HBV evolution and genetic variability: impact on prevention, treatment and development of antivirals

Dieter Glebe\textsuperscript{1,5*}, Nora Goldmann\textsuperscript{1}, Chris Lauber\textsuperscript{2,3,5}, Stefan Seitz\textsuperscript{2,4,5*}

(1) Institute of Medical Virology, Justus Liebig University of Giessen, National Reference Centre for Hepatitis B Viruses and Hepatitis D Viruses, Schubertstr. 81, 35392, Giessen, Germany

(2) Division of Virus-Associated Carcinogenesis, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

(3) Research Group Computational Virology, Institute for Experimental Virology, TWINCORE, Centre for Experimental and Clinical Infection Research, a joint venture between the Helmholtz Centre for Infection Research and the Hannover Medical School; Cluster of Excellence RESIST, Hannover Medical School, 30625 Hannover, Germany

(4) Department of Infectious Diseases, Molecular Virology, University of Heidelberg, 69120 Heidelberg, Germany

(5) German Center for Infection research (DZIF), Partner Sites Giessen, Hannover & Heidelberg

*Corresponding authors:

Dieter Glebe, E-Mail: dieter.glebe@viro.med.uni-giessen.de

Stefan Seitz, E-Mail: s.seitz@dkfz-heidelberg.de
Abstract

Hepatitis B virus (HBV) poses a major global health burden with 260 million people being chronically infected and 890,000 dying annually from complications in the course of the infection. HBV is a small enveloped virus with a reverse-transcribed DNA genome that infects hepatocytes and can cause acute and chronic infections of the liver. HBV is endemic in humans and apes representing the prototype member of the viral family *Hepadnaviridae* and can be divided into 10 genotypes. Hepadnaviruses have been found in all vertebrate classes and constitute an ancient viral family that descended from non-enveloped progenitors more than 360 million years ago. The *de novo* emergence of the envelope protein gene was accompanied with the liver-tropism and resulted in a tight virus-host association. The oldest HBV genomes so far have been isolated from human remains of the Bronze Age and the Neolithic (~7,000 years before present). Despite the remarkable stability of the hepadnaviral genome over geological eras, HBV is able to rapidly evolve within an infected individual under pressure of the immune response or during antiviral treatment. Treatment with currently available antivirals blocking intracellular replication of HBV allows controlling of high viremia and improving liver health during long-term therapy of patients with chronic hepatitis B (CHB), but they are not sufficient to cure the disease. New therapy options that cover all HBV genotypes and emerging viral variants will have to be developed soon. In addition to the antiviral treatment of chronically infected patients, continued efforts to expand the global coverage of the currently available HBV vaccine will be one of the key factors for controlling the rising global spread of HBV. Certain improvements of the vaccine (e.g. inclusion of PreS domains) could counteract known problems such as low or no responsiveness of certain risk groups and waning anti-HBs titers leading to occult infections, especially with HBV genotypes E or F. But even with an optimal vaccine and a cure for hepatitis B, global eradication of HBV would be difficult to achieve because of an existing viral reservoir in primates and bats carrying closely related hepadnaviruses with zoonotic potential.
1 Introduction

Despite the availability of a highly efficient recombinant vaccine for more than 30 years (Huzair and Sturdy, 2017), the number of individuals persistently infected with the human hepatitis B virus (HBV) is still increasing (Thomas, 2019). In 2015, about 260 million people worldwide were chronic HBV carriers living at high risk of developing liver cirrhosis and hepatocellular carcinoma (HCC). The infection with HBV resulted in approximately 890 thousand deaths in 2015, and the trend is on the rise (WHO, 2019). Currently, the annual death toll associated with hepatitis B is estimated to have exceeded that of HIV and malaria (Thomas, 2019). According to the WHO, only roughly 11% of the affected individuals have been diagnosed with hepatitis B (HB) and are thus aware of being infected; less than 2% of the chronically infected patients have access to state-of-the-art antiviral treatment (WHO, 2019). In light of this factual under-diagnosis and under-treatment, it is the inhabitants of low-income countries who suffer most from this plague (O'Hara et al., 2017).

The causative agent is a small enveloped DNA virus which appears in electron micrographs as a double-shelled spherical particle with a diameter of 42 nm (Dane et al., 1970; Dryden et al., 2006; Seitz et al., 2007) (Figure 1A). The inner shell of the infectious virion is an icosahedral capsid assembled from 240 Core (C) protein subunits (Böttcher et al., 1997; Conway et al., 1997; Wynne et al., 1999) (Figure 1B). The capsid contains the viral replication complex composed of a reverse-transcribed, partially double-stranded circular DNA genome, covalently attached to a copy of the viral polymerase P (Seeger and Mason, 2000). This nucleocapsid is enclosed into an outer shell, a lipid bilayer derived from the host ER membrane into which the three envelope proteins L (large), M (middle) and S (small) are embedded – collectively referred to as hepatitis B surface antigen (HBsAg or shortly HBs) (Heermann et al., 1984; Seitz et al., 2016). In the bloodstream of infected persons, intact virions are outnumbered by a $10^3$- to $10^5$-fold excess of subviral particles (SVP) consisting only of empty envelopes (Heermann et al., 1984).
The circular genome of HBV encompasses just 3.2 kilobases (kb) comprising four open reading frames (ORFs) (Gerlich et al., 2020; Seeger and Mason, 2000) (Figure 1C). The C ORF encodes the 183 amino acids (aa) of the capsid protein. Translation initiation at the next in-frame start codon upstream of the C start gives rise to the Precore (PreC/C) protein (Wang et al., 2020), which after proteolytic processing becomes secreted into the serum as e-antigen (HBeAg), a diagnostic marker of the viral replication activity (Kramvis et al., 2018). The viral polymerase is encoded by the P ORF and can be subdivided into three functional domains: The terminal protein (TP) domain serves as protein primer to initiate minus-strand DNA synthesis during reverse transcription of the pregenomic RNA (pgRNA); it is separated from the reverse transcriptase (RT) and RNase H (RH) domains of P by a flexible spacer region (Bartenschlager and Schaller, 1988). The three envelope proteins are expressed from a single ORF (PreS1/PreS2/S) by utilization of alternative in-frame start codons (Heermann et al., 1984). The fourth ORF codes for the X protein (HBx) required for establishment and maintenance of infection and involved in hepatocarcinogenesis (Minarovits and Niller, 2019; Slagle and Bouchard, 2018) (Figure 1C).

2 HBV evolution

2.1 Hepadnaviridae – an ancient viral family

The infectious nature of the certain forms of hepatitis was known at least since the end of the 19th century, but the viral origin of hepatitis B could only be determined in the 60s and 70s of the 20th century (Gerlich, 2013). HBV is the prototype member of the increasingly large viral family Hepadnaviridae. The first non-human hepatitis B viruses had been identified in woodchucks (WHV) (Summers et al., 1978), squirrels (Marion et al., 1980; Testut et al., 1996) and ducks (DHBV) (Mason et al., 1980). In particular WHV and DHBV served as indispensable surrogate systems to study the molecular biology of HBV in the past (Roggendorf et al., 2015; Schultz et al., 2004), before cell lines fully supportive for HBV infection became available (Watashi et al., 2014). In the subsequent years,
several more hepadnaviruses were found in birds (genus *Avihepadnavirus*) by classical methods of molecular virology (Chang et al., 1999; Guo et al., 2005; Jo et al., 2017; Piasecki et al., 2012; Prassolov et al., 2003; Pult et al., 2001). In the same way, a mammalian hepadnavirus (WMHBV) was isolated from the woolly monkey, a New World primate (Lanford et al., 1998).

With the advent of next-generation sequencing techniques, the frequency of newly described hepadnaviral species accelerated dramatically. In the recent years, many novel species from diverse mammalian hosts (genus *Orthohepadnavirus*) have been detected in bats (Drexler et al., 2013; He et al., 2013a; He et al., 2013b; Hiller et al., 2019; Lei et al., 2019; Nie et al., 2018; Wang et al., 2017; Yang et al., 2018), in shrews (Nie et al., 2019; Rasche et al., 2019), in the domestic cat (Aghazadeh et al., 2018), in the capuchin monkey (de Carvalho Dominguez Souza et al., 2018), and in Maxwell’s duiker (Gogarten et al., 2019), a small antelope from West Africa. Hepatitis B viruses meanwhile have been discovered in all other classes of bony vertebrates, too, i.e. in reptiles (Lauber et al., 2017), amphibians (Dill et al., 2016; Lauber et al., 2017) and ray-finned fishes (Dill et al., 2016; Hahn et al., 2015; Lauber et al., 2017) (Figure 1D).

Very intriguing and fundamental to our understanding of the evolutionary history of the whole virus family was the observation that certain hepadnaviral lineages integrated into the genomes of their hosts’ germline and subsequently became fixed as endogenous viral elements (EVEs) (Feschotte and Gilbert, 2012; Patel et al., 2011). Such endogenization events occurred independently in birds (Gilbert and Feschotte, 2010; Lauber et al., 2017; Suh et al., 2013), crocodiles (Suh et al., 2014), snakes (Gilbert et al., 2014; Lauber et al., 2017; Suh et al., 2014), lizards (Lauber et al., 2017) and turtles (Suh et al., 2014) several tens to hundreds of million years ago (mya). The proneness of birds and reptiles to germline invasion of hepadnaviral sequences might result from the looser tissue tropism of these peculiar viruses able to infect a wide range of glandular organs.

A second step toward deeper insight into hepadnavirus evolution was the discovery of a related family of non-enveloped fish viruses, termed nackednaviruses, which share similarities with regard to
the genome organization, the mode of protein-primed reverse transcription and the ultrastructure of the viral capsids (Lauber et al., 2017) (Figure 1E).

2.2. Reconstructing the origin and evolutionary history of hepadnaviruses

Nackednaviruses represented the ideal outgroup to safely root the phylogenetic tree of hepatitis B viruses for the very first time, while the inclusion of an avian EVE of known age (eAHBV-FRY) enabled time-calibrating this phylogeny. These analyses suggest that both virus families separated from a common ancestor in the Silurian about 430 mya concomitant with the divergence of ray-finned fishes (Actinopterygii) and lobe-finned fishes (Sarcopterygii), the latter ones including all terrestrial vertebrates (Tetrapoda). The envelope protein gene in the hepadnaviral lineage emerged \textit{de novo} in a rather short window of time between 380 and 360 mya, concomitant with the rise of the first tetrapods (Figure 2).

Since then, these viruses coevolved in intimate association with their respective host lineages. On the deep evolutionary scale, virus-host cospeciation appears to dominate, although some successful host jumps across vertebrate classes occurred. The gain of the envelope protein gene represents a fundamental transition in viral lifestyle. We assume that it was directly involved in the development of the hepatotropism, which in turn might have resulted in narrowing the host tropism limiting cross-species transmission.

Subsequently, several independent “innovations” appeared in the distinct branches of the viral family. In the lineage leading to avi- and herpetohepadnaviruses, the C ORFs became elongated due to several insertions (Figure 2). The largest C protein to date is found in the skink hepatitis B virus (SkHBV) and comprises 335 aa (vs. 183 aa in human HBV). In a common ancestor of ortho- and metahepadnaviruses, the latter infecting fishes, the \(\alpha\)-determinant emerged. This insertion into the S ORF constitutes the major immunogenic ectodomain on the viral particle surface, against which the
anti-HBs response is directed (see also Figure 5). Finally, the X ORF originated in the ancestry unique to orthohepadnaviruses (Figure 2; see also Figure 1C and D).

The strong association of hepadnaviruses with their hosts over geologic eras suggests that the rate of viral macroevolution approximates that of their hosts. This synchronicity stays in sharp contrast to much faster rates of viral molecular evolution observed within infected individuals (Tedder et al., 2013). In line with the latter are findings from a study on intra- and inter-host evolution in a family of chronic HBV carriers across three generations covering a 100-year period of virus diversification (Lin et al., 2015). The viral evolutionary rate within each family member during the years of chronic infection was found to be significantly faster than that between the carriers in the vertical (mother-to-infant) transmission chains. Importantly, the authors observed an accumulation of non-synonymous substitutions at immune epitopes of structural genes in the viral quasispecies within each individual. To explain the difference in substitution rates within and between hosts, the authors hence proposed the viral mutant spectrum to switch between colonization and adaptation. “Colonizers” are thought to represent optimally replicating viruses that are in advantage early after transmission into an immunologically naive host, particularly in newborns or immune-compromised persons. “Adaptors”, on the other side, diversify under pressure of the host immune system during the late inflammatory phase of chronic infection at the cost of replicative fitness. The fast intra-host evolutionary rates are hence attributed to the divergence of the “adaptors,” whereas the back-selection toward “colonizers” succeeding each transmission event is thought to be responsible for the slow viral macroevolution.

2.3. HBV host range and genotypes

Natural hosts for HBV are humans and non-human apes (=hominoids; superfamily Hominoidea) (Grethe et al., 2000; Robertson and Margolis, 2002; Starkman et al., 2003). Isolates of human origin cluster into at least 9 genotypes (A-I, and the provisional 10th genotype J), and sub-genotypes thereof
(Kramvis, 2014; Schaefer, 2007; Tatematsu et al., 2009; Velkov et al., 2018). The viral strains found to circulate in non-human apes constitute two distinct branches interspersed into the phylogeny of human HBV genotypes (Figure 3). One branch comprises the isolates from African apes, i.e. chimpanzees and gorillas (Hu et al., 2001; MacDonald et al., 2000; Njouom et al., 2010; Takahashi et al., 2000); the other one contains those from Asian apes, i.e. gibbons and orangutans (Noppornpanth et al., 2003; Sa-nguanmoo et al., 2008; Sall et al., 2005; Verschoor et al., 2001; Warren et al., 1999).

Neither the phylogenetic pattern, nor the geographic distribution of the genotypes is compatible with an exclusive mode of virus-host co-divergence and co-spreading (Littlejohn et al., 2016), but rather indicates that independent introductions into the respective host populations occurred repeatedly in the past. The picture gets even more complicated by frequent recombination events between genotypes (Araujo, 2015; Bollyky et al., 1996; Castelhano et al., 2017; Morozov et al., 2000; Shi et al., 2012; Simmonds and Midgley, 2005; Yang et al., 2007).

The advent of techniques to isolate and sequence ancient DNA opened the possibility not only to trace (pre)historic waves of human migration (Stoneking and Krause, 2011), but also to recover genome sequences of pathogens infesting these peoples (Spyrou et al., 2019). The first historic HBV genomes have been detected in a Korean mummy (Kahila Bar-Gal et al., 2012) and an Italian child mummy (Patterson et al., 2018), both from the 16th Century. Recently, two studies isolated ancient HBV from human body remains of individuals who lived in Eurasia in the Medieval, the Bronze Age and the Neolithic up to 7.000 years before present (Krause-Kyora et al., 2018; Muhlemann et al., 2018). Surprisingly, several of these ancient HBV strains are phylogenetically most closely related to those viral isolates endemic in present-day African apes (Figure 3).

2.4 Dating and tracing the origin and dispersal of HBV genotypes

Former divergence time estimates for the extant HBV genotypes were based on molecular clock models in which substitution rates were derived by relating the genetic distances between viral
sequences to the year of virus isolation. These approaches inferred time estimates for the most recent common ancestor (MRCA) ranging from several hundred to 7,000 years before present (Godoy et al., 2013; Mizokami and Orito, 1999; Orito et al., 1989; van Hemert et al., 2011). Inclusion of the Bronze Age strains retrieved a median age estimate for the MRCA of 12,000 years (Muhlemann et al., 2018), pointing out the ambiguity of these molecular clock approaches. Further evidence for a much slower evolutionary rate of virus evolution on deep time scales and thus a higher age of the MRCA came from the observation that the two Neolithic strains were more similar to each other than to any other HBV isolate, although the infected carriers had lived almost 2,000 years apart (7,000 vs. 5,000 years before present) (Krause-Kyora et al., 2018).

The alternative dating method using avihepadnaviral EVEs for phylogenetic tree calibration revealed a more than 1000-fold higher age estimate for the HBV MRCA (Lauber et al., 2017). According to this analysis, the HBV genotypes started to diversify about 30 mya (Figure 4) concomitant with the adaptive radiation and geographic dispersal of early hominoids across Africa and Eurasia (Begun, 2003; Bohme et al., 2019; Springer et al., 2012; Stevens et al., 2013).

Since there is no species barrier for HBV to be transmitted between humans and extant apes, we suggest that the same also applied to the plethora of extinct hominoids during their complex evolutionary history since the late Oligocene. Thus, HBV might have been freely floating within and between ancient ape species, whenever they came into local contact with each other in the past. This assumption is supported by the finding of recombination between chimpanzee and gorilla HBV strains circulating in the wild (Lyons et al., 2012; Yang et al., 2007). Moreover, an isolate from an East African wild-born chimp showed evidence for inter-host recombination with human HBV genotype C (Magiorkinis et al., 2005). Consequently, the diverse genotypes as we see them nowadays might have originated from ancient virus-host co-speciation events and successive (and still ongoing) host jumps accompanied by recombination events.
Recent studies on prehistoric genetic admixture found that anatomically modern humans not only successfully interbred with Neanderthals (Green et al., 2010; Prufer et al., 2014) and Denisovans (Browning et al., 2018; Meyer et al., 2012) in Eurasia, but also with a yet unknown sort of archaic humans in West-Africa ~40,000 years before present (Durvasula and Sankararaman, 2020; Hammer et al., 2011). The ancestors of Neanderthals and Denisovans in turn interbred with a “superarchaic” Eurasian population that separated from other human populations about 2 mya (Rogers et al., 2020). It is more than likely that on such occasions of productively exchanging genetic material between distantly related human populations, a less pleasant exchange of pathogens happened, too. Therefore, we propose that anatomically modern humans, during their expansion out of Africa across the whole Old World, picked up and became the melting pot for a diverse spectrum of preexisting HBV genotypes. Since then, some of these genotypes might have disappeared from mankind again, such as those Neolithic and Bronze Age strains clustering with the isolates from African apes (Krause-Kyora et al., 2018; Muhlemann et al., 2018), while others were reshaped to a certain degree by recombination to form the variability and geographic distribution pattern we observe today (Muhlemann et al., 2018).

2.5 Reservoirs for zoonotic HBV transmissions

A worldwide eradication of human HBV would require the following prerequisites: first of all, a better protection from infection by an even more effective vaccine in combination with improved hygiene practices, and second the de facto cure of the many millions of chronically HBV-infected people as a source of infection. Last but not least, much stricter wildlife protection measures to reduce human-ape-contacts will be necessary both to save our closest relatives from extinction and to protect ourselves from viral reintroduction (e.g. via “bush meat”). As noted above, it should be considered that based on genetic classification, one single HBV species contains all viral isolates endemic in hominoids, including all human HBV genotypes and HBV isolates from apes (chimpanzee, gorilla, orangutan, and gibbon). Although hepadnaviral infections are regarded as highly host-specific,
interspecies transmission events of different HBV genotypes between members of the superfamily Hominoidea can be deduced from phylogenetic data. Interspecies transmission of the two New World primate hepadnavirus species, WMHBV isolated from woolly monkeys (Lanford et al., 1998) and CMHBV from capuchin monkeys (de Carvalho Dominguez Souza et al., 2018), have not yet been described but are most likely, too. Efficient experimental infection with WMHBV has been described in spider monkeys (Lanford et al., 1998) which are closely related to woolly monkeys. No such experimental *in vivo* data are available for CMHBV, but *in vitro* data have shown that CMHBV surface proteins can specifically interact with homologues of the HBV entry receptor sodium-taurocholate cotransporting polypeptide (NTCP), from closely related monkey species (de Carvalho Dominguez Souza et al., 2018). Interestingly, CMHBV surface proteins are also able to specifically interact with the (human) hNTCP and support infection of human hepatocytes via hNTCP (de Carvalho Dominguez Souza et al., 2018). So far, TBHBV from the tent-making bat *Uroderma bilobatum* is the only known non-primate hepadnavirus whose surface proteins interact very efficiently with hNTCP and can also infect primary human hepatocytes *in vitro* (Drexler et al., 2013). The most worrying aspect is that although CMHBV infection of human hepatocytes could be neutralized *in vitro* by high doses of anti-HBs (> 500 IU/L) from successfully vaccinated individuals (de Carvalho Dominguez Souza et al., 2018), this was not possible for TBHBV infections due to the many amino acid exchanges in the antigenic determinant of the surface proteins compared to the HBV vaccine (Drexler et al., 2013). In summary, primates harbor a reservoir of hepadnaviruses with varying zoonotic potential. The only non-primate animals that carry hepadnaviruses with known zoonotic potential are from the large and diverse order of bats (Chiroptera). Further investigations are necessary to characterize the full zoonotic potential of the many newly discovered hepadnaviruses in bats.

### 3 Impact of HBV genetic variability for development of vaccines and antiviral therapy

#### 3.1 The HBV surface proteins
The three co-carboxy-terminal HBV surface proteins L-, M- and SHBs, encoded by a single ORF (Figure 1), can be distinguished by amino-terminal polypeptide extensions (PreS1+PreS2) and co- and posttranslational modifications (Glebe and Bremer, 2013). The shared S domain encompasses 226 aa (Figure 5). The amino-terminal extensions to SHBs are the PreS1 domain (present only in LHBs) and the PreS2 domain (present in L- and MHBs). Across all wildtype HBV genotypes known so far, the number of aa residues of SHBs and the PreS2 domain (55 aa) is constant, while the PreS1 domain of LHBs shows genotypic variations of either 109, 118 or 119 residues (Figure 5) (Glebe and Bremer, 2013; Kramvis, 2014). The HBV surface proteins are translated from two viral mRNA species with different 5’ ends. The longer mRNA (2.4 kb) exclusively facilitates synthesis of LHBs, while M- and SHBs are both translated from a single shorter mRNA species (2.1 kb) (Glebe and Bremer, 2013). Transcription of these two HBV mRNAs is driven by different viral promoter/enhancer systems, thus allowing independent regulation of transcription and thus differential intracellular expression levels of LHBs versus M- and SHBs (Cattaneo et al., 1983; Raney et al., 1991). This is important because an intracellular imbalance in favor of LHBs towards the amount of M- and SHBs can lead to adverse effects in the cell, such as endoplasmic reticulum (ER) stress, which can trigger cellular signals for apoptosis or uncontrolled cellular growth (Montalbano et al., 2016).

3.2 The S domain of HBV surface proteins and its antigenic loop

The S domain is translated at the ER with integration of four hydrophobic transmembrane domains (TM-I-IV) into the ER membrane (Figure 5). Stability of SHBs is provided by formation of intra- and intermolecular disulfide bridges between cysteines in the lumen of the ER, resulting in formation of a 60 aa long external antigenic loop (AGL, aa 99-160) between TM-II and TM-III that carries the major antigenic determinant (called α-determinant) of the viral surface (Glebe and Bremer, 2013). The hydrophilicity of this domain is partly increased by N-glycosylation at the highly conserved residue Asn146, but only about half of the synthesized protein molecules are finally glycosylated at this site.
During translation of LHBs and MHBs, their C-terminal S domains are produced in a similar manner. Intermolecular disulfide bridging between the S domains of LHBs, MHBs, and SHBs leads to formation of protein homo- and heterodimers (Suffner et al., 2018) that in turn self-assemble to form SVP or budding virions (Ning et al., 2018) which become actively secreted from infected hepatocytes without cellular lysis (Blondot et al., 2016). The numerous intra- and intermolecular disulfide bonds within the S domains (Suffner et al., 2018) contribute to the well-known structural stability of SVP and virions and are one of the reasons for the high resistance of HBV to inactivation by dehydration and heat stress (Kobayashi et al., 1984; Konig et al., 2019). The molecular basis of this stability, exceeding that of most other enveloped viruses, is mainly determined by eight cysteines within the external loop of the S domain. The high cysteine density in this region is conserved within the genus Orthohepadnavirus and a feature shared with the piscine metahepadnaviruses (Lauber et al., 2017). Exchange of these cysteines by mutagenesis leads to severe impairment of the conformation of the S domain (Mangold et al., 1997) with a reduced budding competence and infectivity of virions (Abou-Jaoude and Sureau, 2007).

### 3.3 HBV-Vaccines and Vaccine-approaches

The starting biomaterials for the first active HBV vaccines were extracted directly from plasma of chronically HBV-infected patients (for reviews, see: Gerlich, 2015; Huzair and Sturdy, 2017). However, it was not the virions that were utilized by the vaccine manufacturers, but the non-infectious SVP present in huge excess in carrier plasma. These SVP can be separated quite easily from the virions by biophysical methods. Subsequent processing, such as inactivation of remaining viruses and adjuvantation, resulted in safe and immunogenic first-generation HBV vaccines that provided protection against acute and chronic hepatitis B, based on initial clinical trials in high-risk groups (Buynak et al., 1976; Maupas et al., 1976; Szmuness et al., 1980; Thomssen et al., 1982). The cloning of the HBV genome enabled the large-scale production of SVP in recombinant yeast cells, and the
corresponding yeast-derived second-generation vaccines showed comparable protection against HBV as plasma-derived vaccines in early clinical trials (McMahon, Wainwright, 1993). The global use of these vaccines to immunize infants, adolescents and special risk groups has been instrumental in reducing new HBV infections, and thus the global incidence of acute/chronic HB and its sequelae liver cirrhosis and HCC (Gerlich, 2015). While vaccine response rates with yeast-derived vaccines are excellent in healthy infants and adolescents (> 99%), they are insufficient in about 5% of healthy adults. The rate of non-responders can rise to 70% in certain groups of people with pre-existing lifestyle/medical problems or conditions of a lowered immune competence (Gerlich, 2015). For historical reasons, the vast majority of yeast-derived vaccines consist only of the SHBs of the globally underrepresented HBV genotype A2, which is dominant only in Northern Europe and North America. Nevertheless, it also protects against infections with other genotypes, although well-documented clinical cases of breakthrough infections with genetically very distant genotypes (e.g. F) despite prior successful vaccination have been reported (Luongo et al., 2015; O’Halloran et al., 2011; Tacke et al., 2007). The underlying problem with these incidences of vaccine escape is that a predominant fraction of the immune response to the vaccine is subtype-specific. With high anti-HBs titers (> 1,000 IU/L), this may be irrelevant, but low or waning anti-HBs-titers over time increase the risk of breakthrough infection with antigenically distant HBV genotypes. This problem was also evident in studies on newly acquired HBV infections of healthy American blood donors (Stramer et al., 2011). The frequencies of breakthrough infections among those vaccinated blood donors were significantly higher for HBV genotypes B, C, D and F and at anti-HBs titers of less than 100 IU/L. However, all non-vaccinated, freshly infected individuals observed in this study had acquired HBV subgenotype A2, the predominant variant in the USA (Stramer et al., 2011). Although all infections described in this study had been asymptomatic and transient, this still poses a threat to blood safety due to the very high infectivity of HBV. Recent calculations of the minimal infectious dose of HBV are as low as 16 virions (calculated as HBV genomes, or 3 International Units, IU), following accidental transmission through HBV-contaminated blood transfusions (Candotti et al., 2019).
Mutations affecting antigenicity and vaccine-escape

A single amino acid exchange (G145R) within the α-determinant/AGL of the S domain (Figure 5) was one of the first HBV variants found to be associated with breakthrough infections in newborns from chronically infected mothers, despite the combined active/passive vaccination directly after birth (Gerlich, 2015). The obvious concern that this immune-escape variant could spread efficiently in a population of fully vaccinated individuals turned out to be negligible given that the anti-HBs titers in the population are sufficiently high: Successfully vaccinated chimpanzees with high anti-HBs titers were protected against infection with variant G145R (Ogata et al., 1999). Notably, many other immune escape mutations have been identified and characterized since then. For an extended overview of the genetic variability of HBV immune escape variants see also: (Rajoriya et al., 2017; Tong and Revill, 2016).

Emergence of vaccine-escape mutations during specific antiviral therapy

Due to the overlapping ORFs of P and S (Figure 1C), adaptive mutations within the RT domain positively selected under therapy with nucleoside/nucleotide analogues (NA) may concomitantly lead to non-synonymous substitutions of corresponding codon positions in the S-ORF (Torresi, 2002). The development of specific mutations within the α-determinant was indeed observed during NA therapy of chronic HBV carriers; vaccine-escape mutations occurred in particular when lamivudine was used, even when patients were not treated with anti-HBs (Colagrossi et al., 2018). The individual mutations in the S-ORF not only led to specific amino acid exchanges in the α-determinant/AGL (Figure 5) but sometimes also to C-terminally shortened surface proteins resulting from introduction of a stop codon in the S-ORF, which leads to premature translation termination (Colledge et al., 2017). Misfolded, C-terminally truncated forms of HBs are associated with the development of HCC in transgenic mice. It is mediated by direct transcriptional activation through an altered intracellular
topology of the PreS domains of L- and MHBs (Hildt et al., 2002). However, more clinical studies are needed to determine whether these effects observed in mice also manifest themselves in chronically HBV-infected patients. Furthermore, the observed intracellular retention and accumulation of these variants within the infected cell can contribute to ER stress and apoptosis (Colledge et al., 2017; Warner and Locarnini, 2008). Notably, the occurrence of such mutants remains underdiagnosed, since secretion not only of virions but also of SVP is inhibited which may result in detection of low or no HBsAg in the serum of affected chronic carriers. The relevant mutations have been frequently described under therapy with NA that have a low genetic barrier (e.g. lamivudine, adefovir) and often led to therapy breakthroughs (Colagrossi et al., 2018; Colledge et al., 2017; Torresi, 2002). The use of recommended NA with improved efficacy and higher genetic barrier (e.g. tenofovir, entecavir) should reduce the occurrence of these dangerous escape mutants during therapy.

3.4 The PreS1 domain of LHBs determines viral infectivity of HBV

LHBs is the key player for viral envelopment and infectivity (for reviews, see: Glebe and Bremer, 2013; Seitz et al., 2020). While the C-terminal part of PreS1 (extending into the first 5 aa of PreS2) is necessary for envelopment of cytosolic mature core particles during viral budding (Bruss, 1997), the high-affinity virus-receptor binding sites of primate HBV reside in the N-terminal part (Figure 5). During budding, the PreS1/PreS2 domains of LHBs have an entirely inward (cytosolic) orientation facilitating contact with core particles. Consequently, newly secreted virions are functional immature and non-infectious (i.e. not-yet-infectious) due to the absence of PreS1 on the particle surface (Seitz et al., 2020; Seitz et al., 2016). The subsequent topological switch of the PreS domains across the viral membrane onto the outside leads to receptor binding competence, thus rendering the virions infectious. This distinct maturation process initiates spontaneously in virions circulating in the blood stream and proceeds within the particle population with a half-time of 4.7 h (Seitz et al., 2020; Seitz et al., 2016).
The infection-determining regions for high-affinity binding to the liver-specific HBV uptake receptor NTCP are located within the N-terminal 75 amino acids of the PreS1 domain (Figure 5) (for reviews, see: Glebe and Bremer, 2013; Glebe and Urban, 2007; Seitz et al., 2020). A short and highly conserved motif within this determinant (NPLGFFP at aa position 9-15 in genotype D) together with the N-terminal myristoylation at glycine 2 have been shown to be essential for the infection process (Figure 5). This motif is conserved between HBV, the two related viral species from New World monkeys and some isolates from bats (Drexler et al., 2013).

3.5 Antivirals interfering with the early HBV infection process by targeting PreS1

Since the PreS1 domain of LHBs is a central driver of infectivity, many antiviral approaches have been taken to attack its key function in viral entry. These can be divided into two categories: disruption of the virus-receptor interaction by (1) neutralizing antibodies against PreS1, and (2) by blocking of the viral binding sites on NTCP.

Neutralizing antibodies against PreS1

Various anti-PreS1-antibodies have been successfully used to neutralize infection in vitro and in vivo. Most of them target epitopes located at a rather N-terminal part of PreS1 between aa 19-48 (Glebe, 2006; Glebe and Urban, 2007). However, there are known genotypic variations in this region that might reduce a broad-spectrum neutralizing efficacy of such monoclonal antibodies (Bremer et al., 2011; Glebe et al., 2003). On the other hand, the central receptor binding motif NPLGFFP at PreS1 aa 9-15 (Figure 5) is highly conserved between all three primate and some bat hepadnavirus species (Drexler et al., 2013) and single amino acid changes within this small stretch abolish HBV infectivity completely (Engelke et al., 2006). This motif could therefore be considered the Achilles’ heel of the virus because specific antibodies against this particular sequence should be highly efficient in neutralizing hepadnaviral infections across all primate and bat species. The study of Bremer et al.
tested this hypothesis by generating polyclonal antisera against specific polypeptide stretches of PreS1 (Bremer et al., 2011). As expected, polyclonal antisera against the N-terminal (aa 2-48) and the C-terminal region (aa 78-108) showed broad reactivity against all HBV genotypes tested (A, C, D), including genotype F that is most distinct to all others with regard to the PreS1 sequence. However, only antisera against the N-terminal domain showed neutralizing activity (Figure 5). Interestingly, antisera generated against aa residues 2-21, containing the very essential receptor binding motif showed neither reactivity against SVP or virions purified from human sera, nor neutralizing activity during HBV infection in vitro (Bremer et al., 2011). Subsequent analyses with chemically synthesized PreS1-peptides demonstrated that the N-terminal myristoylation at glycine-2 prevented antibody recognition of the receptor binding motif. It turned out that it is not specifically the added myristoyl moiety itself decreasing antibody binding, but simply the increased hydrophobicity: The reactivity of the neutralizing antisera inversely correlated with the length of the carbon chain (C) artificially added to the N terminus of the synthetic peptides ranging from pentanoyl (C5) to the natural myristoyl (C14). Thus, although the receptor binding motif at aa residues 9-15 is immunogenic in principle, antibodies directed against this structure likely cannot reach it on the viral particle surface under natural conditions (Bremer et al., 2011). Nevertheless, anti-PreS1 antibodies targeting the downstream aa residues 19-48 are usually superior to anti-HBs antibodies in the prevention of HBV infection (Glebe et al., 2003), because they are directly interfering with the PreS1-dependent interaction of the virus with NTCP (Figure 5). However, amino acid variations within this domain between HBV genotypes have to be considered carefully during generation of highly cross-reactive anti-PreS1-antisera (Bremer et al., 2011).

**Blocking of the viral binding sites on the cellular HBV-receptor NTCP**

The importance of the first 75 aa of PreS1 for viral entry (Figure 5) was known before the discovery of the cellular bile-acid transporter NTCP as the high affinity HBV uptake receptor (for reviews, see:
The use of highly efficient infection-interfering myristoylated PreS1-peptides of variable length led to the characterization of infection-relevant essential (aa 9-15) and accessory motifs (aa 19-48) within LHBs (Figure 5) (Glebe et al., 2005; Gripon et al., 2005; Gripon et al., 2002, Petersen et al., 2008). These peptides were not only crucial to identify the viral infection determinants, but also led to the discovery and characterization of the NTCP receptor (Yan et al., 2012) and paved the way for developing specific HBV entry inhibitors. Due to the high degree of conservation and the essentiality of the respective amino acid sequence in all currently known HBV genotypes, it is very unlikely that naturally occurring viral mutants will undermine the antiviral effect of these highly active peptides.

As an alternative to acylated peptides, allosteric small molecule HBV-entry inhibitors targeting NTCP have been evaluated. These lead structures comprise natural substrates (bile acids) or derivatives thereof, or synthetic NTCP substrates (e.g. ezetimibe, irbesartan) that either block binding of HBV, inactivate NTCP or interfere with cellular NTCP re-uptake (for a review, see: Tu and Urban, 2018).

3.6 Antivirals interfering with reverse transcription and replication of HBV

Direct acting antivirals are nucleoside/nucleotide analogues (NA) that specifically interfere with HBV replication in infected cells by acting on reverse transcription and/or the DNA-dependent DNA synthesis activity of the HBV polymerase (for a review, see: Zoulim and Locarnini, 2009). Currently approved first line drugs are entecavir (ETV), a deoxyguanosine analogue and the deoxyadenine analogue tenofovir (TDF) including its variant form tenofovir alafenamide (TAF). They are highly effective in significantly lowering viremia virtually down to negativity upon long term treatment of chronic HB (CHB) patients. They have a favorable oral administration route and usually good tolerability with minor side effects. In many cases they have to be taken lifelong, since NA do not directly interfere with the stability of the viral covalently closed circular DNA (cccDNA) persisting as an episome in the nucleus of the host cell (for a review, see: Buti et al., 2018). Nevertheless, years
after continuous antiviral treatment with NA, a decline of the cccDNA levels in the liver can often be observed (Lai et al., 2017). In summary ETV and TDF/TAF efficiently interfere with HBV replication of all present HBV genotypes, because the functional features of the hepadnaviral polymerase are highly conserved within the family of hepadnaviruses.

4 Conclusions

HBV is a genetically diverse virus species composed of 10 extant genotypes and several extinct genotypes that affected historic human populations in the past. Being present in humans and other apes, its origin can be traced back to several million years into the past using molecular phylogenetics approaches and genome-based comparison with related hepadnaviruses infecting diverse vertebrate species. The strong host association and correspondingly low rate of hepadnavirus evolution at the macroevolutionary scale is in stark contrast to high rates of HBV evolution within infected individuals. Diversification of these “adaptors” under pressure of the host immune system poses various challenges for vaccine and antiviral development including the emergence of resistance-conferring substitutions that might be genotype-specific.

Full eradication of HBV appears intractable at the moment due to different reasons. These include a viral source in chronic HBV carriers that together constitute a significant portion of the human population as well as a natural reservoir in apes. Moreover, closely related hepadnavirus species in New World monkeys and diverse bats represent a potential risk for zoonotic spillover infections into humankind. Additional problems are posed by the poor or absent responsiveness and waning antibody titers in a subset of vaccinees. Future research efforts should thus focus on the improvement of vaccines and the development of effective antivirals. In addition, medical surveillance systems have to be established to interrupt the globally predominant vertical mother-to-infant transmission route. For any such effort a focus should be put on low-income and developing countries that suffer most from the high burden of HBV prevalence.
Acknowledgments

We thank Ben Krause-Kyora (University Kiel) and Johannes Krause (Max-Planck-Institute Jena) for providing the nucleotide alignment on which Figure 3 is based on. S.S. acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for project TP23N within TRR 179. D.G. was supported by grants B08/SFB1021 and GL595/9-1 from the DFG. The National Reference Centre for Hepatitis B Viruses and Hepatitis D Viruses at Justus Liebig University Giessen is supported by the German Ministry of Health via the Robert Koch Institute, Berlin. CL is supported by the DFG through the Cluster of Excellence RESIST (EXC 2155), project 390874280.

References

Bartenschlager, R., Schaller, H., 1988. The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. EMBO J 7, 4185-4192.


Figure 1. Morphology and genome architecture of hepadnaviruses. A. Transmission electron micrograph of HBV particles purified from cell culture supernatant. Negative stain with uranyl acetate. Spherical 42-nm virions (Dane particles) and filamentous and spherical 20-nm SVP. Scale bar: 50 nm. B. Schematic representation of the Dane particle structure. C-E. Representative genome maps depicting open reading frames. DR1, DR2: direct repeats. C. HBV. D. Tetra metahepadnavirus (TMDV) from the Mexican tetra (*Astyanax mexicanus*), a blind cavefish. E. Rockfish nackednavirus (RNDV) from the tiger rockfish (*Sebastes nigrocinctus*). Note the absence of a surface protein gene. smORF: small open reading frames coding for proteins of unknown function.
Figure 2. Hepadnavirus phylogeny basing on a multiple sequence alignment of conserved Pol regions and rooted using 13 nakednaviruses as outgroup. The BEAST tool was used to reconstruct this Bayesian tree under the LG+G4+I substitution model, a relaxed molecular clock with log-normally distributed rates and a Yule speciation prior. Size of the circles at internal branching points is proportional to the posterior probability of a split, which ranges from 0.43 to 1.0. The unit of the scale bar is average number of substitutions per site. Major evolutionary innovations are marked by red arrows.
Figure 3. Phylogenetic network of Hepatitis B virus, reconstructed using the SplitsTree tool with default parameters. The network is based on a multiple nucleotide alignment of 552 HBV genome sequences (Krause-Kyora et al., 2018; Muhlemann et al., 2018). HBV genotypes are indicated by capital letters, and the positions of viral lineages infecting non-human apes are marked through host names. Ancient HBV lineages from the Neolithic (blue; 2 isolates), the Bronze Age (orange; 12) and the Medieval period (green; 1) are highlighted by terminal dots.
Figure 4. Time-calibrated phylogeny of HBV genotypes. The tree is a subtree of the phylogeny shown in Figure 4 of Lauber et al. (2017). Woolly monkey hepatitis B virus (WMHBV) serves as outgroup. Numbers indicate divergence time estimates in million years ago (mya). Bars represent 95% confidence intervals.
Figure 5. Organization and function of HBV surface proteins. The three hepatitis B virus surface proteins are depicted in their co-carboxy-terminal orientation. The small hepatitis B virus surface protein (SHBs) consists of the S domain (red), comprising 226 amino acids (aa) and forms at least 4 transmembrane domains (blue arrows). The middle hepatitis B virus surface protein (MHBs) consists of the S domain and the amino-terminal PreS2 domain (55 aa). Addition of the PreS1 domain (yellow) to the PreS2 and the S domains forms the large hepatitis B virus surface protein (LHBs). The length of the PreS1 domain is 108, 118 or 119 aa depending on the HBV genotype. The PreS1 domain is myristoylated at the amino-terminal glycine-2. Envelopment of virions is dependent on a conserved junction of PreS1 and PreS2 within LHBs. HBV infection is initiated by low-affinity binding to cellular heparansulfate proteoglycans (HSPG - *it was reported that this process is mediated by the PreS1 domain (Leistner et al., 2008; Schulze et al., 2007), while another report attributed it to the external antigenic loop (AGL) of the S domain (Sureau and Salisse, 2013)). High-affinity HBV binding leading to infection is mediated by the human liver bile acid transporter NTCP (Na+/taurocholate co-transporting peptide). A short and highly conserved amino acid stretch (PreS1 aa 9-15, depending on the genotype, dark green) within the amino-terminal 75 aa of myristoylated PreS1 (light green) is essential for infection. The AGL within the S domain and the amino-terminal domain of PreS1 (aa 15-75) contain B-cell epitopes that are recognized by neutralizing antibodies. N- and O-glycans of HBV surface proteins were omitted for simplification.