBV) has a slightly longer precore peptide of 32aa, with 11aa predicted to remain on mature e antigen. Strikingly, Cys-7 required for intramolecular disulfide bond of HBeAg is conserved for other members of orthohepadnaviruses with an open precore region (Fig. 6A). Avihepadnaviruses have a much longer precore peptide of 43aa. Only after artificially removing a portion of the N-terminal sequence (10aa for DHBV but 5aa for Ross goose hepatitis B virus) would SignalIP-5.0 program predict cleavage by a signal peptidase, with the last 8aa of the precore peptide remaining on mature e antigen (Fig. 6A). Strikingly, Cys-3 (counting from the N-terminus of the core protein) is conserved among all avihepadnaviruses. It is tempting to suggest that as demonstrated for HBeAg (Nassal and Rieger, 1993; Wasenauer et al., 1993), Cys-7 for orthohepadnaviruses and Cys-3 for avihepadnaviruses participate in intramolecular disulfide bond to prevent particle formation.

6. Both orthohepadnaviruses and avihepadnaviruses probably use furin or another basic endopeptidase for e antigen maturation.

For HBV, both the third and fourth potential furin cleavage sites downstream of ¹⁴⁶TTVV¹⁴⁹ consist of RRRR sequence. These two motifs are conserved in most other orthohepadnaviruses (Fig. 6B). The first and second motifs are fused in WHV, GSHV, domestic cat hepadnavirus, and some BHBV isolates similar to non-A HBV genotypes. In some other BHBV isolates these two motifs are separate just like genotype A (the first motif was expanded to two fused sites in KY962686). In other viruses only one motif is retained (woolly monkey hepatitis B virus) or both are lost (capuchin monkey hepatitis B virus and KC790380). Considering that for HBV, cleavage product at the third motif can be released as HBeAg if further cleavage at motifs 1 and 2 was blocked (Messageot et al., 2003), all non-HBV hepadnaviruses should be able to release e antigen if furin or a related PC is responsible for the cleavage. This cleavage, followed by further trimming of the C-terminal arginine residues by a basic carboxypeptidase, may also remove the nuclear localization signal (Yeh et al., 1990). If furin is indeed the enzyme involved, viruses lacking motifs 1 and 2 (such as capuchin monkey hepatitis B virus) will release an e antigen of larger size than viruses with favorable motifs 1 and/or 2 (such as woolly monkey hepatitis B virus having RRRR sequence immediately downstream of TTVV) (Fig. 6B).

Avihepadnaviruses have a much larger core protein (262aa) than orthohepadnaviruses (182 – 190aa) (Fig. 7). Nevertheless, as previously noticed (Schlicht et al., 1987) the TTVV sequence found near the end of HBeAg (Takahashi et al., 1983) is preserved in core protein of nearly all avihepadnaviruses at a similar distance from the C-terminus (Fig. 6B), thus raising the possibility of cleavage after TTVV sequence for maturation of both HBV and DHBV e antigens. Avihepadnaviruses have two conserved RXXR motifs: site one (²⁰⁷RKPR²¹⁰ for most viruses including DHBV) upstream and site 2 (²³³RERR²³⁶ for DHBV) downstream of the ²²¹TTVV²²⁴ sequence. In crane hepatitis B virus and Ross goose hepatitis B virus an additional ²¹⁵RRRR²¹⁸ site occurs in between, while in parrot hepatitis B virus the ²³³RERR²³⁶ motif is lost. Surprisingly, the ²²¹TTVV²²⁴ sequence in DHBV did not align with ¹⁴⁶TTVV¹⁴⁹ of HBV, and the ²⁰⁷RKPR²¹⁰ motif in DHBV (upstream of TTVV) aligned with one of the two conserved RRRR motifs in orthohepadnaviruses (site 3 for HBV) (Fig. 7). Sequence corresponding to site 1 and site 2 in HBV is lost in DHBV due

to a deletion of 5aa relative to genotype A and 3aa relative to non-A genotypes (Figs. 7A & B). The sequence alignment suggests that the third furin cleavage site in HBV (¹⁶⁴RRRR¹⁶⁷) represents the archaic furin cleavage site for orthohepadnaviruses. The proximal sites (sites 1 and 2 for HBV) were added more recently. The alignment also suggests that for avihepadnaviruses, the e antigen is formed by cleavage at ²⁰⁷RKPR²¹⁰ site instead of the ²³³RERR²³⁶ site. Cleavage at ²⁰⁷RKPR²¹⁰ rather than ²³³RERR²³⁶ will remove the nuclear localization signal for DHBV core protein (Mabit et al., 2001). If so parrot hepatitis B virus can produce e antigen despite absence of the ²³³RERR²³⁶ site. The ²⁰⁷RKPR²¹⁰ and ²³³RERR²³⁶ sites are 52aa and 26aa away from the C-terminus of the precore/core protein, respectively. A previous study found that DHBV e antigen production was prevented by C-terminal deletion of 89aa but not 36aa or 12aa from the precore/core gene construct (Schlicht, 1991). That finding is consistent with mature e antigen being cleavage product of ²⁰⁷RKPR²¹⁰ rather than ²³³RERR²³⁶.

7. Herpetohepadnaviruses and most metahepadnaviruses cannot produce e antigen.

As shown in Fig. 8A, most herpetohepadnaviruses and metahepadnaviruses do not have an open precore region required for e antigen expression, due to absence of a methionine codon or presence of a stop mutation(s). Only white sucker hepadnavirus has an open precore region. Although the 21-aa precore peptide of WSHBV is predicted to be cleaved by a signal peptidase, the only cysteine residue would be clipped off preventing formation of intramolecular disulfide bond. As for the C-terminal furin cleavage site, some herpetohepadnaviruses and metahepadnaviruses have at least one RXXR motif (Fig. 8B). But in the absence of a signal peptide encoded by the precore region e antigen expression would be impossible. On the other hand, this finding suggests that emergence of a precore region with both N-terminal signal peptide sequence and C-terminal cysteine residue is the critical final step in gaining the ability to produce e antigen.

8. Shrew hepatitis B virus: the primitive orthohepadnavirus incapable of e antigen expression.

Very recently, musk shrew hepatitis B virus (MSHBV) was found in China and Ivory Coast, respectively, while crowned shrew hepatitis B virus (CSHBV) was identified in Germany (Nie et al., 2019; Rasche et al., 2019). These viruses are unable to express the e antigen due to one or more stop codons upstream of the core gene (Fig. 8A). In fact the shared nonsense mutation (just one codon apart from the core gene ATG) is equivalent to the G1896A mutation abolishing HBeAg production (Carman et al., 1989; Okamoto et al., 1990; Tong et al., 1990), and the predicted amino acid sequence nearby is similar to that of HBV. This raised possibility that the virus detected represented e antigen-negative mutant, and that the wild-type virus could be detected if liver samples were collected at an earlier time point. Indeed all the three types of SHBV contained a conserved RXXR motif at the C-terminus of precore/core protein required for e antigen maturation (Fig. 6A). However, deep sequencing of liver samples failed to find wild-type precore sequence. Moreover, neither CSHBV nor the Chinese variant of MSHBV has a methionine codon for translation initiation. While the Ivory Coast variant of MSHBV does have the methionine codon, the predicted translation product does not conform to a signal peptide (Rasche et al., 2019). Thus, SHBV most likely has not yet evolved capacity to express the e antigen, and the shared nonsense mutation can

be explained by the base-pairing constraint of the overlapping pregenome encapsidation signal as already demonstrated for HBV (Li et al., 1993; Rasche et al., 2019; Tong et al., 1993).

The Ivory Coast variant of MSHBV cannot express M protein because the putative preS2 domain is too long (104aa) (Table 1; Fig. S2E). Although CSHBV and the Chinese variant of MSHBV have a shorter preS2 domain of 65aa and 53aa, respectively, no N-linked glycosylation site could be found (Fig. 5). Moreover, they have a T or C at the -3 position of the putative preS2 ATG codon, in contrast to A or G for hepadnaviruses known to express M protein (Table 1). Therefore, SHBV most likely does not express M protein.

9. Overall conclusions.

For HBV, M protein and e antigen are byproducts of the SPII promoter/2.1-kb RNA for S protein and core promoter/3.5-kb RNA for core protein, respectively. The same principle most likely can be applied to other hepadnaviruses. This group of hepatotropic DNA viruses are derived from nackdnviruses expressing just core and P proteins essential for genome replication (Fig. 3) (Lauber et al., 2017). Herpetohepadnavirus and metahepadnavirus do not have the X gene and most cannot express the e antigen due to a lack of coding capacity upstream of core gene (Lauber et al., 2017). According to this study, most of these viruses probably do not express M protein and two members even do not have myristoylation signal in the L protein. Thus, they represent immediate descendants of nackdnvirus by acquisition of L and S protein expression (Fig. 3). All orthohepadnaviruses have evolved ability to express the X protein, and most can express M protein. Most orthohepadnaviruses (with the exception of SHBV) and all avihepadnaviruses could express the e antigen, which requires cleavable signal peptide and a precore-derived cysteine residue, as well as C-terminal cleavage by furin or another basic endopeptidase. Ability to express the e antigen most likely promotes chronic infection (especially during perinatal transmission) and at the same time reduces liver pathogenesis. Indeed, SHBV infection was found quite sporadic and restricted, with high viral load in the liver (Rasche et al., 2019). The evolution of the X gene and e antigen does not necessarily occur in a sequential manner, with avihepadnaviruses capable of e antigen but not X protein expression, while SHBV capable of X protein but not e antigen production. On the other hand, M protein expression is most likely the latest addition to hepadnavirus, with even some orthohepadnaviruses (SHBV and cat hepadnavirus) probably incapable of its expression. L protein myristoylation is a nearly universal feature of hepadnaviruses, which for HBV is required for receptor binding. Whether the S protein of BHBV, HHBV, and some other hepadnaviruses is myristoylated and has functional consequences requires further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

This work was supported by NIH grants R21AI142456, R01AI116639, the National Natural Science Foundation of China (81672064), and the Major Science and Technology Special Project of China (2017ZX10202203-003-002). We are grateful to Dr. Stefan Seitz for sharing sequence information.

Abbreviations:

HBV	hepatitis B virus
AMDV	Astatotilapia metahepadnavirus
ASHBV	Ashy-headed Sheldgoose hepatitis B virus
BGHBV	bluegill hepatitis B virus
BHBV	bat hepatitis B virus
CHBV	crane hepatitis B virus
CMHBV	capuchin monkey hepatitis B virus
CSHBV	crowned shrew hepatitis B virus
CWHBV	Chiloe wigeon hepatitis B virus
DCHBV	domestic cat hepadnavirus
DHBV	duck hepatitis B virus
GSHV	ground squirrel hepatitis virus
HBHBV	horseshoe bat hepatitis B virus
HHBV	heron hepatitis B virus
LBHBV	long-fingered bat hepatitis B virus
MSHBV	musk shrew hepatitis B virus
OSHBV	Orinoco shieldgoose hepatitis B virus
PHBV	parrot hepatitis B virus
RBHBV	roundleaf bat hepatitis B virus
RGHBV	Ross goose hepatitis B virus
SGHBV	snow goose hepatitis B virus
SHBV	shrew hepatitis B virus
SkHBV	skink hepatitis B virus
SLHBV	spiny lizard hepatitis B virus
STHBV	stork hepatitis B virus
TBHBV	tent-making bat hepatitis B virus
TFHBV	Tibetan frog hepatitis B virus
TMDV	tetra metahepadnavirus

WHV	woodchuck hepatitis virus
WMHBV	woolly monkey hepatitis B virus
WSHBV	white sucker hepadnavirus
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
L, M, S proteins	large, middle, and small envelope proteins
NTCP	sodium taurocholate cotransporting polypeptide
P	polymerase
SVP	subviral particles

REFERENCES

- Aghazadeh M, Shi M, Barrs V, McLuckie A, Lindsay S, Jameson B, Hampson B, Holmes EC, Beatty JA, 2018 A novel hepadnavirus identified in an immunocompromised domestic cat in Australia. *Viruses* 10 (5), E269. [PubMed: 29772771]
- Bruss V, Hagelstein J, Gerhardt E, Galle PR, 1996 Myristoylation of the large surface protein is required for hepatitis B virus in vitro infectivity. *Virology* 218 (2), 396–399. [PubMed: 8610467]
- Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarvey MJ, Makris A, Thomas HC, 1989 Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* 2 (8663), 588–591. [PubMed: 2570285]
- Chang C, Enders G, Sprengel R, Peters N, Varmus HE, Ganem D, 1987 Expression of the precore region of an avian hepatitis B virus is not required for viral replication. *J. Virol* 61 (10), 3322–3325. [PubMed: 3041052]
- Chang SF, Netter HJ, Bruns M, Schneider R, Frölich K, Will H, 1999 A new avian hepadnavirus infecting snow geese (*Anser caerulescens*) produces a significant fraction of virions containing single-stranded DNA. *Virology* 262 (1), 39–54. [PubMed: 10489339]
- Chen H, Kew M, Hornbuckle W, Tennant B, Cote P, Gerin J, Purcell R, Miller R, 1992 The precore gene of the woodchuck hepatitis virus genome is not essential for viral replication in the natural host. *J. Virol* 66 (9), 5682–5684. [PubMed: 1501300]
- Cote PJ, Korba BE, Miller RH, Jacob JR, Baldwin BH, Hornbuckle WE, Purcell RH, Tennant BC, Gerin JL, 2000 Effects of age and viral determinants on chronicity as an outcome of experimental woodchuck hepatitis virus infection. *Hepatology* 31 (1), 190–200. [PubMed: 10613745]
- de Carvalho Dominguez Souza BF, König A, Rasche A, de Oliveira Carneiro I, Stephan N, Corman VM, Roppert PL, Goldmann N, Kepper R, Müller SF, Völker C, de Souza AJS, Gomes-Gouvêa MS, Moreira-Soto A, Stöcker A, Nassal M, Franke CR, Rebello Pinho JR, Soares M, Geyer J, Lemey P, Drosten C, Netto EM, Glebe D, Drexler JF, 2018 A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. *J. Hepatol* 68 (6), 1114–1122. [PubMed: 29428874]
- Dill J, Camus A, Leary J, Di Giallonoardo F, Holmes E, Ng T, 2016 Distinct viral lineages from fish and amphibians reveal the complex evolutionary history of hepadnaviruses. *J. Virol* 90 (17), 7920–7933. [PubMed: 27334580]
- Drexler J, Geipel A, König A, Corman V, van Riel D, Leijten LM, Bremer CM, Rasche A, Cottontail VM, Maganga GD, Schlegel M, Müller MA, Adam A, Klose SM, Carneiro AJ, Stöcker A, Franke CR, Gloza-Rausch F, Geyer J, Annan A, Adu-Sarkodie Y, Oppong S, Binger T, Vallo P, Tschapka M, Ulrich RG, Gerlich WH, Leroy E, Kuiken T, Glebe D, Drosten C, 2013 Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc. Natl. Acad. Sci. USA* 110 (40), 16151–16156. [PubMed: 24043818]

- Fernholz D, Galle P, Stemler M, Brunetto M, Bonino F, Will H, 1993 Infectious hepatitis B virus variant defective in pre-S2 protein expression in a chronic carrier. *Virology* 194 (1), 137–148. [PubMed: 8480417]
- Ganem D, Prince AM, 2004 Hepatitis B virus infection--natural history and clinical consequences. *N. Engl. J. Med* 350 (11), 1118–1129. [PubMed: 15014185]
- Garcia P, Ou J, Rutter W, Walter P, 1988 Targeting of the hepatitis B virus precore protein to the endoplasmic reticulum membrane: After signal peptide cleavage translocation can be aborted and the product released into the cytoplasm. *J. Cell Biol* 106 (4), 1093–1104. [PubMed: 3283145]
- Garcia T, Li J, Sureau C, Ito K, Qin Y, Wands J, Tong S, 2009 Drastic reduction in the production of subviral particles does not impair hepatitis B virus virion secretion. *J. Virol* 83 (21), 11152–11165. [PubMed: 19706705]
- Gripon P, Le Seyec J, Rumin S, Guguen-Guillouzo C, 1995 Myristoylation of the hepatitis B virus large surface protein is essential for viral infectivity. *Virology* 213 (2), 292–299. [PubMed: 7491754]
- Guo H, Mason WS, Aldrich CE, Saputelli JR, Miller DS, Jilbert AR, Newbold JE, 2005 Identification and characterization of avihepadnaviruses isolated from exotic anseriformes maintained in captivity. *J. Virol* 79 (5), 2729–2742. [PubMed: 15708992]
- Hahn C, Iwanowicz L, Cornman R, Conway C, Winton J, Blazer V, 2015 Characterization of a novel hepadnavirus in the white sucker (*Catostomus commersonii*) from the Great Lakes Region of the United States. *J. Virol* 89 (23), 11801–11811. [PubMed: 26378165]
- He B, Fan Q, Yang F, Hu T, Qiu W, Feng Y, Li Z, Li Y, Zhang F, Guo H, Zou X, Tu C, 2013 Hepatitis virus in long-fingered bats, Myanmar. *Emerging Infect. Dis* 19 (4), 638–640. [PubMed: 23631923]
- He B, Zhang F, Xia L, Hu T, Chen G, Qiu W, Fan Q, Feng Y, Guo H, Tu C, 2015 Identification of a novel orthohepadnavirus in Pomona roundleaf bats in China. *Arch. Virol* 160 (1), 335–337. [PubMed: 25193071]
- Ito K, Kim K-H, Lok A, Tong S, 2009 Characterization of genotype-specific carboxyl-terminal cleavage sites of hepatitis B virus e antigen precursor and identification of furin as the candidate enzyme. *J. Virol* 83 (8), 3509–3517.
- Jones SA, Hu J, 2013 Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. *Emerg. Microbes Infect* 2 (9), e56. [PubMed: 26038488]
- Lanave G, Capozza P, Diakoudi G, Catella C, Catucci L, Ghergo P, Stasi F, Barrs V, Beatty J, Decaro N, Buonavoglia C, Martella V, Camero M, 2019 Identification of hepadnavirus in the sera of cats. *Sci. Rep* 9 (1), 10668. [PubMed: 31337847]
- Lanford RE, Chavez D, Brasky KM, Burns RB 3rd, Rico-Hesse R, 1998 Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc. Natl. Acad. Sci. USA* 95 (10), 5757–5761. [PubMed: 9576957]
- Lauber C, Seitz S, Mattei S, Suh A, Beck J, Herstein J, Borold J, Salzburger W, Kaderali L, Briggs J, Bartenschlager R, 2017 Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. *Cell Host & Microbe* 22 (3), 387–399. [PubMed: 28867387]
- Li J, Tong S, 2015 From DCPD to NTCP: the long journey towards identifying a functional hepatitis B virus receptor. *Clin. Mol. Hepatol* 21 (3), 193–199. [PubMed: 26523264]
- Li JS, Tong S, Wen YM, Vitvitski L, Zhang Q, Trepo C, 1993 Hepatitis B virus genotype A rarely circulates as HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *J. Virol* 67 (9), 5402–5410. [PubMed: 8350403]
- Mabit H, Breiner K, Knaust A, Zachmann-Brand B, Schaller H, 2001 Signals for bidirectional nucleocytoplasmic transport in the duck hepatitis B virus capsid protein. *J. Virol* 75, 1968–1977. [PubMed: 11160696]
- Macrae DR, Bruss V, Ganem D, 1991 Myristylation of a duck hepatitis B virus envelope protein is essential for infectivity but not for virus assembly. *Virology* 181 (1), 359–363. [PubMed: 1994583]
- Magnius LO, Espmark JA, 1972 New specificities in Australia antigen positive sera distinct from the Le Bouvier determinants. *J. Immunol* 109 (5), 1017–1021. [PubMed: 4116763]
- Marion PL, Oshiro LS, Regnery DC, Scullard GH, Robinson WS, 1980 A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. *P Natl Acad Sci USA* 77 (5), 2941–2945.

- Mason WS, Seal G, Summers J, 1980 Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *J. Virol* 36 (3), 829–836. [PubMed: 7463557]
- Messageot F, Salhi S, Eon P, Rossignol J-M, 2003 Proteolytic processing of the hepatitis B virus e antigen precursor. Cleavage at two furin consensus sequences. *J. Biol. Chem* 278 (2), 891–895. [PubMed: 12417589]
- Milich D, Jones J, Hughes J, Price J, Raney A, McLachlan A, 1990 Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance in utero? *Proc Natl Acad Sci USA* 87 (17), 6599–6603. [PubMed: 2395863]
- Milich D, Liang TJ, 2003 Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. *Hepatology* 38 (5), 1075–1086. [PubMed: 14578844]
- Nassal M, Rieger A, 1993 An intramolecular disulfide bridge between Cys-7 and Cys61 determines the structure of the secretory core gene product (e antigen) of hepatitis B virus. *J. Virol* 67 (7), 4307–4315. [PubMed: 8510224]
- Ni Y, Sonnabend J, Seitz S, Urban S, 2010 The pre-s2 domain of the hepatitis B virus is dispensable for infectivity but serves a spacer function for L-protein-connected virus assembly. *J. Virol* 84 (8), 3879–3888. [PubMed: 20130049]
- Nie FY, Lin XD, Hao ZY, Chen XN, Wang ZX, Wang MR, Wu J, Wang HW, Zhao G, Ma RZ, Holmes EC, Zhang YZ, 2018 Extensive diversity and evolution of hepadnaviruses in bats in China. *Virology* 514, 88–97. [PubMed: 29153861]
- Nie FY, Tian JH, Lin XD, Yu B, Xing JG, Cai JH, Holmes EC, Ma RZ, Zhang YZ, 2019 Discovery of a highly divergent hepadnavirus in shrews from China. *Virology* 531, 162–170. [PubMed: 30884426]
- Okamoto H, Yotsumoto S, Akahane Y, Yamanaka T, Miyazaki Y, Sugai Y, Tsuda F, Tanaka T, Miyakawa Y, Mayumi M, 1990 Hepatitis B viruses with precore region defects prevail in persistently infected hosts along with seroconversion to the antibody against e antigen. *J. Virol* 64 (3), 1298–1303. [PubMed: 2304145]
- Piasecki T, Kurenbach B, Chrzastek K, Bednarek K, Kraberger S, Martin DP, Varsani A, 2012 Molecular characterisation of an avihepadnavirus isolated from *Psittacula krameri* (ring-necked parrot). *Arch Virol* 157 (3), 585–590. [PubMed: 22183110]
- Pult I, Netter HJ, Bruns M, Prassolov A, Sirma H, Hohenberg H, Chang SF, Frolich K, Krone O, Kaleta EF, Will H, 2001 Identification and analysis of a new hepadnavirus in white storks. *Virology* 289 (1), 114–128. [PubMed: 11601923]
- Rasche A, Lehmann F, König A, Goldmann N, Corman VM, Moreira-Soto A, Geipel A, van Riel D, Vakulenko YA, Sander AL, Niekamp H, Kepper R, Schlegel M, Akoua-Koffi C, Souza BFCD, Sahr F, Olayemi A, Schulze V, Petraityte-Burneikiene R, Kazaks A, Lowjaga KAAT, Geyer J, Kuiken T, Drosten C, Lukashev AN, Fichet-Calvet E, Ulrich RG, Glebe D, Drexler JF, 2019 Highly diversified shrew hepatitis B viruses corroborate ancient origins and divergent infection patterns of mammalian hepadnaviruses. *Proc. Natl. Acad. Sci. USA* 116 (34), 17007–17012. [PubMed: 31371507]
- Salfeld J, Pfaff E, Moah M, Schaller H, 1989 Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. *J. Virol* 63 (2), 798–808. [PubMed: 2463383]
- Schlicht HJ, 1991 Biosynthesis of the secretory core protein of duck hepatitis B virus: intracellular transport, proteolytic processing, and membrane expression of the precore protein. *J. Virol* 65 (7), 3489–3495. [PubMed: 2041077]
- Schlicht HJ, Salfeld J, Schaller H, 1987 The duck hepatitis B virus pre-C region encodes a signal sequence which is essential for synthesis and secretion of processed core proteins but not for virus formation. *J. Virol* 61 (12), 3701–3709. [PubMed: 3682059]
- Seeger C, Mason W, 2015 Molecular biology of hepatitis B virus infection. *Virology* 479–480: 672–686.
- Sprengel R, Kaleta EF, Will H, 1988 Isolation and characterization of a hepatitis B virus endemic in herons. *J. Virol* 62 (10), 3832–3839. [PubMed: 3418788]
- Standing D, Ou J, Masiarz F, Rutter W, 1988 A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in xenopus oocytes. *Proc. Natl. Acad. Sci. USA* 85 (22), 8405–8409. [PubMed: 3186731]

- Summers J, Mason W, 1982 Replication of the genome of a hepatitis B-like virus by reverse transcription. *Cell* 29 (2), 403–415. [PubMed: 6180831]
- Summers J, Smolec JM, Snyder R, 1978 A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proc. Natl. Acad. Sci. USA* 75 (9), 4533–4537. [PubMed: 212758]
- Takahashi K, Machida A, Funatsu G, Nomura M, Usuda S, Aoyagi S, Tachibana K, Miyamoto H, Imai M, Nakamura T, Miyakawa Y, Mayumi M, 1983 Immunochemical structure of hepatitis B e antigen in the serum. *J. Immunol* 130 (6), 2903–2907. [PubMed: 6189903]
- Tong SP, Brotman B, Li JS, Vitvitski L, Pascal D, Prince AM, Trepo C, 1991a In vitro and in vivo replication capacity of the precore region defective hepatitis B virus variants. *J. Hepatol* 13 Suppl 4, S68–73. [PubMed: 1668333]
- Tong SP, Diot C, Gripon P, Li J, Vitvitski L, Trepo C, Guguen-Guillouzo C, 1991b In vitro replication competence of a cloned hepatitis B virus variant with a nonsense mutation in the distal pre-C region. *Virology* 181(2), 733–737. [PubMed: 2014646]
- Tong SP, Li JS, Vitvitski L, Kay A, Trepo C, 1993 Evidence for a base-paired region of hepatitis B virus pregenome encapsidation signal which influences the patterns of precore mutations abolishing HBe protein expression. *J. Virol* 67 (9), 5651–5655. [PubMed: 8350418]
- Tong SP, Li JS, Vitvitski L, Trepo C, 1990 Active hepatitis B virus replication in the presence of anti-HBe is associated with viral variants containing an inactive pre-C region. *Virology* 176 (2), 596–603. [PubMed: 2345966]
- Wang B, Yang XL, Li W, Zhu Y, Ge XY, Zhang LB, Zhang YZ, Bock CT, Shi ZL, 2017 Detection and genome characterization of four novel bat hepadnaviruses and a hepevirus in China. *Virol. J* 14 (1), 40. [PubMed: 28222808]
- Wasenauer G, Kock J, Schlicht HJ, 1993 Relevance of cysteine residues for biosynthesis and antigenicity of human hepatitis B virus e protein. *J Virol* 67 (3), 1315–1321. [PubMed: 8437217]
- Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z, Huang Y, Qi Y, Peng B, Wang H, Fu L, Song M, Chen P, Gao W, Ren B, Sun Y, Cai T, Feng X, Sui J, Li W, 2012 Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife* 1, e00049. [PubMed: 23150796]
- Yeh CT, Liaw YF, Ou JH, 1990 The arginine-rich domain of hepatitis B virus precore and core proteins contains a signal for nuclear transport. *J. Virol* 64 (12), 6141–6147. [PubMed: 2243390]

Highlights:

- Bat hepatitis B virus has myristoylation signal in its small envelope protein
- Most herpeto- and meta-hepadnaviruses probably express no middle envelope protein
- e antigen from both ortho- and avi-hepadnaviruses has a precore-derived cysteine
- Furin is probably the shared host enzyme for hepadnaviral e antigen maturation
- Shrew HBV plus most herpeto- and meta-hepadnaviruses cannot produce e antigen

Figure 1. Genetic organization of the HBV genome and expression of the 7 viral proteins.

(A) The HBV genome, genes, and transcripts. Four open reading frames (ORFs) cover the entire 3.2-kb circular HBV genome. The largest ORF, P (polymerase), overlaps with the entire envelope gene, the 3' end of the core (C) ORF, and the 5' end of the X ORF. The core gene is preceded by the precore (pre-C) region, while the envelope gene is subdivided into preS1, preS2, and S regions. Since each region starts with an ATG codon, such arrangement enables expression of two proteins from precore/core gene (HBeAg and core protein) and three proteins from envelope gene (L, M, S proteins). (B) Co-expression of HBeAg and core protein from precore/core gene is achieved by heterogeneity of the 3.5-kb RNA: the longer version for HBeAg and the shorter version for core protein e (see Fig. 2). Similarly, expression of both M and S proteins is achieved by heterogeneity of the 2.1-kb RNA: the longer version (with the entire preS2 region covered at its 5' end) for M protein and the shorter version (lacking preS2 AUG codon at its 5' end) for S protein.

Figure 2. Alternative translation initiation from HBV envelope and core genes generates multiple proteins.

(A) L, M, and S proteins as alternative translation products from AUG codons in the preS1, preS2, and S regions of the envelope gene, respectively. They are co-terminal, with M protein having extra preS2 domain than S protein, and L protein having extra preS1 domain than M protein. (B) HBeAg and core gene as alternative translation products from the precore/core gene. The core promoter drives transcription of 3.5-kb mRNA with heterogeneous 5' end. The longer one (pcRNA) has the entire precore region covered at its 5' end and is hence used for translation of fused precore/core protein (p25). P25 is cleaved by a signal peptidase to generate p22, which is further cleaved by furin or related proprotein convertase (PC) in the secretory pathway to form p17, the mature HBeAg released to bloodstream. The shorter pgRNA is used for core protein translation.

Figure 3. Possible evolutionary history of hepadnaviruses.

Metahepadnaviruses and herpetohepadnaviruses possibly originated from nakednavirus by insertion of the preS region together with coding capacity in the S region, thus enabling L and S protein expression to become enveloped hepatotropic virus. Coding capacity in the precore region enabled e antigen production, which probably promotes viral persistence in birds (for avihepadnavirus) and mammals (for orthohepadnavirus) following perinatal transmission. Most orthohepadnaviruses further gained X gene and can express M protein. An exception is shrew hepatitis B virus, which is e-minus and most likely M-minus. The e-minus and M-minus mutants are frequently selected at the immune clearance phase of chronic HBV infection. It is uncertain whether shrew hepatitis B virus represents the immediate ancestor to other orthohepadnavirus.

BHBV	KC790378	<u>MALPAAGLLGPLAVLQVG</u> YFLWTKIQEIGKTVDWWISLSSPGGPTGCGGQNSLSPTSNNHSPISCPVTC
	KF939648	<u>MGITSSSELLAPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFLGGTPECNGLNLLSPTCKHSPTSCPATC
	JX941466	<u>MGITTSGLVPLAGLQVGC</u> FLLTILEIGKSLDWWWTSLSFPGGIPVCIGQNSQSQTSSHSPTSCPPTC
	KY962705	<u>MGNISSELLAPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFLGGTPECNGLNLLSPICKHSPTSCPATC
	NC_024444	<u>MGNISSELLAPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFPGGIPECFGLSLQSQTCCKHSPTYCPPTC
	KX513949	<u>MGNISSELLGPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFLGGTPECNGLGLQYPTCKHSPTSCPPTC
	KC790373	<u>MGNISSELLGPLVGFQVV</u> FFLWTKILTIAQSLDSWWTSLSFPGGIPECFGLNLQSQTCCKHSPTSCPPTC
	KY962693	<u>MGNISSELLAPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFLGGTPECNGLSLLSPICKHSPTSCPATC
	KY962686	<u>MGNISSELLEPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFLGGTPECNGQNLQFQTCRHSPTSCPATC
	KY962701	<u>MGNISSELLAPLVGFQVV</u> FFLWTKILTIGQSLDSWWTSLSFLGGTPECNGLSLLSPICKHSPTSCPATC
	DCHBV	<u>MGNITSGPFGFLLGLQVG</u> SFLWTKIQTIGQSADSWWTSLSFPGAIPGCIGQDSQYQTSKHSPTSCPPTC
	HHBV	<u>MGATFGGILAGLIGLLVG</u> FFLLTKILEILRKLDWWISLSSPKKMLCAFQNTGAQTSPHYVGSCPWGC
	STHBV	<u>MGASFGGILAGLIGLLVG</u> FFLLTKILEILRKLDWWISLSSPKKMLCAFQSTGHQTSPHYVGSCPWGC
	AMDV	<u>MGSTLSNAGPGFLGGLPV</u> VSFLWTKIQETLERLDLWWTSLSSPDCLLGSCGPGHLRQILGHSPSCCPQE
	WSHBV	<u>MGSFAQGLSFDLWVFLV</u> VLLVVFLLIKIHTILKAADWWLISRSLGTQTCFRTNTEPTSMHFKTDC

Figure 4. Presence of myristoylation signal in the S protein of most BHBV isolates and some other hepadnaviruses.

Shown are N-terminal 69 residues of the S protein, with MG dipeptide at the N-terminus underlined. From top are 11 orthohepadnaviruses, two avihepadnaviruses, and two metahepadnaviruses. BHBV (bat hepatitis B viruses) are identified by their GenBank accession numbers. DCHBV: domestic cat hepatitis B virus; HHBV: heron hepatitis B virus; STHBV: stork hepatitis B virus; AMDV: Astatotilapia metahepadnavirus; WSHBV: white sucker hepadnavirus.

		Length (aa)
HBV	MQ WNSTTFHQTLDPRVRGLYFPAGGSSSGTVNPVLTASPLSSIFSRIGDPALN	55
WMHBV	MT WNTSSSFQSYLQNPVKVRGLYFPAGGSTSSIVNPVPTTASTSSSFSTTGVPVST	55
CMHBV	MT WNISNFQKALQDPRVRGLYFPAGGSSSGIPTAPTASTISSLFATTGDPVEN	55
WHV	MKNQ TFRLQGFVDGLRDLTTERYHNAYGDPFTTLLSPVPTVSTILSPSTTGDPALSPE	60
GSHV	MKNQ TGHLQGFAEGLRALTTSDHNSAYGDPFTTLLSPVPTVSTILSPPLTIGDPVLSTE	60
BHBV	KY962705 MN STWRQRYSSLLNPVPGSTTATHQPAASTASSTASFSTGDPDPAID	47
	NC_024444 MS NQWRERYGHLISALGSSSGTTPADPTASDSAPFFWTGDPAPIS	47
	KX513949 MN TTWRQRYGHLSPAGSSSTNLPPAASTASSAPSFSTGDPVID	47
	KC790373 MT NQWRERYGHLSPAGPTSAPIQPADPTVSSAPFSWTGDLALD	47
	KF93964 MN KTWQRYASLITPVGSSSTILRPAEDTASDSVFYYSPIGDPDPAID	47
	KY962701 MN STWRQRYSSLLNPVPGSTTATHQPAASTASSTASFSTGDPDPAID	47
	KC790378 MN ITALKENPEVQKLFAGGFPSFLPLPVETTASDTSSPTGTGTPGAPAH	47
KY962686	MT TWREKYASLLTPVPGSSSSTPTAGNTASRTVSSSWKTGESVTG	46
JX941466	MP SWREKYGHLITAGGSSSKLQIPAVTAASVTSSASSTIGDPDPAID	45
DCHBV	MS PSSWRQKYGYLLLPKTTAAPGPTSPPTTAVAPTASSTSSPPLTIGAPAA	52
MSHBV (CN)	ML RPDHHTPLSKIEIKTPPPVPSSASKLLVPVENTASGPSLEKTETGDPVLET	53
CSHBV	MP SDHPVHAALSQFDEELRESERERSLVGTAGAPSDYFPSRLPAGNIASAPSSVSEPTGDPVLVN	65
WSHBV	MQ IEQPLQSPPLASKVSPLSPIKAVPFLAPLLIPVAHQTVRTSYNMTTSNISEIMDEARTATT	64
SLHBV	MR ERRRIPAVRKKQVPVRKPTRPTPATPTFPGTPTPQPDTLNDSTEAPTEAPQVPVKTTETQPESQN	67
SKHBV	MN NQPTREHQRLQEAIRSNQSDPDNGTWWVTTVEGEVISVPPFPVIPSFPSATLAPSTKPDAPLATRPPPTQK	74

Figure 5. The preS2 domain of some hepadnaviruses and the potential to express M protein.

Shown are amino acid sequences of the preS2 domain or putative preS2 domain, with the methionine residues in boldface and N-linked glycosylation sites underlined. The top 12 entries are mammalian hepadnaviruses capable of M protein expression (top 5) or expected to (next 7). They have a short preS2 domain of 47–60aa with an N-linked glycosylation motif (NXS/T) near the N-terminus. The next 5 entries are orthohepadnaviruses lacking the glycosylation site in the putative preS2 domain. At the bottom are a metahepadnavirus (WSHBV) and two herpetohepadnavirus (SLHBV and SKHBV). Their glycosylation site is not located near the N-terminus, and their preS2 domains are a little long. WMHBV: woolly monkey hepatitis B virus; CMHBV: Capuchin monkey hepatitis B virus; WHV: woodchuck hepatitis virus; GSHV: ground squirrel hepatitis virus; BHBV: bat hepatitis B virus (each one identified by its GenBank accession number); DCHBV: domestic cat hepadnavirus; MSHBV CN: musk shrew hepatitis B virus (Chinese variant); CSHBV: crowned shrew hepatitis B virus; WSHBV: white sucker hepadnavirus; SLHBV: spiny lizard hepatitis B virus; SKHBV, skink hepatitis B virus.

HBV	MQLFHLCLIISC	SCPTVQASKL	CLGWLWG
WMHBV	-H-----L-	-----	-----L-
CMHBV	-H-----L-	-----	Q-----L-
KC790373	-Y-----F-	-Y-----	-----D
KY962693	-Y-----LF-	-----L-	-----
KX513949	-Y-----LF-	-Y-----	-----
KF939648	-Y-----LF-	-Y-----	Q-----
DCHBV	-H-----VLF-	-I-----	-----R-----
NC_024444	-Y-----VF-	-Y--L-	-----D
WHV	-Y-----VFA-VP-	-F-----	-----
GSHV	-Y-----VFA-VP-	-----	-----D
KC790380	MRL--YLFH-CLFLC-S--F-	PE-	-R-----
DHBV	MWNLRLITPLSFGAACQGIFTSTLLLS	CVTVPLVCTIVYDS	CLY
SGHBV	-----	L-----	-----T---
ASHBV	-----	FF-----	-----F---
HHBV	--S--LH-SP-----	S-FL-----	-----
CHBV	--S--LH-SP-----	FA-----	-----W
STHBV	--S--LH-SP-----	S-FL-----	-----T---
RGHBV	--D--LH-SP-----	S-FL-----	-----C
PHBV	--R--LH-SP--V-----	C--AL--S-----	-----Y---

		site 1	site 2	site 3	site 4
HBVgtnonA	WIRTPPAYRPPNAPILSTLPETT TVVR		<u>RRGRSPRRRTTPSPRRRRSQSPRRRRSQSRESQC</u>		
WHV	WIRTPAPYRPPNAPILSTLPETT TVIRRRGGA		<u>RASRSPRRRTTPSPRRRRSQSPRRRRSQSPSANC</u>		
GSHV	WIRTPAPYRPPNAPILSTLPETT TVIRRRGGS		<u>RAARSPRRRTTPSPRRRRSQSPRRRRSQSPASNC</u>		
DCHBV	WIRTPAPYRPPNAPILSTLPETT TVIRARGAARRPARSPRRRTTPSPRRRRSQSPRRRRSQSPTQSN		<u>C</u>		
NC_024444	WMRTPAAYRPPNAPILSTLPETT TVIRRRGNP		<u>RAPRSPRRRTTPSPRRRRSQSPRRRRSQSPAPSNC</u>		
HBVgtA	WIRTPPAYRPPNAPILSTLPETT TVVR	<u>RRDRG</u>	<u>RSPRRRTTPSPRRRRSQSPRRRRSQSRESQC</u>		
KF939648	WIRTPQPYRPPNAPILSTLPETT TVRRRGGR		<u>AATRSPRRRTTPSPRRRRSKSPRRRRTPSPAPAC</u>		
KY962705	WIRTPQPYRPPNAPILSTLPETT TVIRRRGGR		<u>AATRSPRRRTTPSPRRRRSKSPRRRRTPSPAPSSC</u>		
KY962693	WIRTPQPYRPPNAPILSTLPETT TVIRRRGGR		<u>AATRSPRRRTTPSPRRRRSKSPRRRRTPSPAPSSC</u>		
KY962686	WIRTPQAYRPPNAPILSTLPETT TVIRGRGR		<u>ATSRPPRGRTTPSPRRRRSQSPRRRRSSSPASTNC</u>		
WMHBV	WIRTPPAYRPPNAPILSTLPETT TV	<u>RRRR</u>	<u>PSGRRTTPSPRRRRSQSPRRRRSQSPASSC</u>		
CMHBV	WIRTPPAYRPLNAPILSTLPETT TV	<u>RRRP</u>	<u>GSSRGRTTPSPRRRRSQSPRRRRSQSPASSC</u>		
KC790380	WIRTPAAYRPPNAPIPVNAPIL TTL PETSVIRR		<u>RPASRRSTTPSPRRRRSQSPRRRRSPSPRPASNC</u>		
MSHBV (CN)	WIRTPAPYRPPNAPIL TTL PETCHIRGR TPR		<u>GTSPAPRRRRSQSPGKRRSPSPRGK</u>		
MSHBV (IC)	WIRTPPIYRPPNAPIL TTL PETCHLRGR APR		<u>GSSPAPSRRRSKSPGKRRSPSPARK</u>		
CSHBV	WIRTPAAYRPPNAPIL TTL PETCHLRGR TPR		<u>GSSPAPSRRRSKSPGKRRSPSPAKK</u>		
DHBV	RKTTTGT RKPR GLEP RRRKVKTTVVY GRRRSKSRERRAPT Q RAGSPLPRSSSSSHRSPSPRK				
HHBV	RNQTKGT RKPR GLEP RRRKVKTTVVY GRRRSKSRGRRSSPSQAGSPLPRNRGNQTRSPSPRE				
SGHBV	RKTSTAT RKPR GLEP RRRKVKTTVVY GRRRSKSRERRASSPQAGSPLPRSSSSSHRSPSPRK				
STHBV	RNQTKG V R KPR GLEP RRRKVKTTVVY GRRRSKSRGRRSSPSQAGSPIPRNRENQSRSSSPRE				
ASHBV	RNKTQGS RKPR GLQ RRRKVKTTVVY GRRRSKSRDRRAPSPQAGSPLPRPSTSHRSPSPRK				
CHBV	RNKTQGV RKPR GLEP RRRKVKTTVVY GRRRSKSRDKRATSPQAGSPIPRTSRDYHRSTSPRQ				
RGHBV	RNKTQGV RKSR GLEP RRRKVKTTIVY GRRRSKSRERRAPT Q RAGSPLPRTSRDHHRSPSPRE				
PHBV	RTKTQGT RKPR GLEP RRRKVKTTVVY GRRRSKSRERSSSPQAGSPIPRTSGGSGRPSTTRR				
		site 1	site 2		

Figure 6. N-terminal and C-terminal sequences of precore/core protein of orthohepadnaviruses vs. avihepadnaviruses.

(A) The precore peptide of 12 orthohepadnaviruses (top) and 8 avihepadnaviruses (bottom). The complete amino acid sequence is shown for HBV and DHBV, respectively. For other viruses only different residues are given, with sequences identical to HBV or DHBV indicated by hyphens. A gap indicates a missing residue. Underlined is the precore peptide remaining on mature e antigen. Boxed is the conserved cysteine residue on the e antigen possibly involved in intramolecular disulfide formation (to block particle formation). (B) The C-terminus of precore/core protein of 16 orthohepadnaviruses (top) and 8 avihepadnaviruses (bottom). The TTVV sequence (or its homolog) is shown in boldface, while potential furin cleavage sites are underlined. Gaps indicate missing residues. HBVgtnonA and HBVgtA represent non-A HBV genotypes and genotype A, respectively. BHBV isolates are identified by their GenBank accession numbers. WHV: woodchuck hepatitis virus; GSHV: ground squirrel hepatitis virus; DCHBV: domestic cat hepadnavirus; WMHBV: woolly monkey hepatitis B virus; CMHBV: Capuchin monkey hepatitis B virus; MSHBV CN and MSHBV IC: the Chinese and Ivory Coast variants of musk shrew hepatitis B virus; CSHBV: crowned shrew hepatitis B virus; DHBV: duck hepatitis B virus; HHBV: heron hepatitis B virus; SGHBV: snow goose hepatitis B virus; STHBV: stork hepatitis B virus; ASHBV: Ashy-headed Sheldgoose hepatitis B virus; CHBV: crane hepatitis B virus; RGHV: Ross goose hepatitis B virus; PHBV: parrot hepatitis B virus.

A	AMDV		*CRFSCICFQPYTTGVPQALHVKAG
	TMDV	*RKHLTSVSGNEGGLPPFRNPPLTGAPPPNCPRGGTKLSWNKQIELVVFSRSV	
	SKHBV		*LFSLL*S
	SLHBV		*AYRTIWSFLWDVLSVCS*LNV
	TFHBV	MEPTPFSIITHLLRSTRSVVITETIVV*LSCACNS	
	BGHBV	MFSY*SFSPYFSRPPGCGSCLCGEK	
	WSHBV	MFAFVLILLRLWSGCLNGSST	
	CSHBV	*IACLRTGCQYFFTFANRCSNPTFQAPKLCLEWL*G	
	MSHBV CN	*HCSCVFCIFFTFAILCSNPTFQAPKLCLEWL*G	
	MSHBV IC	MGAVGINSGTACVLHTFSPLPTYISNPTFQAPKLCLEWL*G	
	HBV PCM	MQLPHLCLIIISCSCPTVQASKLCLEWL*G	
B	AMDV	AFLISMATWLQTPAAYRPPNAPILSHMGGSNPNIYIRPSSRSATARSKSRSPSPRRRSKSPAKR	
	TMDV	VCFITSFVTWLQTPPQYRPPNAPILSCMASTAGNYFGPRATSSAPRPSRSPSPRRRTSPPKR	
	SKHBV	AAFTPVRAKPTTTAVRSSST <u>QRTGRSRRRS</u> RANNYKSQKIYVSTRKSPSPQGRSKSKSPKRH	
	SLHBV	EQVTVPERVTVQTTGPSATPSRSGS <u>RQGRSR</u> KKGNFYRKTVRGFGKSPPRRSKSPNQQKS	
	TFHBV	APLIEEIVEPQRQRTSRPNSRRR <u>ARGG</u> SHPGGDRGLYYQRTVRRFNSP <u>PRERGS</u> RRRSPPAGRT	
	BGHBV	FIVSFATWLQTPPAYRPPNAPILSVMAGGPNIYITRTRGRSPTPSRRRSKSPRRSGSRGSRGK	
	WSHBV	KLRSLWMTPTSYRNQYAPTIEALTRTVKATHLSS <u>RRSR</u> PTARSPGVRARSSSKSKSKSPRRKT	
	CSHBV	YLVSGVWIRTPAAYRPPNAPILTLTPETCHLRG <u>RTPR</u> GSSPAPSRRRSKSPGKRRSPSPAKK	
	MSHBV CN	YLVSGVWIRTPAPYRPPNAPILTLTPETCHIRG <u>RTPR</u> GTSAPAPRRRSQSPGKRRSPSPRGK	
	MSHBV IC	YLVSGVWIRTPIPYRPPNAPILTLTPETCHLRG <u>APRG</u> SSPAPSRRRSKSPGKRRSPSPARK	

Figure 8. Predicted translation product upstream of the core gene (A) and the 3' core gene (B) for metahepadnaviruses, herpetohepadnaviruses, and shrew hepatitis B viruses.

For panel A, a star indicates a stop codon. Only WSHBV has a small open reading frame upstream of core gene. Shrew hepatitis B viruses (CSHBV, MSHBV) are the only orthohepadnavirus having a non-functional precore region. For the core gene translation product, potential furin cleavage sites are underlined. AMDV, Astatotilapia metahepadnavirus; TMDV, tetra metahepadnavirus; SKHBV, skink hepatitis B virus; SLHBV, spiny lizard hepatitis B virus; TFHBV, Tibetan frog hepatitis B virus; BGHBV, bluegill hepatitis B virus; WSHBV, white sucker hepadnavirus. CSHBV: crowned shrew hepatitis B virus; MSHBV: musk shrew hepatitis B virus; CN and IC: China and Ivory Coast. HBV PCM: the G1896A precore mutant of HBV, with the nonsense mutation preventing HBeAg production.

Table 1.

Potential for different groups of hepadnaviruses to express M protein

virus	domain upstream of S	N-linked glycosylation	S domain	Kozak sequence	M protein expression
Orthohepadnaviruses¹					
HBV	55	N4	226	GCCATGC	yes
WMHBV	55	N4	226	GCCATGA	yes
CMHBV	55	N4	226	GCCATGA	yes
WHV	60	N3	222	ACTATGA	yes
GSHV	60	N3	222	ACAATGA	yes
DCHBV	52	no	226	CAAATGT	no? (poor Kozak)
HBHBV (NC_024444)	47	N3	224	AACATGA	yes
RBHBV (KC790373)	47	N3	224	AACATGA	yes
TBHBV (KC790378)	50	N2	223	AAAATGA	yes
LBHBV (JX941466)	45	no	224	AACATGC	yes?
KY962705	47	N2	224	AACATGA	yes
HBHBV (KX513949)	47	N2	223	AATATGA	yes
CSHBV	65	no	225	CTCATGC	no ? (poor Kozak)
MSHBV (China)	53	no	225	TCCATGC	no ? (poor Kozak)
MSHBV (Ivory Coast)	104	yes	225	AACATGA	no (too long preS2)
Meta- and herpetohepadnaviruses²					
AMDV	107	N23	216	GAAATGG	no (too long preS2)
BGHBV	no ATG	NA	212	NA	no (no preS2)
TMDV	96	no	214	CAAATGA	no (too long preS2)
WSHBV	19 or 64?	N5 or N50?	175	GTAATGC	?
SLHBV	67	N42	240	TGCATGA	no? (poor Kozak)
SkHBV	74	N19 + N25	172	CCCATGA	no? (poor Kozak)
TFHBV	163	N43 + N62	171	CGGATGG	no (too long preS2)
Avihepadnaviruses³					
ASHBV	119	ND	167	NA	no
CHBV	99	ND	167	NA	no
CWHBV	110	ND	167	NA	no
DHBV	109	ND	167	NA	no
HHBV	117	ND	169	NA	no
OSHBV	108	ND	167	NA	no
PHBV	122	ND	171	NA	no
RGHBV	102	ND	167	NA	no
SGHBV	110	ND	167	NA	no
STHBV	119	ND	169	NA	no

¹HBV: hepatitis B virus; WMHBV: woolly monkey hepatitis B virus; CMHBV: Capuchin monkey hepatitis B virus; WHV: woodchuck hepatitis virus; GSHV: ground squirrel hepatitis virus; DCHBV: domestic cat hepatitis B virus; HBHBV: horseshoe bat hepatitis B virus; RBHBV: roundleaf bat hepatitis B virus; TBHBV: tent-making bat hepatitis B virus; LBHBV: long-fingered bat hepatitis B virus; CSHBV: crowned shrew hepatitis B virus; MSHBV: Musk shrew hepatitis B virus.

²AMDV: astatotilapia metahepadnavirus; BGHBV: bluegill hepatitis B virus; TMDV, tetra metahepadnavirus; WSHBV: white sucker hepadnavirus; SLHBV: spiny lizard hepatitis B virus; SkHBV: skink hepatitis B virus; TFHBV, Tibetan frog hepatitis B virus.

³ASHBV: Ashy-headed Sheldgoose hepatitis B virus; CHBV: crane hepatitis B virus; CWHBV: Chilo wigeon hepatitis B virus; DHBV: duck hepatitis B virus; HHBV: heron hepatitis B virus; OSHBV: Orinoco sheldgoose hepatitis B virus; PHBV: parrot hepatitis B virus; RGHBV: Ross goose hepatitis B virus; SGHBV: snow goose hepatitis B virus; STHBV: stork hepatitis B virus.