ELSEVIER

Journal of Clinical Virology 36 Suppl. 1 (2006) S3–S11

JOURNAL OF CLINICAL VIROLOGY

www.elsevier.com/locate/jcv

Genetic variation in HBV infection: genotypes and mutants

Stephan Günther*

Bernhard-Nocht-Institute of Tropical Medicine, Hamburg, Germany

Keywords: Hepatitis B virus; Mutants; Genotypes; Occult hepatitis B

1. Main types of sequence variability and their origin

Three main types of HBV sequence variability are known. The first type of variability reflects the HBV genotype or local strain. Genotype-specific variability is – on a short time scale – stably transmitted within the host population and is present from the beginning of infection in an individual. Little is known about the driving forces behind the divergence of HBV into genotypes, the reason for their specific geographic distribution, and the timing of the evolutionary events. Genotypes may result from neutral evolutionarily drift of the virus genome, from recombination, or as a consequence of a long-term adaptation of HBV to genetic determinants of specific host populations.

The second type of variability arises spontaneously during replication of HBV due to biochemical processes of the host cell or due to infidelity of the viral replication machinery. These mechanisms provide a basis for both evolutionary virus drift and selection of specific variants in a patient. Point mutations are presumably introduced by the HBV polymerase during reverse transcription but may also occur during RNA pregenome synthesis by the cellular RNA polymerase. Like other reverse transcriptases, that of HBV probably lacks a proofreading function. Another potential mechanism for introduction of point mutations is editing of viral DNA by cellular cytidine deaminases, resulting in G to A and C to T substitutions (Noguchi et al., 2005; Suspene et al., 2005). The occurrence of G–A hypermutated HBV genomes *in vivo* points to a possible role of this mechanism in generation of HBV mutations (Günther et al., 1997b; Noguchi et al., 2005). Deletions and insertions in HBV genomes can be generated by splicing of the pregenomic RNA (Su et al., 1989; Suzuki et al., 1989; Günther et al., 1997a; Sommer et al., 2000). Experiments with related hepadnaviruses suggest that topoisomerase I mediated cleavage/ligation of HBV DNA (Wang and Rogler, 1991) or recombination of linear HBV DNA molecules (Yang and Summers, 1995; Yang et al., 1996) contribute to generation of mutations.

The third type of viral sequence variability is the result of a selection process due to a selective advantage of variants during the course of HBV infection in a patient. If host conditions do not favour selection of specific strains, the circulating HBV genomes are genetically extremely stable (Okamoto et al., 1987; Bozkaya et al., 1997; Hannoun et al., 2000a). However, during specific phases of chronic infection we observe a complete replacement of the major virus population by specific variants within weeks or months, for example by those containing a defect in pre-core protein expression (pre-C mutants), amino-acid changes in core protein (HBc) and hepatitis B e antigen (HBeAg), mutations in pregenomic/core promoter, deletions in C gene, pre-S1 or pre-S2 region, and/or amino-acid changes in a-determinant of hepatitis B surface antigen (HBsAg). This can hardly be explained without assuming a selective advantage of these variants during particular stages of the chronic infection. However, it is currently unknown or merely speculative which, if any, of the multiple functional and immunological phenotypes documented *in vitro* for these mutants are relevant for their selection *in vivo*. Replication advantage has been demonstrated for variants with a novel hepatocyte nuclear factor 1 (HNF1) site in core promoter or with a deletion in C gene. Selection of other variants, for example those with a pre-C mutation, HBc/e amino-acid changes, or HBsAg a-determinant mutations may be an immune escape phenomenon. Selection of specific variants is independent of their genotype background, unless genotype-specific mutations interfere with the introduction of mutations as exemplified by the common pre-C mutation.

2. HBV Genotypes and regional strains

The first indication for sequence heterogeneity of HBV was the existence of different antigenic determinants of

* Correspondence address: Stephan Günther, Bernhard-Nocht-Institute of Tropical Medicine, Department of Virology, Bernhard-Nocht-Strasse 74, D-20359 Hamburg, Germany. Tel.: +49 40 42818 421; fax: +49 40 42818 378.

E-mail address: guenther@bni.uni-hamburg.de (S. Günther).

HBsAg which led to the classification of HBV into four major serological subtypes: adw, adr, ayw, and ayr (Le Bouvier, 1971; Bancroft et al., 1972). On the basis of sequence similarity, HBV is currently divided into 8 genetic groups, the genotypes A through H (Okamoto et al., 1988; Orito et al., 1989; Noreder et al., 1992, 2004; Kidd-Ljunggren et al., 2002). The inter-type differences in the nucleotide sequence of complete genomes are ⩾8%, while the intra-type differences are ⩽4% (Norder et al., 1994). There are some characteristic length differences between the genotypes which facilitate their detection and discrimination. Genotype A genomes have a 6-nucleotide insertion at the 3′ end of the C gene, and genotype D genomes have a 33-nucleotide deletion in the pre-S1 region. A typical feature of genotype G genome is a 36-nucleotide insertion at the 5′ end of the C gene. Due to this insertion, genotype G expresses a 24-kD core protein with 12 extra amino acids at the N terminus instead of the usual 21-kD protein. Nevertheless, it is fully competent for autonomous replication (Tran et al., 1991; Kremsdorf et al., 1996). Another feature of genotype G is the presence of stop codons in the pre-C region at codons 2 and 28 (Bhat et al., 1990; Tran et al., 1991; Stuyver et al., 2000; Kato et al., 2002b; Vieth et al., 2002). They prevent synthesis of HBeAg. Paradoxically, HBeAg can be detected in serum of patients with genotype G infection which seems to originate from coinfecting genotype A strains (Tran et al., 1991; Kato et al., 2002a,b).

Genotypes show a specific geographic distribution (Norder et al., 1993, 2004; Lindh et al., 1997): genotype A genomes are prevalent in the USA, in Northern and Middle Europe, and South Africa; genotypes B and C in the Far East; genotype D predominates in the Mediterranean area, and in the Near and Middle East; genotype E in Africa; and genotype F is most prevalent in South and Central America. Because genotype F is largely restricted to South and Central America and is almost unique among the Amerindians this type of HBV is probably indigenous to the native population of the New World (Blitz et al., 1998; Nakano et al., 2001). All strains of the recently identified genotype H, which is closely related to genotype F, are derived from North and Central America (Arauz-Ruiz et al., 2002). The existence of genotype G has been noted in 2000 (Stuyver et al., 2000). However, genotype G-specific sequences were first described in the early 1990s (Bhat et al., 1990; Tran et al., 1991), but the virus was not recognised as a new genotype. Genotype G is found in Europe and North America and all known genomic sequences are closely related to each other (Kremsdorf et al., 1996; Stuyver et al., 2000; Kato et al., 2002b; Vieth et al., 2002), suggesting either an epidemiological link between the isolates or high genetic stability of this viral genotype. Genotype G was isolated from white males, some of whom were homosexual and HIV-coinfected (Bhat et al., 1990; Kato et al., 2002a; Vieth et al., 2002). Whether there is a specific route of transmission of genotype G remains to be studied.

The geographic distribution of genotypes and strains can change due to human population migration or due to a specific risk behaviour. In the USA, genotypes B and C result from immigration from Asia, while genotype A is prevalent among white and black patients (Chu et al., 2003a). Another example for a change in the molecular epidemiology of HBV is the widespread dissemination of a regional HBV strain in England, which was primarily associated with injection drug use (Hallett et al., 2004).

Accumulation of sequence information and analysis of complete genomes currently allows the sub-division of genotypes into several sub-genotypes: A1 to A3; B1 to B4; C1 to C4; D1 to D4; F1 and F2 (Norder et al., 2004; Kurbanov et al., 2005). Patients can be coinfected with different genotypes, for example G+A or D+A, and the predominant genotype may shift during infection (Tran et al., 1991; Günther et al., 1992; Bahn et al., 1997; Gerner et al., 1998; Kato et al., 2002a,b). Genotype coinfection raises the possibility of recombination between genotypes. There is indeed evidence for recombination involving genotypes A/C, A/D, A/E, B/C, and B/D (Bowyer and Sim, 2000; Hannoun et al., 2000b; Fares and Holmes, 2002; Kurbanov et al., 2005).

The history of diversification of HBV into human genotypes and HBV of non-human primates is poorly understood, and a number of controversial hypotheses have been put forward (Bollyky and Holmes, 1999; Simmonds, 2001; Fares and Holmes, 2002). One hypothesis is based on the rate of neutral nucleotide substitutions in the HBV genome in infected patients (10^{-5} to 10^{-4} per position and year) (Okamoto et al., 1987; Hannoun et al., 2000a). Using this estimate as a phylogenetic clock, it was calculated that HBV genotypes diverged less than 6000 years ago (Orito et al., 1989; Fares and Holmes, 2002).

Although the existence of genotypes has been known for a long period of time, only recently evidence was obtained for an association of genotype and clinical course. Genotypes B and C have been studied most extensively due to their co-circulation in Asia, eliminating differences in ethnic or racial background of the patients. These studies show that genotype B, compared to genotype C, is associated with a higher rate of seroconversion from HBeAg to anti-HBe, less active liver disease, and a lower rate of progression to cirrhosis (Lindh et al., 1999; Kao et al., 2000; Orito et al., 2001; Chu et al., 2002; Sumi et al., 2003). Correspondingly, genotype C is more prevalent among HBeAg-positive patients than genotype B (Chu et al., 2003a). Similarly, compared to genotype D, genotype A is more prevalent in HBeAg-positive than in anti-HBe-positive patients (Li et al., 1993; Rodriguez-Frias et al., 1995; Mangia et al., 1996; Chu et al., 2003a). One possible explanation is the replacement of genotype A by coinfecting genotype D during infection, as has been observed along with HBeAg or HBsAg seroconversion (Günther et al.,