Analysis of the complete genome of subgroup A’ hepatitis B virus isolates from South Africa

Anna Kramvis, Louise Weitzmann, William K. B. A. Owiredu and Michael C. Kew

MRC/CANSA/University Molecular Hepatology Research Unit, Department of Medicine, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg 2193, South Africa

A phylogenetic analysis is presented of six complete and seven pre-S1/S2/S gene sequences of hepatitis B virus (HBV) isolates from South Africa. Five of the full-length sequences and all of the pre-S2/S sequences have been previously reported. Four of the six complete genomes and three of the five incomplete sequences clustered with subgroup A’, a unique segment of genotype A of HBV previously identified in 60% of South African isolates using analysis of the pre-S2/S region alone. This separation was also evident when the polymerase open reading frame was analysed, but not on analysis of either the X or pre-core/core genes. Amino acids were identified in the pre-S1 and polymerase regions specific to subgroup A’. In common with genotype D, 10 of 11 genotype A South African isolates had an 11 amino acid deletion in the amino end of the pre-S1 region. This deletion is also found in hepadnaviruses from non-human primates.

Hepatitis B virus (HBV) is the prototype of the family Hepadnaviridae. Using phylogenetic analysis of the complete genome of HBV, the virus has been classified into genotypes A–G, with an intergenotypic diversity of at least 8% (Okamoto et al., 1988; Norder et al., 1994; Stuyver et al., 2000). The seven 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 and genotype F in the aboriginal populations of South America (Norder et al., 1993; Arauz-Ruiz et al., 1997). To date, isolation of the recently identified genotype G has been limited to HBV carriers in France and Georgia, USA (Stuyver et al., 2000).

Although HBV is hyperendemic in sub-Saharan Africa, HBV sequencing data from this region are limited. We have recently published the first seven full-length sequences of South African HBV isolates: two isolates from anti-HBs-positive asymptomatic carriers of HBV were found to be genotype A/D recombinants (Owiredu et al., 2001a), and five HBV isolates from fulminant hepatitis patients (Owiredu et al., 2001b) belonged to genotype A. Moreover, three of the latter isolates clustered with subgroup A’, the unique segment of genotype A that we have previously identified in isolates from South Africa using phylogenetic analysis of pre-S2/S genes (Bowyer et al., 1997). We report here one additional complete genotype A sequence (#A20) and seven pre-S1 sequences (#A18, #A26, #A28, #A29, #A30, #B05, #C25) of HBV from South Africa (Fig. 1) and provide a comparative molecular analysis of the six full-length genomes, four of which cluster with subgroup A’.

The Human Ethics Committee of the University of the Witwatersrand approved the study and serum samples were obtained, with informed consent, from five fulminant and eight acute hepatitis B patients. All patients were seropositive for HBsAg and HBeAg except for #78 who was HBeAg-negative. DNA extraction, amplification of subgenomic fragments of HBV by PCR, direct sequencing and phylogenetic analysis were carried out as described previously (Bowyer et al., 1997; Owiredu et al., 2001a, b). The sequences of the whole genome (when available) and of the four individual open reading frames (ORFs) were compared with corresponding sequences of HBV obtained from GenBank. Serotypes for all samples were deduced from sequence data of the S ORF and some of them were confirmed using monoclonal antibodies (Bowyer et al., 1997).

The complete sequences of HBV isolates AF297621 (#78), AF297622 (#79), AF297624 (#80), AF297623 (#83) and AF297625 (#84) from fulminant hepatitis patients reported previously (Owiredu et al., 2001b) belonged to subtype adw2 and all except #83 had a genomic length of 3200 bp. Isolate #83 was 3149 nucleotides long. Additional sequencing data of the pre-S1 region of HBV isolates from acute hepatitis patients extend our previous study in which only the pre-S2/S genes

Author for correspondence: Michael Kew.
Fax +27 11 643 4318. e-mail mkew@chiron.wits.ac.za

GenBank accession numbers of sequences reported in this paper: AF364333, U87740, U87741, U87743–U87745, U87747 and U87748.
A. Kramvis and others

Fig. 1. Dendrograms obtained by neighbour-joining phylogenetic analysis of full-length HBV sequences rooted on genotype F (a), showing the relationship of six South African isolates (boxed) to the seven genotypes including subgroup A', and by neighbour-joining phylogenetic analysis of pre-S1/S2/S sequences (nt 2848–814 from EcoRI site) of HBV rooted on genotype F (b), showing the relationship of 13 pre-S1/S2/S sequences of South African HBV isolates (boxed) to the seven genotypes including subgroup A'.

were sequenced (Bowyer et al., 1997). The pre-S1 gene has now been sequenced for the following HBV isolates: U87740 (#A16), U87741 (#A26), U87744 (#A28), U87745 (#A29), U87743 (#A30), U87747 (#B05) and U87748 (#C25). The complete genomic sequence has been obtained for isolate AF364333 (#A20) for which the partial pre-S2/S sequence (U87742) was previously reported (Bowyer et al., 1997). The complete sequence is 3231 nucleotides long. There is a three amino acid insertion at positions 152–154 of the X gene overlapping the amino end of the pre-core region. In common with all the South African genotype A HBV isolates analysed, this sequence has the six nucleotide insert (position 2354–2359) in the core region that is characteristic of genotype A.

Phylogenetic analysis was performed using neighbour-joining and maximum likelihood algorithms (Owiredu et al., 2001a), and these did not differ significantly in the relative positions of the sequences. Moreover, phylogenetic analyses of nucleotide and amino acid sequences were in general agreement. Regardless of whether the complete genome sequence or the sequences of the individual ORFs were analysed, isolates #78, #79, #80, #83, #84 and #A20 clustered with genotype A (Fig. 1a and b), concurring with divergence determination (data not shown). Four isolates, #78, #83, #84 and #A20, clustered with subgroup A'. The only non-South African strain of HBV previously documented as belonging in subgroup A', and for which a complete sequence is available, is a strain isolated from the Philippines, GenBank number M57663 (Estacio et al., 1988). The splitting of genotype A into two subgroups, namely subgroups A' and A — A' (genotype A excluding A'), was well supported by bootstrap analysis when the complete genome (Fig. 1a), pre-S1/S2/S (Fig. 1b) and polymerase genes were analysed. The tree topology following phylogenetic analysis of the polymerase gene was essentially the same as that for the complete genome (tree not shown). However, the separation of subgroup A' from the remainder of genotype A did not occur upon phylogenetic analysis of either
Table 1. Comparison of amino acid residues of S and polymerase ORFs of subgroup A isolates with amino acid sequences of other human HBV genotypes and non-human primate hepadnaviral isolates.

<table>
<thead>
<tr>
<th>IDH</th>
<th>S/T/R/G</th>
<th>P/L/A/S (not T, which defines subgroup A — A')</th>
</tr>
</thead>
</table>

- Amino acids found in subgroup A and in other non-A genotypes (B—F) are shown in bold and those found predominantly in subgroup A and not other genotypes are in bold and are shaded.
- South African isolates are shown in bold and those found predominantly in subgroup A and not other genotypes are in bold and are shaded.
- Analysis of the complete genome obtained from GenBank, Vietnamese A refers to a consensus of AF241407—AF241409. The isolates that cluster with subgroup A after phylogenetic analysis of the complete genome (Fig. 1a) and/or the S ORF (Fig. 1b) are shaded.
- Deleted amino acid residue.
- Not sequenced.
the X or core genes. This is to be expected because 66% of the polymerase gene and the entire S gene contain overlapping reading frames, whereas the major portion of both the core and X genes is single coding regions. Thus, the molecular evolution of both the polymerase and S genes is more constrained than that of the core and X genes (Mizokami et al., 1997).

The mean nucleotide divergences (%) of the full-length sequences of subgroup A’ isolates compared with sequences representative of genotypes A–G were: 4.61 (A–A’), 4.31 (A), 9.31 (B), 8.61 (C), 10.25 (D), 9.63 (E), 13.42 (F) and 10.93 (G). The intragroup divergence for subgroup A’ was 3.80%, subgroup A–A’ 1.48% and the whole of genotype A 3.54%. Similar values were obtained when the nucleotide divergence of the S ORF alone was determined. The intergroup divergence of the complete genome between subgroup A’ and A–A’ is less than 8% and therefore too low to warrant a genotype separation of these two groups.

The inclusion of pre-S1 sequences in the phylogenetic analysis did not alter the genotypic segregation previously obtained when the pre-S2/S gene sequences were analysed alone (Bowyer et al., 1997). Isolates #A18, #A29 and #A30 were found in subgroup A’, #A26 and #A28 in subgroup A–A’, whereas #B05 and #C25 clustered with genotypes B and C, respectively.

By examining the translated sequences of the four ORFs of HBV we identified a number of amino acid residues that distinguished subgroup A’ from the remainder of genotype A in the S and polymerase genes (Table 1) but not in the X and pre-core/core genes. A number of these residues were unique to subgroup A’ (boldface and shaded) whereas others were shared with other non-A genotypes (B–G) (boldface only). In addition to amino acids 32, 35, 47 and 54 of the pre-S2 region and residues 207 and 209 of the surface gene, which we had previously found to set subgroup A’ apart from the rest of genotype A (Bowyer et al., 1997), we have now identified subgroup A’-specific amino acids in the pre-S1 and polymerase genes (Table 1). These included 54Q, 74V, 86A and 91V in the pre-S1 gene, and 236T, 256C, 268G, 333T and 334Q in the polymerase gene. Moreover, amino acid residues previously found only in non-A genotypes have been identified in subgroup A’ isolates. These include 67F, 89P and 90A in pre-S1, 35V, 47S and 54P in pre-S2, and 87H, 120N, 121S, 220F, 271A, 273S, 308S, 309F and 315R in the polymerase gene. Six published sequences, the Philippine isolate M57663 (Estacio et al., 1988), three genotype A isolates from Brazil (M52220–M52222) (Moraes et al., 1996), M74498 from France (Tran et al., 1991) and X69458 from Zimbabwe (Chirara & Chetsanga, 1994), all share amino acids with subgroup A’ in the S and polymerase ORFs (Table 1). They also cluster within this subgroup after phylogenetic analysis of the S gene (Fig. 1b), confirming the association of these amino acids with subgroup A’ isolates. The amino acids specific to subgroup A’ were concentrated in the pre-S1 region. This region has a role to play in attachment of HBV to hepatocytes (Neurath et al., 1986; Pontisso et al., 1989) and its sequence is well conserved within a given HBV subtype (Uy et al., 1992). Therefore, it is possible that the molecular evolution of the pre-S1 sequence is constrained by the host population. It is of interest to note that a number of the amino acids identified in subgroup A’ isolates are also found in the aberrant genotype A HBV recognized in Vietnam (Hannoun et al., 2000) (Table 1). These authors have suggested that this aberrant genotype may be a link between the European/African A and the Asian B and C genotypes. Moreover, a valine at position 91 of the pre-S1 gene that is characteristic of subgroup A’ is found in gibbon and orangutan hepadnaviral isolates.

In common with genotype D, 10 of 11 genotype A South African isolates (#A18, #A26, #A28, #A29, #A30, #78, #79, #80, #83 and #84) had an 11 amino acid deletion in the amino end of the pre-S1 region (Table 1). This deletion was described previously in genotype A isolates adto V00866 (Ono et al., 1983) and in X69458 from Zimbabwe (Chirara & Chetsanga, 1994). It is therefore possible that the genotype A isolates with this pre-S1 deletion are genotype A variants prevalent in the Southern African black population. This deletion is also found in hepadnaviruses from non-human primates (Table 1).

By analysing the complete genomes of South African HBV isolates, we have confirmed the predominance of subgroup A’ in this area. Furthermore, we showed that the separation into the two subgroups of A results from changes in amino acid residues in the S ORF overlapping the polymerase gene. As more sequencing data are generated in the future, it is possible that other genotypes will be shown to contain similar subgroup divisions. These subdivisions could aid in tracing the molecular evolution and the transmission routes of HBV and facilitate our understanding of the relationship between virus isolates from different geographical regions and their respective roles in viral pathogenesis.

This study was supported by grants from the H. E. Griffin Cancer Trust awarded to Dr Anna Kramvis and Mr William Owiredu. Dr Kramvis is a recipient of a grant from the Poliomyelitis Research Foundation.

References


Received 23 July 2001; Accepted 5 December 2001

839