Available online at www.sciencedirect.com
ScienceDirect
Virus Research 127 (2007) 164–176
Virus Research
www.elsevier.com/locate/virusres

Hepatitis B virus genetic variability and evolution

Alan Kay [a,b,*], Fabien Zoulim [a,b,c]

[a] *INSERM, U871, Physiopathologie moléculaire et nouveaux traitements des hépatites virales, 151 cours A. Thomas, Lyon F-69424, France*
[b] *Université Claude Bernard Lyon 1, Faculté de Médecine Laennec, IFR62, Lyon F-69008, France*
[c] *Hospices Civils de Lyon, Hôtel Dieu, Service d'Hépatologie, Lyon F-69002, France*

Available online 26 March 2007

Abstract

Hepatitis B virus has been evolving gradually over a long period of time, resulting in a large amount of genetic diversity, despite the constraints imposed by the complex genetic organization of the viral genome. This diversity is partly due to virus/host interactions and partly due to parallel evolution in geographically distinct areas. Recombination also appears to be an important element in HBV evolution. Also, human intervention in the form of mass vaccination and antiviral treatment will reduce the burden of HBV-related liver disease but may also be accelerating evolution of the virus.

Keywords: HBV; Genetics; Evolution; Serotypes; Genotypes; Mutations

1. Introduction

The association between man and hepatitis B virus (HBV) is certainly very ancient. The Hepadnavirus family comprises representatives from non-human primates that are very similar to HBV, from rodents (woodchuck hepatitis B virus, or WHV, etc.) that share about 80% similarity to HBV (Galibert et al., 1982) and from birds (duck hepatitis B virus, or DHBV, etc.) that share only about 40% similarity to HBV (Mandart et al., 1984). If one supposes a common ancestor and separate evolution within the different lineages, then hepadnaviruses must have existed before the speciation between birds and mammals. This longevity has permitted the emergence of a great deal of diversity, not only between the hepadnaviruses of the different species but also among HBV isolates, despite the severe constraints imposed by the genomic organization (Mizokami et al., 1997), the genomic structure and the replication strategy of Hepadnaviruses (Ganem and Schneider, 2001). The genomes of hepadnaviruses are very small, and DHBV, with only 3021 bp, possesses the smallest genome of known animal DNA viruses. This has forced the virus to optimize the genomic organization. There is much overlap of genes and every nucleotide participates in the coding of at least one viral protein. The regulatory and structural sequences necessary for viral transcription and replication are therefore automatically included within coding regions. This extensive overlap limits the diversity that the virus can tolerate. A mutation may have little effect on one viral protein but may have severe consequences on an overlapping gene or on regulatory and structural sequences. The structure of the genome is a direct reflection of the viral replication strategy. The genome exists in fact in two different forms (Ganem and Schneider, 2001). In the virions, the genome is a relaxed circular DNA molecule (RC-DNA) that is only partially double-stranded (Fig. 1). One strand is complete (the L or minus strand) and even has a short terminal redundancy with a protein, the viral polymerase, covalently attached to the 5'-end. The other strand (the S or plus strand) is incomplete, about two-thirds complete for HBV, almost complete for DHBV. The 5'-end is fixed and starts with a short oligoribonucleotide and the 3'-end is variable. The plus strand overlaps the 5' and 3' extremities of the minus strand, thereby assuring the circularity of RC-DNA. RC-DNA therefore contains all of the genetic information of the virus but is unsuitable as a replication template. After infection of a hepatocyte, plus strand DNA is completed, the oligoribonucleotide at the 5'-end of plus strand DNA and the terminal redundancy of minus strand DNA, along with the attached viral polymerase, are eliminated and the two strands are ligated. Importantly, these steps

* Corresponding author. INSERM U871, 151 cours Albert Thomas, 69424 Lyon Cedex 03, France. Tel.: +33 4 72 68 19 76; fax: +33 4 72 68 19 71.
E-mail address: kay@lyon.inserm.fr (A. Kay).

0168-1702/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2007.02.021

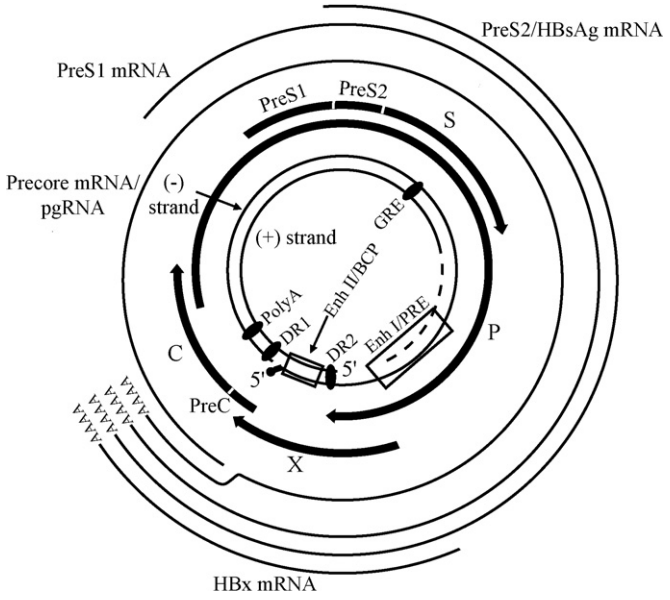

Fig. 1. Genetic organization of the HBV genome (RC-DNA form). GRE, glucocorticoid response element; Enh, enhancer; PRE, post-transcriptional regulatory element; BCP, basal core promoter; DR, direct repeat.

are apparently performed by cellular enzymes, independently of the viral polymerase. The genome is then found in the nucleus of the infected hepatocyte in the form of a covalently-closed circular DNA molecule (cccDNA) that is the real replication matrix. The mRNAs for the viral proteins are transcribed from the cccDNA. This includes pregenomic RNA (pgRNA) which is of more-than-genome length, approximately 1.1 genomes. In effect, transcription of pgRNA is initiated upstream of the unique viral polyadénylation signal. At the first passage of the polyadénylation signal, the nascent RNA is not, or only poorly, polyadenylated and efficient polyadenylation occurs only after the second passage. The pgRNA therefore possesses a terminal redundancy that is essential if the virus is not to lose genetic information during replication and can be compared to the LTRs of retroviruses. After export to the cytoplasm, the pgRNA is encapsidated along with the viral polymerase and minus strand DNA is synthesized via reverse transcription. The synthesis of plus strand DNA is initiated, but at some point there is maturation of the nucleocapsid containing RC-DNA and plus strand DNA synthesis stops. The nucleocapsid can then either be recycled back to the nuclease to amplify or replenish the cccDNA pool or it can be enveloped and secreted as a new virion. The duality of the genomic forms of hepadnaviruses has important consequences. The mutations that are generated by reverse transcription are initially found in RC-DNA. Mutations can be stably transmitted only if the mutated RC-DNA genome is recruited into the cccDNA pool of the patient, either by recycling to the nucleus or infection of a new hepatocyte by the mutant virus. In both cases, the mutant virus will be in competition with a vast excess of other viral genomes, either "wild type" or containing other mutations. The emergence of HBV mutants is therefore much slower, even in the presence of selective pressure, than with other viruses, such as HIV or HCV, where mutations generated by an error-prone replication step, reverse transcription or RNA-dependant transcription, can have immediate phenotypic effects. A second consequence of the dual nature of the HBV genome is the presence in a chronically infected patient of two quasispecies. The first is the quasispecies of the cccDNA pool and the second is the quasispecies of RC-DNA that reflects both the cccDNA quasispecies and new mutations generated during replication. Finally, the diversity of HBV genomes can be divided into two categories, a genotypic variability that is the result of gradual evolution of the genome in the absence of selective pressure and phenotypic variability that results from adaptation of the virus to selective pressures, either the host immune response or antiviral treatments, including vaccination. With genotypic variation, viral fitness is the most important factor. A mutant virus that is significantly less fit than other circulating HBV will eventually be eliminated even if the mutated RC-DNA genome integrates the cccDNA pool. With phenotypic variation, the driving force is selection, since the ability to resist antiviral pressure usually far outweighs the lowered fitness of a mutant compared to a wild type virus that cannot resist.

2. Genotypic variation

2.1. HBV serotypes

Very quickly after the discovery of "Australia antigen", or HBsAg, the major envelope protein of HBV, it became evident that sera of patients who had seroconverted to anti-HBs did not react in the same way with HBsAg from different chronic carriers and that this was due to viral variability. The first classification of HBV isolates was therefore done by serotyping, the reactivity of the HBsAg of the isolate with standard panels of antisera (Bouvier and Williams, 1975; Courouce et al., 1983). The major immunogenic region, the "*a*" determinant spanning residues 124–147 of HBsAg and probably composed of several conformational epitopes, is common to the almost all HBV isolates and is therefore not informative for classification. Classification is therefore done using subtypes, the molecular basis of which are now known (Table 1). The two major subtype epitopes are the *d*/*y* and *r*/*w* determinants. Both of these determinants are comprised of two mutually exclusive epitopes that depend upon the nature of the amino acids at positions 122 and 160 of HBsAg respectively. If the amino acid at position 122 is Arg (122R)

Table 1
Molecular basis of HBV serotypes

Serotype	HBsAg sequence
ayw1	122R + 160K + 127P + (134F and/or 159A)
ayw2	122R + 160K + 127P
ayw3	122R + 160K + 127T
ayw4	122R + 160K + 127L
ayr	122R + 160R
adw2	122K + 160K + 127P
adw3	122K + 160K + 127T
adw4q−	122K + 160K + 127L + 178Q
adrq+	122K + 160R + 177V + 178P
adrq−	122K + 160R + 177A