Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa

Gerald C. Kimbi, Anna Kramvis and Michael C. Kew

MRC/CANSA/University Molecular Hepatology Research Unit and Department of Medicine, University of the Witwatersrand, Parktown 2193, Johannesburg, South Africa

Phylogenetic analysis of hepatitis B virus (HBV) has led to its classification into eight genotypes, A to H. The dominant genotype in South Africa is genotype A, which consists of two subgenotypes, A1 and A2. Subgenotype A1 (previously subgroup A9) predominates over subgenotype A2 (previously subgroup A minus A9). The complete genome of HBV isolated from 18 asymptomatic carriers of the virus and five acute hepatitis B patients was amplified; the resulting amplicons were cloned and sequenced. All acute hepatitis isolates belonged to subgenotype A1 and had no distinguishing mutations relative to the isolates from asymptomatic carriers, which had a distribution of ten subgenotype A1, two subgenotype A2 and six genotype D. The presence of the previously described amino acid residues that distinguish subgenotype A1 (subgroup A9) from the remainder of genotype A in the S and polymerase genes was confirmed. Moreover, the large number of subgenotype A1 isolates sequenced allowed identification in the other open reading frames of additional nucleotide and amino acid changes that are characteristic of subgenotype A1. In particular, nucleotide mutations at positions 1809–1812 that alter the Kozak sequence of the precore/core open reading frame, and A1888 in the precore region, were found exclusively in subgenotype A1 isolates. Unique sequence alterations of the transcriptional regulatory elements were also found in subgenotype A1 isolates. The mean nucleotide divergence of subgenotype A1 was greater than that of subgenotype A2, suggesting that this subgenotype has been endemic for a longer time in the South African black population than had subgenotype A2.

INTRODUCTION

Hepatitis B virus (HBV), a DNA virus, is a member of the family Hepadnaviridae. Infection with HBV is a public health problem of worldwide importance, with some 387 million people being chronically infected (WHO, 2002). In sub-Saharan Africa the virus is hyperendemic, with a carrier rate of 5 to 20% (Mphahlele et al., 2002). HBV infection causes a spectrum of liver diseases, including subclinical, acute self-limited, and fulminant hepatitis, an asymptomatic carrier (ASC) state, chronic hepatitis progressing to cirrhosis, and hepatocellular carcinoma.

Hepadnaviruses have an unusual mechanism of viral DNA replication involving reverse transcription of pregenomic RNA by the virus-encoded polymerase (Nassal & Schaller, 1996). The virus polymerase lacks proofreading activity, and sequence heterogeneity is a feature of HBV. Phylogenetic analysis of HBV full-length genomes has led to the classification of HBV into eight genotypes. The separate genotypes are arbitrarily defined by an intergroup divergence in the complete HBV genome sequence of more than 8% (Norder et al., 1992a; Okamoto et al., 1988) and at the level of the S gene of more than 4% (Norder et al., 1992b). Early studies enabled the identification of four genotypes, A to D (Okamoto et al., 1988), with the genotypes E, F (Norder et al., 1992b, 1994), G (Stuyver et al., 2000) and H (Arauz-Ruiz et al., 2002) being identified later.

The eight genotypes show a distinctive geographical distribution. Genotype A is prevalent in northwestern Europe, North America and Africa (Norder et al., 1993). Genotypes B and C are characteristic of Asia (Okamoto et al., 1988), whereas genotype D has a worldwide distribution but predominates in the Mediterranean area. Genotype E is found in Africans (Odemuyiwa et al., 2001), genotype F in the aboriginal populations of South America (Arauz-Ruiz et al., 1997; Norder et al., 1993) and genotype H is confined to the Amerindian populations of Central America (Arauz-Ruiz et al., 2002). To date, the isolation of genotype G has been limited to HBV carriers in France and Georgia, USA (Stuyver et al., 2000) and Germany (Vieth et al., 2002).
Genotypes A and D coexist in southern Africa, with genotype A predominating. Furthermore, a unique segment of genotype A, subgroup A' (renamed subgenotype A1) has been identified in this region (Bowyer et al., 1997; Kramvis et al., 2002). This subgroup diverges from subgroup A minus A' (renamed subgenotype A2). Because subgenotype A1 is widespread in southern Africa it is important to study the influence of this subgenotype on virus replication and disease outcome (Mayerat et al., 1999). Our previous study on subgroup A' (subgenotype A1) was limited to isolates from five patients with fulminant hepatitis and one with acute hepatitis (AH) (Kramvis et al., 2002). Therefore, the aim of the present study was to carry out full genome analysis on HBV isolates from a larger number of AH patients and from ASCs of the virus. A comparison of these sequences with HBV isolates from various parts of the world and with various disease conditions may help us in understanding the pathogenesis of HBV-induced disease.

**METHODS**

Subjects and serum samples. Serum samples were collected from 275 hepatitis B surface antigen (HBsAg)-positive southern African blacks: 260 were ASCs and 15 AH patients. The ASCs were randomly selected, unrelated factory workers and labourers from the Gauteng Province. The AH patients were treated at the Johannesburg Academic Hospital as sporadic cases during recent years. These samples were stored at −70 °C until analysed. Commercially available kits (Abbott) were used to detect HBV markers in the serum. The study was approved by the Human Ethics Committee of the University of the Witwatersrand and informed consent was obtained from all subjects.

DNA extraction, amplification, cloning and sequencing. Total DNA was extracted from the sera using the QIAamp DNA Mini Kit (Qiagen), according to the manufacturer’s instructions. When the virus concentration was high enough, the complete genome of the virus was amplified using a single amplification method (Gunther et al., 1995) using primers P1 (5′-TTTTTCAACCTTGCTAATCA-3′) (1821–1843 from EcoRI site) and P2 (5′-AAAAAGTTGCACTGRTGTCG-3′) (1825–1801 from EcoRI site). However, when the virus load was too low for complete genome amplification using single-round PCR, a modification of two subgenomic PCRs was used (Takahashi et al., 1998); this involved the amplification of two overlapping fragments of HBV, fragment A (135 kb) and fragment B (2-2 kb) (Table 1). This PCR was designed so that the overlap occurred over the variable regions of the S and precore/X genes, which would allow us to conclude that the amplified DNA was from a single genome, when overlapping regions were identical.

The reaction mix for the amplification consisted of 2·25 µl 10× Ex Taq buffer with 20 mM MgCl2, 2 µl 2·5 mM dNTP mix, 1·25 µl each of the appropriate primers (Table 1), 2·5 µl DNA, made up to 22·5 µl with water. The enzyme mix was made up of 1·875 µl water, 0·25 µl 10× Ex Taq buffer and 0·375 µl TaKaRa Ex Taq polymerase. The 22·5 µl reaction mix was preheated to 94 °C for 2 min and 2·25 µl TaKaRa Ex Taq enzyme mix was added at the first annealing step. This was followed by 40 cycles of amplification with the cycling profile shown in Table 1.

Amplicons were cloned into a pPCR-Script Amp SK+ vector (Stratagene) according to the protocol provided by the manufacturer. The positive clones containing the correct size amplicons were prepared for direct sequencing using the BigDye Terminator v3·0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and sequenced on an Applied Biosystems 377 DNA automated sequencer using vector-specific primers T3 (5′-CTATCAGCTATATAGGGC-3′) and T7 (5′-CTATCAGCTATATAGGGC-3′) as well as HBV-specific primers (Owiredu et al., 2001). All sequences were analysed in both the forward and reverse directions.

Phylogenetic analyses. Complete HBV genomes sequences were compared with corresponding sequences of HBV from GenBank. Multiple sequence alignments were carried out using ClustalW (Thompson et al., 1994). The complete genomes of HBV isolates from 24 ASCs were sequenced fully, 20 using the single round Gunther method (Gunther et al., 1995) and four a modification of the method described by Takahashi et al. (2000). The complete genomes of five HBV isolates from

**RESULTS**

Amplification, cloning and sequencing

Using subgenomic nested PCR assays, 223 of 260 (86%) ASC sera and all 15 (100%) AH patient sera were HBV DNA positive. The complete genomes of HBV isolates from 24 ASCs were successfully amplified, 20 using the single round Gunther method (Gunther et al., 1995) and four a modification of the method described by Takahashi et al. (2000). The complete genomes of five HBV isolates from

Table 1. Oligonucleotide primers and polymerase chain reaction cycling profiles

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Primer</th>
<th>Position*</th>
<th>Sequence</th>
<th>Denaturation</th>
<th>Annealing</th>
<th>Extension</th>
<th>Size†</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>455(+):</td>
<td>455-474</td>
<td>AGTATGTTGGCGCTTG3&quot;</td>
<td>94 °C 30 sec</td>
<td>62 °C 30 sec</td>
<td>72 °C 90 sec</td>
<td>1345</td>
</tr>
<tr>
<td></td>
<td>1800(-):</td>
<td>1800-1773</td>
<td>AGCAATTTATGCTACGCTA3′</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1687(+):</td>
<td>1687-1708</td>
<td>CGACCGCCCTTGAAGCATA3′</td>
<td>94 °C 30 sec</td>
<td>63 °C 30 sec</td>
<td>72 °C 120 sec</td>
<td>2198</td>
</tr>
<tr>
<td></td>
<td>685(-):</td>
<td>704-685</td>
<td>GAAACCTGAAACATTGGAC3′</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Denotes the nucleotide position of hepatitis B virus adw genome (GenBank accession #V00866) where the EcoRI cleavage site is position 1.
†Size of the amplicons in base pairs.
AH patients were successfully amplified using the method of Takahashi et al. (2000). Of these, 18 amplicons from ASCs and the five from AH patients were successfully cloned and sequenced. These sequences have been deposited in GenBank/EMBL/DDJB databases as AY233274–AY233296. Subgenomic PCR and direct sequencing of the amplicons confirmed that the clones sequenced were representative of the major HBV strain in the serum.

Phylogenetic analysis

The length of the complete genomes of all South African (SA) genotype A isolates sequenced in the present study was 3221 bp. The serological subtype of all SA genotype A isolates, except AY233288, was deduced from the sequence to be adw2. Isolate AY233288 belonged to serological subtype ayw2. The complete genome sequences of the 23 HBV isolated sequences were aligned with 32 sequences from GenBank and phylogenetic analysis was carried out (Fig. 1). All AH isolates belonged to subgenotype A1 and had no distinctive mutations relative to the isolates from the ASCs. The HBV genotype distribution among the 18 isolates from ASCs was ten subgenotype A1, two subgenotype A2 and six genotype D. The isolates clustered in the same positions when phylogenetic analysis of the individual open reading frames (ORFs) was performed. There was no unique clustering of the isolates from ASCs when compared to those from AH patients.

The complete genomes of subgenotype A1 had a mean nucleotide divergence from subgenotype A2 of >4% (Table 2), validating the existence of a separate subgroup of genotype A. HBV isolates belonging to subgenotype A1 sequenced in this study had a genome length of 3221 bp and did not have the 21 bp deletion, which we had found previously in the pre-S1 region (Kramvis et al., 2002). This deletion was also absent from subgenotype A2 of the 23 HBV isolates sequenced in the present study was compared with sequences obtained from GenBank (Fig. 3). Nucleotides found in subgenotype A1 but not A2 are shown in bold, and those that are found predominantly in subgenotype A1 but not other genotypes are shown in bold and shaded. Table 3 summarizes the mutations in the cis-acting regulatory elements characteristic of subgenotype A1 and the functional elements that are affected by these mutations.

Comparison of amino acid sequences of subgenotype A1 to those of subgenotype A2 and other genotypes

Fig. 2 provides a comparison of the translated sequences of the 15 subgenotype A1 and two subgenotype A2 isolates, sequenced in the present study, to sequences obtained from GenBank. The amino acids found only in subgenotype A1 are shaded in grey, whereas those that are found in subgenotype A1 and in other non-A genotypes, but not in subgenotype A2, are shown in bold. Confirming our previous analyses, amino acids Gln54, Val74, Ala86 and Val91 in the pre-S1 region; Leu12 in pre-S2 and Thr236, Gly268, Tyr269, Gln334, Lys338 in the spacer of the polymerase gene were unique to subgenotype A1 (Bowyer et al., 1997; Kramvis et al., 2002). In addition, in the present study, we identified residues Ser11, Ala31, Ser47, Ser146, Ser147 in the X ORF to be characteristic of subgenotype A1. Residues His182 and Ser251 in the polymerase ORF are shared by subgenotype A1 and genotype E, the genotype found only in Africa.

By corollary, subgenotype A2 also has signature amino acids that distinguish it from subgenotype A1 and all other genotypes. These are circled in Fig. 2. They are Ala34, Ser89, Thr96 and Ile91 in the pre-S1; Ala77 in the pre-S2; Val209 in the S region. In the polymerase region, subgenotype A2 had the following unique amino acids: Gly18, Ala33 and Thr210 in the priming region and Val271, Asp273, Cys308 and Arg348 in the spacer of the polymerase.

Comparison of the nucleotide sequences of cis-acting elements of subgenotype A1 with those of subgenotype A2 and other genotypes

The sequences of the cis-acting elements of the 15 subgenotype A1 and 2 subgenotype A2 isolates sequenced in the present study were compared with sequences obtained from GenBank (Fig. 3). Nucleotides found in subgenotype A1 but not A2 are shown in bold, and those that are found predominantly in subgenotype A1 but not other genotypes are shown in bold and shaded. Table 3 summarizes the mutations in the cis-acting regulatory elements characteristic of subgenotype A1 and the functional elements that are affected by these mutations.

DISCUSSION

In our previous study we presented a phylogenetic analysis of four complete and three pre-S1/S2/S gene sequences of subgenotype A1 (previously called subgroup A’) HBV isolates from fulminant and AH patients from SA (Kramvis et al., 2002). Here we extend the study to isolates from both ASCs and additional AH patients with the analysis of 15 complete genomes belonging to subgenotype A1 and two to subgenotype A2 (Fig. 1). This allows confirmation of our previous findings regarding subgenotype A1 (Bowyer et al., 1997; Kramvis et al., 2002), as well as a more comprehensive analysis of subgenotype A1 from Africa and identification of features unique to this subgenotype.

All isolates belonging to subgenotype A1 sequenced in this study had a genome length of 3221 bp and did not have the 21 bp deletion, which we had found previously in the isolates from fulminant hepatitis patients (Kramvis et al., 2002). This deletion was also absent from subgenotype A1 isolates from chronic carriers from Malawi (Sugauchi et al., 2003).

The amino acids differentiating the subgenotypes of genotype A from each other and from other genotypes were concentrated in the pre-S1 region overlapping the spacer of the polymerase (Fig. 2). The sequence of the pre-S1 region has been shown to be well conserved within a given HBV subtype (Uy et al., 1992) and this region may play a role in the attachment of HBV to hepatocytes (Neurath et al., 1986; Pontisso et al., 1989). Therefore, it is possible that the molecular evolution of the pre-S1 sequence is constrained by the host population (Kramvis et al., 2002). It is of interest to note that Gln54 and Val91 in the pre-S1 region and Thr236 in the spacer of the polymerase, unique to subgenotype A1 isolates, are also found in the aberrant
Table 2. Mean nucleotide divergence (%) of complete genome and individual open reading frame (ORF) sequences of HBV obtained using DAMBE*  

<table>
<thead>
<tr>
<th></th>
<th>Intragroup</th>
<th></th>
<th>Intergroup</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subgenotype A1</td>
<td>Subgenotype A2</td>
<td>Genotype A</td>
<td>A1 vs A2</td>
</tr>
<tr>
<td>Complete genome</td>
<td>2.53 ± 1.00</td>
<td>1.39 ± 0.87</td>
<td>3.36 ± 1.47</td>
<td>4.74 ± 0.56</td>
</tr>
<tr>
<td>polymerase</td>
<td>2.52 ± 0.95</td>
<td>1.25 ± 0.70</td>
<td>3.35 ± 1.48</td>
<td>4.77 ± 0.49</td>
</tr>
<tr>
<td>Pre-S1/S2</td>
<td>3.39 ± 2.07</td>
<td>1.45 ± 1.02</td>
<td>4.72 ± 2.36</td>
<td>6.99 ± 1.02</td>
</tr>
<tr>
<td>HBsAg</td>
<td>1.21 ± 0.56</td>
<td>0.84 ± 0.62</td>
<td>1.53 ± 0.74</td>
<td>2.06 ± 0.59</td>
</tr>
<tr>
<td>Precore/core</td>
<td>2.40 ± 1.19</td>
<td>1.75 ± 1.67</td>
<td>3.26 ± 1.65</td>
<td>4.58 ± 1.08</td>
</tr>
<tr>
<td>X</td>
<td>1.78 ± 1.31</td>
<td>1.55 ± 0.71</td>
<td>2.38 ± 1.26</td>
<td>3.23 ± 0.69</td>
</tr>
</tbody>
</table>

*http://web.hku.hk/~xxia/software/software.htm. The sequences compared are those included in figure 1 [21 subgenotype A1 and 9 subgenotype A2].
†The mean nucleotide divergence (%) ± standard deviation and the range in parentheses.

genotype A HBV recognized in Vietnam, and it has been suggested that this aberrant genotype may be a link between the European/African A and the Asian B and C genotypes (Hannoun et al., 2000). Val91 of the pre-S1 gene that is characteristic of subgenotype A1 is also found in gibbon (Grethe et al., 2000; Norder et al., 1996) and orangutan (Verschoor et al., 2001) hepadnavirus isolates.

The subgenotype A1 unique amino acids Gln334 and Lys338 are found in the fingers of the HBV polymerase within the DNA-binding cleft that is positively charged (Das et al., 2001). Gly334 is uncharged and replaces basic Lys in subgenotype A2 and acidic Glu in other genotypes. On the other hand, basic Lys338 replaces Glu and Asp, both of which are acidic and found in subgenotype A2 and other genotypes, respectively. The change in the charge caused by the alternate amino acids within this region could possibly affect the binding of the DNA to the polymerase and influence reverse transcription. Subgenotype A1, which is the subgenotype prevalent in southern Africa is associated with low HBV DNA levels (Kramvis et al., 1997).

The following signature amino acid motif, ‘Ser11, Ala31, Ser47, Ser146, Ser147’ (verschoor et al, 2003). The Ser146, Ser147 in the X ORF are a result of mutations at nucleotide position 1809 and 1812 and have an effect on the overlapping Kozak sequence preceding the precore/core start codon.

Double or triple point mutations at positions 1809–1812 were found only in subgenotype A1 isolates and not in subgenotype A2 or other genotypes (Fig. 3). In a previous study, we reported that 80% of SA HBV strains harbour similar mutations immediately upstream of the precore AUG codon (Baptista et al., 1999), which might impair HBeAg expression as a result of suboptimal translational initiation (Kozak, 1986, 1987; Kramvis & Kew, 1999). We tested this hypothesis using site-directed mutagenesis and transfection experiments and showed that hepatitis B e antigen expression was severely impaired by the 180918111812 and 180918111812 triple mutations, and moderately reduced by the 18091812 and 18091812 double mutations (Ahn et al., 2003). The effect of the double mutations on hepatitis B e antigen expression was comparable with that of the common core promoter mutations (1762-1764) and independent of HBx expression (Ahn et al., 2003). These mutations are not a result of an adaptive change under immune pressure because they are found in HBV isolates obtained from children and in acute hepatitis patients (Ahn et al., 2003). These mutations have previously been reported to occur only occasionally in other regions of the world (Estacio et al., 1988, Kidd-Ljunggren et al., 1995, Laskus et al., 1994), supporting our observation that they are characteristic of subgenotype A1 isolates.

Regulatory cis-acting elements are embedded within the protein coding ORFs of HBV. Therefore any changes in
### Fig. 2.
For legend see page 1218.
**Fig. 3.** For legend see page 1218.
Table 3. Mutations within the cis-acting regulatory elements found predominantly in subgenotype A1

<table>
<thead>
<tr>
<th>cis-acting element</th>
<th>Mutation</th>
<th>Functional element affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhancer I/X promoter</td>
<td>A/T963C</td>
<td>5' modulator element</td>
</tr>
<tr>
<td></td>
<td>C/T1464T</td>
<td>NRE-γ</td>
</tr>
<tr>
<td></td>
<td>C/T1464G</td>
<td>NRE-β</td>
</tr>
<tr>
<td></td>
<td>A/G1512T</td>
<td>NRE-β</td>
</tr>
<tr>
<td></td>
<td>G1809T, A1811T, C1812T/G</td>
<td>precore Kozak Sequence*</td>
</tr>
<tr>
<td></td>
<td>G1888A</td>
<td>encapsidation signal</td>
</tr>
<tr>
<td>S1 promoter</td>
<td>T/G2720A</td>
<td>HNF-1 binding site</td>
</tr>
<tr>
<td>S2 promoter</td>
<td>[A/G3013C,C/T3014A]</td>
<td>NFI transcription factor-binding site</td>
</tr>
<tr>
<td></td>
<td>T/A/G3045C</td>
<td>region A</td>
</tr>
<tr>
<td></td>
<td>C/A3109G, A/C3111T</td>
<td>region E</td>
</tr>
<tr>
<td></td>
<td>[T3132C,C3133A]</td>
<td>region F</td>
</tr>
</tbody>
</table>

*Shown to reduce HBeAg translation by ribosomal leaky scanning mechanism (Ahn et al., 2003).

The silent G to A nucleotide mutation at position 1888 was unique to subgenotype A1 (Fig. 3). This mutation occurs rarely in other genotypes and in HBV isolates from outside Africa. In addition to stabilizing the encapsidation signal (ɛ) (Kramvis & Kew, 1998) and possibly affecting reverse transcription, this mutation could affect the translation of the core protein. The 1888 G to A mutation introduces an out-of-frame AUG start codon, 13 nucleotides upstream of the core AUG start codon and a minicistron that can potentially be translated into seven amino acids: ‘Met-Ala-Leu-Gly-His-Gly-His’. Therefore, the newly introduced start codon at 1888 may have an important role in the regulation of the translational efficiency of a downstream start codon (Rogozin et al., 2001), in our case the start codon for translation of the core protein. The presence of small upstream ORFs in the leader sequence has also been found to have a modulating role in the translation of proteins from downstream cistrons in a number of viruses (Biegalke & Geballe, 1990; de Smit & van Duin, 1993; Degnin et al., 1993; Ozawa et al., 1988; Ryabova et al., 2002). Similarly, the translation of HBV polymerase gene is
controlled by a leaky scanning mechanism together with a termination-reinitiation mechanism involving an upstream minicistron (Fouillot et al., 1993; Hwang & Su, 1998). Therefore, it is conceivable that the introduction of the start codon by the 1888 G to A mutation, seen in subgenotype A1, may play a modulating role in the translation of the core protein and needs further investigation.

In conclusion, it can be seen that subgenotype A1 HBV isolates from SA differ from subgenotype A2 in two ways. Firstly, subgenotype A1 isolates have distinctive sequence characteristics that may affect both the replication of the virus and the expression of its proteins. Secondly, the mean nucleotide divergence of subgenotype A1 is greater than that for subgenotype A2 suggesting that subgenotype A1 has been endemic and has a very long natural history within the South African black population.

ACKNOWLEDGEMENTS

Anna Kramvis and Gerald Kimbi received grants from the Poliomyelitis Research Foundation of South Africa and the H.E. Griffin Cancer Trust.

REFERENCES


