Characterization of hepatitis B virus genotypes among Yucpa Indians in Venezuela

Tatsunori Nakano, Ling Lu, Xiaolei Hu, Masashi Mizokami, Etsuro Orito, Craig N. Shapiro, Stephan C. Hadler† and Betty H. Robertson

1 Hepatitis Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA
2 Second Department of Medicine, Nagoya City University Medical School, Nagoya 467-8601, Japan

The complete genome sequences of hepatitis B virus (HBV) from 12 HBV-infected Yucpa Indians of Venezuela, a group with highly endemic HBV, were amplified and sequenced. The 12 isolates were closely related to each other, with 98.6–100% nucleotide identity. A phylogenetic tree based on the complete genome indicated clearly that they were genotype F. Three individuals had evidence of infection with two different HBV deletion mutants. In two individuals, a three amino acid deletion was identified just prior to the 'a' determinant loop of the S region. A third individual was infected with virus that contained a complete core reading frame and a population that contained a deletion in the middle of the core region. These results indicate that genotype F HBV is present in the Venezuelan Yucpa Amerindians and the complete genome sequence allowed the identification of two unique deletion mutants in a limited set of samples.

Hepatitis B virus (HBV) strains have been classified into seven genomic groups or genotypes, designated A to G (Okamoto et al., 1988; Norder et al., 1992; Naumann et al., 1993; Stuyver et al., 2000). The predominant HBV genotypes cluster in geographical regions where the populations are known to have high carrier rates resulting from perinatal and childhood transmission. Genotype F HBV is found in South and Central America (Norder et al., 1993; Arauz-Ruiz et al., 1997) and is genetically divergent from the other human HBV genotypes. There are only three full genome sequences of genotype F strains that have been reported. One is from a Brazilian blood donor in Rio de Janeiro (Naumann et al., 1993), while the other two are from chronic HBV-infected individuals from Colombia and France (Norder et al., 1994).

It has been estimated that one-third of the projected 10 million HBV carriers in South America reside in and around the Amazon basin, including parts of Colombia, Venezuela and Peru (Fay, 1994). A majority of the infections in these regions are among indigenous Amerindians, with reported carrier rates of 5–30% (Hadler et al., 1984; Torres & Mondolfi, 1991; Torres, 1996; Blitz-Dorfman et al., 1996). The Yucpa are an indigenous group living in the Perija Mountains in western Venezuela on the border with Colombia. The Yucpa have high mortality from fulminant and rapidly progressive chronic hepatitis due to hepatitis delta virus (HDV) superinfection among the HBV carriers in this high HBV endemic population. Both HBV and HDV infection occur primarily in childhood and early adult life (Hadler et al., 1984).

We determined complete HBV genome sequences from the sera of 12 Yucpa individuals (Table 1). The sera were obtained during serosurveys conducted during the 1980s to determine the prevalence of HBV infection in the area (Hadler et al., 1984, 1992). These serum samples were all positive for hepatitis B surface antigen (HBsAg) and hepatitis B e antigen (HBeAg) in commercial radioimmunoassays (Abbott Laboratories). DNA was extracted from 50 µl serum with the MasterPure DNA purification kit (Epicentre). A full-length 3.2 kb HBV genome product was amplified by PCR by using primers P1 and P2, as described by Günther et al. (1995). This product was used to generate seven overlapping fragments by nested PCR amplification with primers and conditions designed for chimpanzee HBV amplification (Hu et al., 2000) in addition to P1 and P2. Amplicons were purified by using a QIAquick purification kit (Qiagen) and sequenced by direct sequencing using dRhodamine terminators and a 377A DNA Sequencer (Applied Biosystems) by using the approach described by Hu et al. (2000). Sequences of one fragment (nt 56–1120) of VNZ8248 and VNZ8339 and another fragment (nt 1821–2839) of VNZ8375 could not be determined by direct sequencing of

[Author for correspondence: Betty H. Robertson. Fax +1 404 639 1563. e-mail bjr1@cdc.gov]

† Present address: National Immunization Program, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.

The GenBank accession numbers of the sequences reported in this paper are AB036905–AB036920.

SHORT COMMUNICATION
Table 1. Samples used for nucleotide sequencing

<table>
<thead>
<tr>
<th>Sample</th>
<th>Area</th>
<th>HBeAg</th>
<th>HBsAg</th>
<th>S region deletion</th>
<th>C region deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNZ8248/VNZ8339*</td>
<td>Central</td>
<td>+</td>
<td>+</td>
<td>+/−†</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8251</td>
<td>North</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8255</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8323</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8337</td>
<td>Central</td>
<td>+</td>
<td>+</td>
<td>+‡</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8346</td>
<td>Central</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8349</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8351</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8371</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8375</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>+/−†</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8381</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>VNZ8624</td>
<td>South</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

* VNZ8248 and VNZ8339 were collected from the same person at different times.
† VNZ8248/VNZ8339 and VNZ8375 had mixed virus populations with and without deletions.
‡ VNZ8337 had the same deletion as VNZ8248/VNZ8339.

PCR amplicons. For sequencing of these fragments, the PCR amplicons were ligated and cloned into the plasmid vector pCR II using a TA cloning kit (Invitrogen). To investigate the relationship between the sequences obtained in this study and previously reported strains, algorithms in the GCG package (Genetics Computer Group, Madison, WI, USA) were used for alignment and sequence comparison. Alignment was performed with PILEUP, followed by construction of a neighbour-joining phylogenetic tree based on genetic distances calculated using Kimura’s two-parameter method (Kimura, 1980) within GROWTREE.

The 12 strains studied were closely related to each other, with 98.6–100% identity between the sequences. This close relationship is evident in the phylogenetic tree containing Venezuelan HBV sequences and other complete HBV genome sequences that are available (Fig. 1) and is consistent with circulation of closely related viruses within this relatively isolated population. The sequences are most closely related to a genotype F sequence derived from Colombia, with approximately 97% identity.

Three sets of sequences (VNZ8248-1 and -2, VNZ8339-1 and -2 and VNZ8375-1 and -2) are from two individuals (Fig. 1). VNZ8248 and VNZ8339 are serum samples collected from the same person at two different times. In both of these samples, two different sequences were detected in the S region encoding the HBsAg protein. Ten clones were evaluated for this region of VNZ8248 and VNZ8339 and five and four clones, respectively, had a deletion (labelled VNZ8248-2 and VNZ8339-2 and described in more detail below), while the remaining five and six clones (VNZ8248-1 and VNZ8339-1) did not (Table 1). The respective -1 and -2 sequences were identical in both samples collected at different times, ruling out adventitious contamination. The other individual, VNZ8375, also had a mixed virus population with and without a core region deletion (described in more detail below). Ten clones were evaluated for this region of VNZ8375 and three had the deletion (labelled VNZ8375-2), while the remaining seven clones did not (VNZ8375-1) (Table 1).

The amino acid sequences for the HBsAg protein (Fig. 2a) from the Venezuelan strains were highly conserved, with only five positions of three strains containing amino acid changes (Leu→Pro, Leu→Val, Gln→Arg, Leu→Thr, Pro→Asn). All Venezuelan sequences had Gly145, Lys122, Leu127 and Lys169 which are the amino acid determinants for subtype adw4 (Ashton-Rickardt & Murray, 1989; Okamoto et al., 1987; Naumann et al., 1993; Norder et al., 1994; Arauz-Ruiz et al., 1997). As mentioned above, in one individual, we identified two sequences with (VNZ8248-2/VNZ8339-2) and without (VNZ8248-1/VNZ8339-1) a nine nucleotide deletion in the S region gene. This mutation resulted in a three amino acid deletion just prior to the ‘a’ determinant region of HBsAg. In another individual from the same geographical region (VNZ8337), this deletion mutant was detected as a single infection, without the virus containing the complete S region genome (Table 1).

Published reports regarding double infection of HBV are limited and are subject to limitations. Some studies have relied upon subtype differences to document mixed infections; however, subtype determinants are dictated by Lys/Arg differences at single positions and therefore a single nucleotide change would result in different subtypes. Different subtypes in this context may reflect the same master virus population and the differences then reflect a mixed population of Lys- and Arg-containing viruses (van Kok-Doorschot et al., 1972; Paul et al.,
Fig. 1. Phylogenetic tree constructed by the neighbour-joining method, based on the complete sequences of HBV strains. Venezuelan sequences determined in this study are indicated by the VNZ prefix. Samples in which two sequences were detected are designated -1 and -2. The previously published sequences used in this comparison are indicated by their accession numbers. The scale bar indicates substitutions per site.
Fig. 2. Amino acid sequences of Venezuelan HBsAg protein (a) and the core protein (b) predicted from nucleotide sequences and published sequences that contain deletions in the same region. Dashes indicate identity to the consensus sequences and changed amino acids are indicated by the appropriate single letter code. Shaded regions indicate positions of deletions that were detected in this study or from previously published data. Stop codons are indicated by *.

(a) Ten sequences from Venezuela (VNZ8248-1/VNZ8339-1, VNZ8251, VNZ8255, VNZ8233, VNZ8349, VNZ8351, VNZ8375, VNZ8381 and VNZ8624) are identical and are shown as Consensus of Venezuela. VNZ8248-2/VNZ8339-2, VNZ8337 and VNZ8346 have sequences that differ from the consensus. X in VNZ8346 indicates Leu9 or Pro9 translated from the ambiguous nucleotide sequence. AJ003026, AJ003028 and AJ003027 are HBV sequences that were reported in an HBsAg-negative individual (Grethe et al., 1998). Proposed epitopes of antigenic regions of the hydrophilic core of HBsAg protein, HBs1 to HBs5, are shown. Residues Gly145, Lys122, Leu127 and Lys160, which are conserved among all Venezuelan sequences and are determinants of subtype adw4, are indicated by $g$. The ‘a’ determinant loop is shown by double underlining.

(b) Eleven sequences from Venezuela (VNZ8248/VNZ8339, VNZ8251, VNZ8255, VNZ8323, VNZ8346, VNZ8349, VNZ8351, VNZ8375, VNZ8375-1, VNZ8381 and VNZ6624) are identical and are shown as Consensus of Venezuela. VNZ8337 and VNZ8375-2 have sequences that differ from the consensus. A BLAST search identified 17 sequences within GenBank/DDBJ/EMBL, L42358–L42366 (Guńther et al., 1996) and AF143298–AF143307 (Preikschat et al., 1999), that had deletions in approximately the same region of the core. X indicates frame shifts. The deletion of Venezuelan strains between 152 and 153 is generally observed in all genotypes except genotype A, and is not specific to Venezuelan strains. The major B cell epitope of HBcAg and HBsAg (HBc/e1) is shown.

On the other hand, the detection of two different genotypes based upon nucleotide sequencing (Yamanaka et al., 1990) suggests that true mixed infections may occur. There is no way to determine whether the presence of the two HBsAg sequences in the same individual (VNZ8248/VNZ8339) is the result of a double infection or the result of mutation during the course of infection. However, our data indicate that this S region deletion mutant is transmissible and can support virus replication, as it was the only sequence detected in one individual (VNZ8337). Antigenic epitopes in the hydrophilic central core of HBsAg, from residue 99 to 169, can be subdivided into five regions, HBs1 to HBs5 (Carman, 1997; Stirk et al., 1992). Amino acid substitutions in HBs1 or HBs5 have been proposed to result in a lack of reactivity in some HBsAg assays (Carman et al., 1997). The S region deletion that we identified in two individuals did not cause a frame shift or termination but was located in HBs1 (Fig. 2a; highlighted in the underlined region). Deletions in the same region of HBsAg have been described in an HBsAg-negative individual whose virus was genotype D (Fig. 2a) (Grethe et al., 1998). However, the ‘a’ determinant (HBs3 and HBs4 regions) of this virus contained numerous other mutations that could affect the immunological reactivity. The sample in which we detected an S region deletion had no detectable HBsAg serological differences from the other samples in this collection.

The amino acid sequence of the precore and core region (Fig. 2b) from the Venezuelan strains reveals that the amino acid sequences were well conserved, with a single amino acid change (Gln162 → Lys) found in one sequence. One individual (VNZ8375) had two distinct core sequences, one with a deletion (VNZ8375-2) and one without a deletion (VNZ8375-1). The VNZ8375-1 sequence is identical to the core sequence in all the other Venezuelan strains within this region. The
**Genotype F of HBV in Venezuela**

**Fig. 2.** For legend see facing page.
deletion of VNZ8375-2 (86 nucleotides) resulted in a frame shift and termination codon in the core region. These two strains, found in the same serum specimen, had identical sequences surrounding the deleted region, as determined by cloning. Moreover, the remaining genome sequence, determined by direct sequencing, had a single consensus sequence. This indicates that this mixed virus population did not result from a double infection and probably represents a de novo mutation within this individual.

Deletion mutants in this same region have been reported in other sequences (Fig. 2b) (Günther et al., 1996; Prekschat et al., 1999) and all of these deletion mutants co-exist with a sequence containing the complete core region. The majority of these deletions (which vary in length from eight to 79 amino acids) start near the major B cell epitope of hepatitis B core antigen (HBCAg) and HBeAg (HBC/e1). The HBV polymerase (P) reading frame overlaps the end of the core reading frame, and these core region deletions occur prior to the P gene start codon and do not affect the P gene reading frame. However, the deletions are predicted to interfere with the translational regulation of P protein expression by deletion of two conserved ATG codons prior to the P gene start ATG (Günther et al., 1999). In the current model of P protein translation, the presence of these two conserved start codons, located upstream of the P gene start ATG, are responsible for suppressing P gene translation (Fouillot et al., 1993; Hwang & Su, 1998). Most of the core region deletion mutants that have been reported, including the Venezuelan core region deletion mutant, have deletions that include these two ATGs (data not shown). Even though the relationship between the core region deletions and regulation of P protein translation has been elucidated in vitro, the clinical significance of these mutations is still unclear. The appearance of these mutants in immunocompetent individuals is commonly associated with a favourable outcome (Marinos et al., 1996; Nakayama et al., 1995) while, in the absence of a humoral immune response, the outcome is less favourable (Uchida et al., 1994; Zoulim et al., 1996; Günther et al., 1996), resulting in liver failure or progressive liver disease. Furthermore, the deletion mutants in patients with progressive liver disease reported by Günther et al. (1996) were genotypes A and B, and ours is the first report of a similar deletion mutant within genotype F HBV infections.

It has long been recognized that fulminant hepatitis in Amerindian populations is a major problem (Hadler et al., 1984) and there is a clear link between HDV superinfection and severe disease in this population. The interaction between HDV and HBV and the effect of HDV infection on the genome of HBV is an area that has not been explored. Further studies are needed to evaluate the role of the core region deletions and their relationship to the development of fulminant hepatitis observed in these populations.

Genotype F HBV strains have been proposed to be representative of the HBV strains of Amerindian populations of the New World (Norder et al., 1994). The ade4 subtype, which is encoded by genotype F, has been found in French Polynesia, in Argentina, in Venezuelan Cuiva Indians (Cournou-Pauty et al., 1983) and in other native populations from the Amazon basin (Gaspar & Yoshida, 1987). In this study, we have characterized 12 complete HBV sequences present in Venezuelan Yukpa Indians as genotype F and determined that the predicted subtype from the translated amino acid sequences is ade4. The sequences are most closely related to genotype F HBV derived from Colombia, a country bordering Venezuela. These 12 strains included two novel deletion mutants that have not been identified previously among genotype F strains. Our observations indicate that complete genome information, from such a limited population, may reveal unique HBV variants and provide valuable molecular information that may correlate with clinical disease features.

References


Received 7 September 2000: Accepted 3 November 2000

Published ahead of print (10 November 2000) in JGV Direct as DOI 10.1099/vir.0.17393-0