The Complete Nucleotide Sequence of the Genome of a Hepatitis B Virus Isolated from a Naturally Infected Chimpanzee

By MARK VAUDIN,† ADRIAN J. WOLSTENHOLME, KWESI N. TSIQUAYE, ARIE J. ZUCKERMAN AND TIM J. HARRISON*

Department of Medical Microbiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

(Accepted 24 February 1988)

SUMMARY

The complete nucleotide sequence of a strain of hepatitis B virus, originally isolated from a naturally infected chimpanzee, has been determined. Interesting features of the sequence include the presence of an in-phase stop codon in the 'pre-core' region of the core antigen open reading frame. The sequence shows approximately 10% nucleotide divergence from all of the other hepatitis B virus sequences previously published and the possibility that this divergence is the result of passage through chimpanzees is discussed.

The occurrence of naturally acquired hepatitis B virus (HBV) infection in the colony of chimpanzees of the Zoological Society of London has been reported previously (Zuckerman et al., 1978). One of the animals, chimpanzee K, was persistently infected with a high titre of hepatitis B surface antigen (HBsAg), and was also positive for both hepatitis B e antigen (HBeAg) and circulating virus. The subtype of the surface antigen was adw. The mother of this animal died from natural causes in February, 1973, and we have been unable to test her serum. However, a female sibling was also HBsAg-positive and it is possible that both became infected perinatally. We describe here the cloning and sequencing of the genome of this virus.

Virus was pelleted from the plasma of chimpanzee K and further purified by sedimentation through sucrose. Virus DNA was purified by standard methods following completion of the plus strand by the endogenous DNA polymerase. Because the DNA was found to lack the single EcoRI site common to most strains of HBV and a convenient single-cut enzyme was not readily identifiable, it was cloned as two separate PstI fragments of 1758 and 1424 nucleotides (nt) into pBR328. The PstI fragments were later subcloned into M13mp19 so that clones containing both orientations of the fragments were obtained. DNA sequencing was performed using the chain termination method (Sanger et al., 1977). A nested set of templates was obtained from each clone by digestion with exonuclease III according to the method of Henikoff (1984). Any gaps in the sequence were filled by subcloning and sequencing specific restriction enzyme fragments, so that a complete sequence was obtained for both strands of the genome.

The nucleotide sequence of this strain of hepatitis B virus is shown in Fig. 1. The sequence reveals the four major open reading frames of HBV (surface, core, polymerase and X genes as indicated) and is numbered from the ATG of the HBcAg open reading frame (Fig. 2). The PstI sites used in the cloning process are located at nucleotide positions 1305 and 3063. We believe that the two individual sequences are contiguous because the junctions result in intact genes (for HBsAg/P and X respectively), the sequences across the junctions are homologous to other published sequences and the total number of nucleotides (3182) is in agreement with that found for other subtypes of HBV.

† Present address: Plant Biotechnology Section, Imperial Chemical Industries plc, Jealott's Hill Research Station, Bracknell RG12 6EY, U.K.
The complete nucleotide sequence of the LSH strain of HBV. Position 1 is the start of the core region. The predicted amino acid sequences of the four major open reading frames are shown above the nucleotide sequence. The following sequences are underlined: the stop codon in the pre-core region (nt 3177), the start of pre-S1 (nt 948), pre-S2 (nt 1272) and the major sAg gene (nt 1437) and the direct repeats DR1 (nt 3106 to 3116) and DR2 (nt 2872 to 2882).

Fig. 1. The complete nucleotide sequence of the LSH strain of HBV. Position 1 is the start of the core region. The predicted amino acid sequences of the four major open reading frames are shown above the nucleotide sequence. The following sequences are underlined: the stop codon in the pre-core region (nt 3177), the start of pre-S1 (nt 948), pre-S2 (nt 1272) and the major sAg gene (nt 1437) and the direct repeats DR1 (nt 3106 to 3116) and DR2 (nt 2872 to 2882).

The amino acid sequences deduced for the four main open reading frames are also shown in Fig. 1 and a schematic diagram of the genome organization forms Fig. 2. The adw subtype may be confirmed from the sequence of the major HBsAg protein, lysine at amino acid position 122 (nt 1801) confirms the d subtype and lysine at amino acid position 160 (nt 1915) the w subtype (Okamoto et al., 1987). A six base pair insertion (which would occur between nt 455 and 456 in our clone) is present in both the adw-1 (Valenzuela et al., 1980) and adw-2 (Ono et al., 1983) sequences but not in this sequence, despite the fact that the adw-2 sequence is otherwise conserved for around 40 base pairs either side of this position. The region encoding the pre-S1 domain of HBsAg displays considerable size variation and the sequences of this clone and others reported in the literature are compared in Fig. 3. It may be seen that the size of the deletion in this clone is identical to that of the ayw sequence reported by Galibert et al. (1979) and the amino-terminal amino acid sequence of pre-S1 most closely resembles that of the ayw subtype. Comparison of overall nucleotide homology with other strains (Table 1) shows 9-4 to 10% nucleotide differences compared to other published HBV sequences. These other sequences themselves fall into three groups: adw-1 and adw-2, ayw, ayw-2 and adyw [the adyw sequence was derived from a mixed clone bank (Burrell et al., 1979) and has the characteristics of the ayw subtype according to the definition of Okamoto et al. (1987)]; and adr, adr-4 and ayr (references for these sequences are given in Table 1).

Some of the effects of these changes on the amino acid sequences of the HBV proteins are noteworthy. Near to the carboxyl terminus of the core protein the sequence Pro-Ala (nt 535)...
Short communication

Fig. 2. A diagrammatic representation of the genome of the LSH strain of HBV. The inner circle shows the numbering system (clockwise, with the start of the HBcAg open reading frame as position 1) and the two PstI sites used in the original cloning. The locations of the four major open reading frames are shown outside this circle, the broken lines preceding the HBcAg gene denote the position of the pre-core region with the in-phase stop codon.

Subtype  
LSH clone  
ayw  
adw-2  
adw-1  
ad and ayr  

Table 1. Percentage variability between the nucleotide sequence of the LSH strain of HBV and other strains*  

<table>
<thead>
<tr>
<th></th>
<th>adw-1</th>
<th>adw-2</th>
<th>ayw</th>
<th>ayw-2</th>
<th>adyw</th>
<th>adr</th>
<th>adw-4</th>
<th>ayr</th>
</tr>
</thead>
<tbody>
<tr>
<td>adw-LSH</td>
<td>9.8</td>
<td>9.5</td>
<td>9.5</td>
<td>9.4</td>
<td>9.7</td>
<td>9.6</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>ayr</td>
<td>9.7</td>
<td>9.6</td>
<td>10.2</td>
<td>10.3</td>
<td>10.9</td>
<td>2.4</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>adw-4</td>
<td>9.4</td>
<td>9.9</td>
<td>10.5</td>
<td>10.7</td>
<td>11.4</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adr</td>
<td>8.9</td>
<td>9.9</td>
<td>10.6</td>
<td>11.5</td>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adyw</td>
<td>11.0</td>
<td>10.6</td>
<td>3.6</td>
<td>4.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ayw-2</td>
<td>11.0</td>
<td>10.1</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adw-2</td>
<td>10.8</td>
<td>9.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Comparisons are with the sequences of Valenzuela et al. (1980; adw-1), Ono et al. (1983; adw-2 and adw), Galibert et al. (1979, ayw), Bichko et al. (1985, ayw-2), Fujiyama et al. (1983, adw-4), Okamoto et al. (1986, ayr), and Will et al. (1982, adyw).
replaces the Arg–Glu found in all other published sequences. These changes also affect the amino acid sequence of the polymerase. A protease-like sequence near the amino terminus of the core protein which may be involved in the generation of HBeAg (Miller, 1987) is conserved in the LSH strain (nt 79 to 111). The sequence of the major surface antigen is highly conserved when compared to other published sequences. As mentioned above, the pre-S1 amino acid sequence is found to be highly variable between different HBV strains and a number of unique changes occur in this region of the LSH strain. If the receptor-binding site for the hepatocyte is located in the pre-S1 region (Neurath et al., 1986) it is conceivable that some of these changes may represent adaptation of the virus to the chimpanzee host. However, the pre-S2 region is also diverged in our strain and, taken together, these changes affect the amino acid sequence of the polymerase more than that of the pre-S proteins. The middle third of the polymerase is most diverged, in our strain, but the region that shows homology to other viral RNA-dependent DNA polymerases (Toh et al., 1983) is highly conserved. A region of the X protein near the amino terminus (nt 2757 to 2803) which is somewhat variable among HBV strains is less conserved in the LSH sequence. The X protein appears to be a transcriptional trans-activator (Twu & Schoelmer, 1987) and, again, it is possible that changes in this protein may be the result of adaptation to a chimpanzee host.

Atypically, the LSH sequence reveals an in-phase stop codon in the 'pre-core' region of the core antigen open reading frame (position 3177). In order to try to eliminate the possibility that we had cloned a defective virus we obtained a second clone of the 1424 nt PstI fragment from a different plasma sample from the same animal (taken over 12 months later) and sequenced it from the PstI site through the stop codon. The presence of the stop codon was confirmed and the sequence of that strand found to be identical to that in Fig. 1 over 250 nt.

The presence of this in-phase stop codon is particularly noteworthy. A cloned HBV genome with a stop codon in the pre-core region was found in The Netherlands to be non-infectious for chimpanzees (Will et al., 1985). The clone, however, had other anomalies including a six base pair duplication at the end of the X gene which may interfere with its replication and render it defective. Chang et al. (1987) have introduced a frameshift mutation into the pre-core region of an avian hepadnavirus (duck HBV) and have shown that the resultant clone, when inoculated intrahepatically into newborn ducklings, gives rise to an apparently normal, productive infection. If both the pre-core and core start codons are present on a transcript of this gene, translation starts preferentially at the pre-core AUG (Weimer et al., 1987). However, core antigen without the pre-core domain is the major product in vivo and this may reflect heterogeneity of the 5' ends of these transcripts, with the most abundant family of transcripts initiated in the pre-core region itself. The pre-core region resembles a signal sequence and it has been argued that it may be responsible for the secretion of HBeAg (Ou et al., 1986; Roossinck et al., 1986; Uy et al., 1986) though, in this study, the chimpanzee was highly positive for HBeAg (using a monoclonal antibody-based assay; Sorin-Diagnostics, Italy). We have used site-directed mutagenesis to remove this stop codon from our clone (M. Donati, A. J. Wolstenholme & T. J. Harrison, unpublished observations) and plan to investigate the secretion of HBeAg by wild-type and mutant virus in eukaryotic expression systems. Although we cannot rule out the possibility that we have sequenced the genomes of two defective viruses, it appears likely that expression of the pre-core region is not required for infectivity in chimpanzees.

The sequence of this clone shows approximately 10% nucleotide differences compared to other published HBV sequences (Table 1) including the other clone containing a pre-core stop codon (Will et al., 1982). It has been suggested that the chimpanzee responsible for infecting the colony may itself have been infected naturally in Africa (Zuckerman et al., 1978) and this may be the first reported complete sequence of an African isolate of the virus. Alternatively, the sequence divergence may in part represent adaptation of the human virus to a chimpanzee host as discussed above. A sequence of the hepatitis delta virus cloned from an infected human (Makino et al., 1987) was found to be 11% divergent from that of virus cloned from infected chimpanzees (Wang et al., 1986; Kos et al., 1986) and Makino et al. suggested adaptation to the chimpanzee host as one possible explanation of this divergence. Since HBV replicates via an RNA intermediate (Summers & Mason, 1982) it may have the high mutation rate typical of...
RNA viruses. It is now clear that HBsAg subtype specificity depends upon relatively minor changes in protein sequence (Okamoto et al., 1987) and divergence of HBV nucleotide sequence may be otherwise largely independent of these specificities.

We are grateful to Subhan Shah for help in confirming the pre-core stop codon and to Amanda Turner for performing the HBeAg tests. This study was supported by generous awards from the Department of Health and Social Security and the Wolfson Foundation.

REFERENCES


(Received 10 November 1987)