A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness

Lieven Stuyver,† Sija De Gendt, Caroline Van Geyt, Fabien Zoulim, Michael Fried, Raymond F. Schinazi and Rudi Rossau

1 Innogenetics NV, Industriepark 7, Box 4, B-9052 Gent, Belgium
2 INSERM U271, Lyon, France
3 University of North Carolina, Chapel Hill, NC 27599, USA
4 Emory University School of Medicine and Veterans Affairs Medical Center, Decatur, GA 30033, USA

The hepatitis B virus (HBV) genotype was determined in a total of 121 plasma samples collected in France and the US from patients chronically infected with HBV. HBV genotype A was predominant in this collection, appearing in 66 samples (54%), while genotypes B, C, D, E and F occurred in 4 (3%), 14 (12%), 23 (19%), 1 (1%) and 0 (0%) of samples, respectively. However, the genotype of a total of 13 (11%) samples (2 from France, 11 from the US) could not be determined with the methodology used. Sequence analysis, and subsequent phylogenetic analysis of the complete genome and the individual open reading frames, showed that the virus isolate from these samples was 3248 bp long and, phylogenetically, did not cluster with any of the known genotypes. This strain was provisionally called HBV genotype G. Virus isolates that were obtained from geographically separated regions like France and the US were closely related to each other. All virus strains analysed contained some characteristic differences when compared to genotype A: a translational stop codon at aa 2 and 28 of the preCore region; a 36 nt (12 aa) insert in the amino-terminal part of the Core antigen (HBcAg); a 2 aa deletion in the carboxy-terminal part of HBcAg; and a 1 aa deletion in the preS1 open reading frame. The deduced amino acid sequence of HBsAg suggests that this newly discovered genotype G strain belongs to serological group adw2.

Introduction

Human hepatitis B virus (HBV), which is the prototype member of the family Hepadnaviridae, is a circular, partially double-stranded DNA virus of approximately 3200 nt (Magnius & Norder, 1995). This highly compact genome contains the four major open reading frames (ORFs) encoding the envelope (preS1, preS2 and surface antigen HBsAg), core (preCore precursor protein, HBeAg and HBcAg), polymerase (HBPol) and X (HBX) proteins, respectively.

By using subtype-specific antibodies against HBsAg, nine different serological subtypes were defined, reflecting the genetic variability of HBV. Of the defined determinants, one is common to all subtypes (a determinant), but also two pairs of mutually exclusive subdeterminants (d or y, and w or r) were commonly found. By using this tool in epidemiological studies, nine serological subtypes have been identified: ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adrq + and adrq — (Swenson et al., 1991; Blitz et al., 1998).

Genotypically, HBV genomes have been classified into six groups, designated A–F, based on an intergroup divergence of 8% or more in the complete nucleotide sequence (Okamoto et al., 1988; Norder et al., 1992; Magnius & Norder, 1995). These six different genotypes show a characteristic geographical distribution: genotype A is pandemic, but most prevalent in north-west Europe, North America and Central Africa; genotype B is mostly found in Indonesia, China and Vietnam; genotype C is found in East Asia, Korea, China, Japan,
Polynesia and Vietnam; genotype D is also more or less pandemic, but is predominant in the Mediterranean area and the Middle East extending into India; genotype E is typical for Africa; and genotype F is found in American natives and in Polynesia (Van Geyt et al., 1998; Magnus & Norder, 1995). Some studies have shown that, in certain populations where HBV is endemic, a higher variability of HBV might be expected (Bowyer et al., 1997; Carman et al., 1997). However, in areas where HBV is not recognized as endemic, less HBV genotypic data are available. In the US for example, there are an estimated 1–1.25 million chronically infected persons, but these HBV infections are most often associated with groups at high risk (intravenous drug users, those with a history of other sexually transmitted diseases, prisoners, others; Alter & Shapiro, 1998). There is a paucity of data concerning the distribution of HBV genotypes in North American infected persons.

In this study, we report the HBV genotype prevalence in Atlanta (Georgia, USA) and Lyon (France) and describe a complete genome sequence of a new human HBV genotype, provisionally named genotype G. This genotype was found in patients chronically infected with HBV.

**Methods**

**Sample collection.** A total of 121 HBV-positive plasma samples was collected in France (n = 39) and the US (n = 82); the samples were divided into aliquots and stored at −20 °C until use. Samples were taken from chronic HBV carriers and were randomly selected as they became available.

**HBV DNA extraction and amplification.** HBV DNA was extracted from 100 µl serum samples using the High Pure PCR Template Preparation kit (Boehringer Mannheim) essentially as previously described (Stuyver et al., 1999). The complete genome of HBV was amplified using the Expand High Fidelity PCR system (Boehringer Mannheim). The amplification was performed on 5 µl extracted DNA with the primers HBPr108 and HBPr109 (Table 1). A 45 µl reaction mix was made, containing 5 µl 10 × Expand High Fidelity PCR system buffer, 2.5 U Expand High Fidelity PCR system enzyme mix, 200 µM dNTPs, 300 mM of each primer and sterile H₂O. Amplification was performed with denaturation at 94 °C for 40 s, annealing (after shifting to 60 °C in 50 s) for 1 min and elongation (after shifting to 72 °C in 15 s) for 4 min, with an increment of 5 s/cycle (Günther et al., 1998).

Two shorter PCR fragments were also generated: (i) the first amplicon covered the preS1, preS2 and HBsAg region and was amplified using primers HBPr1 and HBPr135 (outer PCR) followed by a nested reaction using HBPr2 and HBPr94 (Table 1); and (ii) the second amplicon covered the preCore/Core region and was amplified by means of a semi-nested set of PCR primers (HBPr86 and HBPr303, followed by HBPr87 and HBPr303). Outer PCR amplified the viral DNA over 40 cycles, with denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s and elongation at 72 °C for 30 s. Samples negative in first-round PCR were further amplified with nested PCR primers for 35 cycles with the same thermal profile.

**Sequencing.** Sequencing was performed on an automated DNA sequencer ABI 377 (PE Applied Biosystems), using fluorescence-labelled dideoxynucleotide chain terminators (ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq DNA polymerase).

### Table 1. Overview of the amplification and sequencing primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer sequence 5′–3′</th>
<th>Polarity*</th>
<th>Domain</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBPr1</td>
<td>GGGTCACCATTATTTCTTGGG</td>
<td>S</td>
<td>HBPol</td>
<td>2850–2868</td>
</tr>
<tr>
<td>HBPr2</td>
<td>GAAACAAGACTACAGCATGGGG</td>
<td>S</td>
<td>HBPol/preS1</td>
<td>2867–2888</td>
</tr>
<tr>
<td>HBPr3</td>
<td>CCACTGATGGGGCTGAGATG</td>
<td>AS</td>
<td>preS1/S2/HBPol</td>
<td>3226–3246</td>
</tr>
<tr>
<td>HBPr7</td>
<td>CTCACAG(A/T)AGCTCAATTC</td>
<td>AS</td>
<td>Core</td>
<td>1958–1978</td>
</tr>
<tr>
<td>HBPr14</td>
<td>TGGGGTGGAGCCCTCAG</td>
<td>S</td>
<td>preS1/HBPol</td>
<td>3104–3120</td>
</tr>
<tr>
<td>HBPr33</td>
<td>CTAGGGCTCCACCCCA</td>
<td>AS</td>
<td>preS1/HBPol</td>
<td>3104–3120</td>
</tr>
<tr>
<td>HBPr86</td>
<td>ACATAAGAGAGCTTCTTGGAC</td>
<td>S</td>
<td>HBX</td>
<td>1652–1671</td>
</tr>
<tr>
<td>HBPr87</td>
<td>TACCTCAAGAGCATTGTTTGA</td>
<td>S</td>
<td>HBX</td>
<td>1704–1723</td>
</tr>
<tr>
<td>HBPr94</td>
<td>GGTCT(A/T)AAAGGACTAC/AGATG</td>
<td>AS</td>
<td>HBPol/HBsAg</td>
<td>775–795</td>
</tr>
<tr>
<td>HBPr108</td>
<td>TTTTTCACCTCTGGCTAATC</td>
<td>S</td>
<td>HBX/preCore</td>
<td>1821–1840</td>
</tr>
<tr>
<td>HBPr109</td>
<td>AAAAACCTTCATGTTGCTCGG</td>
<td>AS</td>
<td>HBX/preCore</td>
<td>1806–1825</td>
</tr>
<tr>
<td>HBPr110</td>
<td>CCTCTGGCAATCTACTTCTGCA</td>
<td>S</td>
<td>HBPol</td>
<td>1255–1279</td>
</tr>
<tr>
<td>HBPr111</td>
<td>CTGAGGGCTCCAGCCCGAA</td>
<td>AS</td>
<td>Core/HBPol</td>
<td>2406–2430</td>
</tr>
<tr>
<td>HBPr113</td>
<td>CCGGCAATGAGAAGGCCAGACGG</td>
<td>AS</td>
<td>HBX/HBPol</td>
<td>1549–1574</td>
</tr>
<tr>
<td>HBPr134</td>
<td>TGCTCTAGGCTCACCTTTC</td>
<td>S</td>
<td>HBPol/HBsAg</td>
<td>414–433</td>
</tr>
<tr>
<td>HBPr135</td>
<td>CA/GAGACAAAAAAATTTGG</td>
<td>AS</td>
<td>HBPol/HBsAg</td>
<td>803–822</td>
</tr>
<tr>
<td>HBPr303</td>
<td>CCCACCTATAGCTCCTACCT</td>
<td>S</td>
<td>HBPol</td>
<td>2493–2512</td>
</tr>
<tr>
<td>HBPr374</td>
<td>GTTCGCGATTATCGAGCGCAGAG</td>
<td>AS</td>
<td>HBPol</td>
<td>1255–1279</td>
</tr>
<tr>
<td>HBPr440</td>
<td>TATGGAATGTTGTTATTGG</td>
<td>S</td>
<td>HBPol/HBsAg</td>
<td>738–758</td>
</tr>
<tr>
<td>HBPr446</td>
<td>GGAGTGTGATTTGCTGATCC</td>
<td>S</td>
<td>Core</td>
<td>2303–2323</td>
</tr>
<tr>
<td>HBPr448</td>
<td>CCCATGCTGATCTTTGCTTC</td>
<td>AS</td>
<td>HBPol/preS1</td>
<td>2868–2888</td>
</tr>
</tbody>
</table>

* S, Sense primer; AS, antisense primer.
FS; PE Applied Biosystems). The primers used for sequencing the complete genome are summarized in Table 1. The primers used to sequence the shorter PCR fragments are the same as the amplification primers.

**Data analysis.** Phylogenetic trees were created and analysed by distance matrix comparison using DNADIST, Neighbor and Drawgram software programs of PHYLIP version 3.5c (Felsenstein, 1993). Statistical analysis (t-test for the mean) was performed using the MedCalc program (Mariakerke). A P value of 0.05 was considered to be statistically significant. Complete genome sequences representing the different genotypes were retrieved from GenBank; their accession numbers are indicated in Fig. 2. In addition, the sequences from Tran et al. (1991) describing the core (HBcAg) and surface antigen (HBsAg) of an atypical variant of HBV were also included (M74499 and M74501).

**Results**

**Prevalence of the different HBV genotypes**

A total of 121 serum samples was collected and genotyped by using a research version of the HBV genotyping Line Probe Assay (LiPA) (Van Geyt et al., 1998). The results are summarized in Table 2. A very typical but previously unrecognized reactivity pattern was obtained for two samples from Europe and 11 samples from the US: hybridization reactions were observed on probe 140 (specifically designed for genotype A), probe 148 (designed for genotypes A and B), probe 80 (designed for genotypes C, D and E) and probe 239 (designed for genotypes B and E) (Van Geyt et al., 1998). Because this mixed hybridization pattern did not allow a unique type recognition, further characterization was needed via sequence analysis.

**Genetic relatedness to other genotypes**

One European virus isolate (FR1) was selected for whole genome sequencing, whereas the preCore/PreC and S genes were sequenced from another seven samples.

In order to compare the genetic relatedness of the FR1 strain with 36 other complete HBV genomes, homology percentages, as well as phylogenetic distances, were calculated.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td>Georgia, USA</td>
<td>46</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>82</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>4</td>
<td>14</td>
<td>23</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>121</td>
</tr>
</tbody>
</table>

54% (121 (100%))

**Table 2. Summary of the genotyping result obtained from 121 individuals infected with HBV**

**Fig. 1.** Phylogenetic distances, obtained by the program DNADIST, between the different HBV genotypes. In total, 36 complete genomes (accession numbers given in Fig. 2) were compared with the FR1 (genotype G) sequence. The mean for each group is indicated (A, ■; B, ○; C, ▲; D, ◆; E, •; F, ◆; G, ♦) including the standard error of the mean; the latter gives a 95% confidence interval around these distances.
The FR1 sequence showed an average homology of: 87·1% (min. 86·2%, max. 87·7%) with genotype A (five sequences); 86·6% (min. 86·5%, max. 86·6%) with genotype B (four sequences); 86·5% (min. 86·0%, max. 87·0%) with genotype C (14 sequences); 86·9% (min. 86·3%, max. 87·3%) with genotype D (eight sequences); 88·3% (min. 88·2%, max. 88·4%) with genotype E (two sequences); and 84·7% (min. 84·5%, max. 84·8%) with genotype F (three sequences). Phylogenetic distances between the six recognized HBV genotypes (36 sequences) were compared to each other and to the FR1 strain (Fig. 1). A clear difference emerged between the phylogenetic distances (i) within one genotype (distance range of 0·01–0·06) and (ii) between different genotypes (distance range 0·08–0·17). Using a t-test for the mean and a distance of 0·08 as a border value between genotypes, FR1 was found to be significantly different from genotypes A–F (range 0·11–0·17; P < 0·019). Phylogenetic trees of the complete genome sequences (Fig. 2 A), as well as of the individual ORF (Fig. 2 B for the surface region (preS1, preS2, HBsAg); data not shown for the other ORFs) were constructed, illustrating that FR1 is indeed located on a separate branch. Fig. 2 (B) further illustrates that the HBV samples from the US and France, including the B1 isolate, are closely related to each other. Based on these calculations and illustrated by means of phylogenetic trees, FR1 and the ten other virus strains shown belong to a new HBV genotype (called genotype G).

Characterization of the genome structure and ORFs

The complete genome structure of FR1 was similar to that described for the known HBV genotypes, but was found to be 3248 bp long (Fig. 3).

The preCore region has translational stops at codon 2 (TAA instead of CAA) and codon 28 (TAG instead of TGG) in all eight isolates sequenced. Based on the presence of this dual stop codon, the presence of HBeAg is generally not expected. Paradoxically, sample FR2 showed the presence of HBeAg in the plasma. The Core region is 585 nt long and encodes a Core protein of 195 aa (Fig. 3). In contrast to the other genotypes, the Core region had a nucleotide insert of 36 bp, located after the fifth nucleotide following the Core translation initiation point (A at position 1901). Genotype G, as well as all other genotypes except A, showed a 6 nt deletion in the carboxy-terminal part of the HBCag ORF (Figs 3 and 4).
The preS1 region contains 354 bp (118 aa), the preS2 region 165 bp (55 aa) and the HBsAg region 678 bp (226 aa) (Fig. 5). Like genotype E, genotype G strains showed a 3 nt deletion (1 aa at position 11; Fig. 5) in the amino-terminal part of preS1. Based on the presence of a lysine (K) at HBsAg position 122, a lysine (K) at position 160 and a proline (P) at position 127, the serological subtype of the genotype G strain was predicted to be adw2.

The HBPol region of this genotype contains 2526 bp (842 aa). The deletion in the carboxy-terminal part of HbcAg, as well as the deletion at the amino terminus of preS1 affects the numbering of the HBPol protein. The methionine residue which is prone to changes during lamivudine therapy is located at position 549 in the highly conserved YMDD motif (Bartholomeusz et al., 1998). Fig. 3 also shows the exact numbering for this methionine residue in the other genotypes.
L. Stuyver and others

Fig. 5. Amino acid sequence alignment of the preS1, preS2 and HBsAg open reading frame of the different HBV genotypes. For each genotype, only one representative genome is included (genotype A, X70185; B, D00331; C, X01587; D, X72702; E, X75664; F, X75663; G, FR1). Top, preS1; middle, preS2; bottom, HBsAg. The amino acid sequence was derived from the nucleotide sequence.

Discussion

Based on numerous prevalence studies, a fairly complete picture of the extent of HBV infections world-wide is available (Alter & Shapiro, 1998; Xu et al., 1998; Attia, 1998; Carrilho & Corrêa, 1998; Alvarado-Esquivel et al., 1998). An ‘individual genetic group’ (phylogenetically different genotype) of HBV was defined if a certain virus strain differed by more than 8% from all other HBV genomes (Okamoto et al., 1988; Norder et al., 1994), or 4–1% divergence when comparing S gene sequences (Norder et al., 1992). In the context of the findings described in this manuscript, there might be a need to further differentiate between ‘genetic variants’ versus ‘genotypes’. Virus strains with, for example, core insertions or preS1 deletions are considered to be genetic variants of known genotypes. Virus strains differing by more than 8% over their entire genome from other isolates (not focussing on possible deletions or insertions) belong to another genotype. Based on results obtained from several epidemiological studies (Blitz et al., 1998; Norder et al., 1994; Magnus & Norder, 1995; Telenta et al., 1997; Alvarado-Esquivel et al., 1998), six major genetic groups of HBV were recognized. The HBV genotype G strain was not found in any of these studies. In this manuscript, HBV genotype prevalence was studied on samples from chronic HBV carriers from Atlanta, USA, and Lyon, France. Apart from the detection of the commonly found genotypes (genotypes A–D), a new HBV strain with a minimum of 11.7% (genotype G versus E) and a maximum of 15.3% divergence over the complete genome was found. The prevalence of this viral variant in the US exceeded 11% of all infections. As with all other HBV genotypes, the particular clinical outcome and the epidemiological features of the genotype G strain need further attention.

In addition to the prominent prevalence of genotype G virus in the US samples, this genotype was also found in samples originating from France. The prevalence in other geographical areas needs further exploration. The complete HBV genome sequence that is presented in this study was determined from a sample taken from a chronic but asymptomatic carrier living in the Lyon area, France. The homology of this FR1 strain with the previously published B1 sequence (another French isolate), which covers the complete S gene as well as the preCore and Core genes, is surprisingly high (Tran et al., 1991; Fig. 2B). This atypical HBV strain was already detected and partially sequenced in 1991, but never recognized as a new genotype. In order to study the genotype G sequence with respect to recombination events (as described by Tran et al., 1991), the complete FR1 genome was inspected for such events at the nucleotide level and phylogenetically for co-segregation with other known genotypes; evidence for recombination was not found. Despite the absence of recombination in FR1, recombination events in HBV were presented as a more common event than previously thought (Tran et al., 1991; Bollyky et al., 1996).
The expression of HBeAg in HBV genotype G needs additional attention. In all isolates studied here (geographically unrelated samples were included), a virus variant with two preCore translational stop codons (TAA at codon 2, TAG at codon 28) was found. This observation suggests that the dual variant is a naturally occurring configuration for genotype G. Due to the presence of an insert in the Core, the stability of the encapsidation signal (Lok et al., 1994) might be altered. There might be a need for a compensatory change, possibly resulting in selection for this dual variant. As a consequence, however, this finding makes all these viruses incapable of expressing the HBeAg. Paradoxically, HBeAg was found in at least one French patient (FR2). Similar observations were also made for HBeAg. Paradoxically, HBeAg was found in at least one subgenotype C in Latin America. In Therapies for Viral Hepatitis, pp. 25–34. Edited by R. F. Schinazi, J.-P. Sommadossi & H. C. Thomas. London, UK: International Medical Press.

In this study, evidence was provided for the existence of a seventh HBV genotype, called genotype G. The virus structure is essentially identical to that of the other genotypes, but has some unique features (like an insert of 36 nt in the core region). The prevalence of this strain was found to be more than 11% in the Georgia area, USA, but needs to be further determined for other geographical regions. These findings may have an impact on the immunological and genetic diagnosis of HBV, as well as on the treatment of the ubiquitous disease it causes.

The authors thank Dr Hans Pottel (Innogenetics) for assistance with the statistical analysis, and Ann Meert for excellent technical support. The suggestions from Fred Shapiro were very useful for manuscript editing. R.F.S. is partially supported by the Department of Veterans Affairs.

References


Received 28 June 1999; Accepted 23 August 1999