Typing Hepatitis B Virus by Homology in Nucleotide Sequence:
Comparison of Surface Antigen Subtypes

By HIROAKI OKAMOTO, 1 FUMIO TSUDA, 2 HIROSHI SAKUGAWA, 3
RETNO I. SASTROSOEWIGNJO, 4 MITSUNOBU IMAI, 1
YUZO MIYAKAWA 5 AND MAKOTO MAYUMI 1,*

1 Immunology Division, Jichi Medical School, Minamikawachi-Machi, Tochigi-Ken 329-04,
2 Section of Immunology, The Kitasato Institute, Tokyo 108, 3 First Department of Internal
Medicine, University of Ryukyus, Okinawa-Ken 903-01, Japan, 4 Department of Microbiology,
University of Indonesia, Jakarta, Indonesia and 5 Institute of Immunology, Tokyo 112, Japan

(Accepted 16 June 1988)

SUMMARY

The complete nucleotide sequences of the DNA of three hepatitis B virus (HBV) genomes of subtype adw, cloned from plasma samples of asymptomatic carriers living in the mainland and Okinawa Prefecture of Japan and Indonesia were determined. All three comprised 3215 bp and differed in sequence by only 3.9 to 5.6%. When these isolates were compared with the reported sequences of two HBV genomes of the same subtype derived from American carriers, however, the differences were greater (8.3 to 9.3%) to an extent comparable with the nucleotide divergence between an HBV genome of subtype adw and that of a heterotypic subtype, such as adr, ayw or ayr. A total of 18 HBV genomes of various subtypes, including the three described here, 10 reported previously and five unpublished ones, were classified into four groups based on an inter-group divergence in nucleotide sequence of 8% or greater: group A (two adw genomes), group B (four adw), group C (three adw, four adr and one ayr) and group D (four ayw). Thus, the nine genomes of HBV subtype adw were distributed into three groups with considerably different sequences. These results indicate that the four major antigenically defined subtypes of envelope polypeptide do not reflect true genotypic variation of HBV. The fact that d to y, as well as w to r, subtypic change can be induced by an A → G point mutation at nucleotides 365 and 479 in the S gene, respectively, supports this view.

INTRODUCTION

The nucleocapsid of hepatitis B virus (HBV) is a 27 nm sphere that bears HBV core antigen, and contains HBV e antigen as well as viral DNA and DNA polymerase. It is covered by an envelope of HBV surface antigen (HBsAg), which has the common, group-specific determinant a. In addition, it carries one member of each of the two pairs of mutually exclusive subtype determinants d and y (Le Bouvier, 1971) and w and r (Bancroft et al., 1972). As a result, the four major subtypes of HBsAg, adw, adr, ayw and ayr, are generated, and they have been proposed to represent the phenotypic expression of the four major genotypes of HBV (Le Bouvier et al., 1972). HBsAg subtypes show distinct geographical distributions (Courouce-Pauty et al., 1983; Mazzur et al., 1974), and may provide historical information on the migration of ancestors carrying HBV of a particular subtype (Yamashita et al., 1975).

The complete nucleotide sequences have been determined for the genomic DNA of HBV strains of various subtypes (Galibert et al., 1979; Valenzuela et al., 1980; Fujiyama et al., 1983; Ono et al., 1983; Kobayashi et al., 1984; Bichko et al., 1985; Sastrosoewignjo et al., 1985; Okamoto et al., 1986, 1987a), allowing analysis of HBV genotypes at the nucleotide level. In general, the extent of divergence in nucleotide sequence between two HBV genomes of the same subtype (intra-subtypic divergence) is much smaller than that between two HBV genomes of
different subtypes (inter-subtypic divergence), with a score of 1-1 to 2-7% compared to 8-5 to 10-0% (Okamoto et al., 1986). The only exception to this rule is the genome of subtype ayw which is different from those of adr by only 2-2% (Okamoto et al., 1986).

We have determined the complete nucleotide sequences of three HBV genomes of subtype adw cloned from plasma samples from Asian countries, and compared them with those of 15 HBV genomes of various subtypes. Four HBV types were recognized on the basis of absolute differences in their complete nucleotide sequences; these groupings did not necessarily correlate with the HBsAg subtypes.

METHODS

Plasma samples (20 ml) were obtained from three asymptomatic carriers of HBsAg living in mainland Japan, in the Okinawa Prefecture (located at the southernmost part of Japan) and in Indonesia. They all contained HBsAg of subtype adw as determined by solid-phase enzyme immunoassay with monoclonal antibodies (Institute of Immunology, Tokyo, Japan), and were positive for HBV DNA by dot blot hybridization (Scotto et al., 1983). Dane particles were isolated from plasma by the method described previously (Takahashi et al., 1976).

The gapped region in HBV DNA was repaired with the aid of endogenous DNA polymerase (Kaplan et al., 1973), and viral DNA was then extracted with phenol/chloroform. HBV DNA was cloned at the BamHI site of pSP65 by a method similar to that used for the cloning of HBV subtype ayr (Okamoto et al., 1986).

Appropriate restriction fragments of HBV DNA clones were subcloned in M13 mp10 and mp11 phage vectors (Amersham). The nucleotide sequences of both plus and minus strands were determined by the dideoxy chain termination method (Sanger et al., 1977).

Comparisons were made between these three subtype adw HBV DNA clones and 10 HBV DNA clones of subtypes adw, adr, ayw and ayr, previously reported (Galibert et al., 1979; Fujiyama et al., 1983; Ono et al., 1983; Kobayashi & Koike, 1984; Bichko et al., 1985; Sastrosoewignjo et al., 1985; Okamoto et al., 1986, 1987a; Valenzuela et al., 1980). Also included in the study were five HBV DNA clones (two adw, one adr and two ayw) the complete nucleotide sequences of which were determined by the method described above. Since the primary purpose of this study was to correlate the four major subtypes of HBsAg with genotypic variation of HBV, DNA clones of compound HBV subtypes, such as adyw (Will et al., 1982), were not included even though the entire nucleotide sequences were available for them.

Nucleotide sequences were aligned at the EcoRI site for HBV DNA clones that possessed it, or at an equivalent position for those without it, in accordance with the numbering system already reported (Galibert et al., 1979; Ono et al., 1983; Okamoto et al., 1986). A minimal number of gaps (deletions and insertions) that had to be introduced in each of these alignments were excluded from the comparison.

RESULTS

In Fig. 1, the entire nucleotide sequences are shown for three HBV subtype adw genomes originating from mainland Japan (pJDW233), the Okinawa Prefecture (pODW282) and Indonesia (pIDW420). They all comprised 3215 bp and possessed four open reading frames (ORFs) (Tiollais et al., 1985), i.e. the pre-S region with the S gene (1200 bp), the pre-C region with the C gene (636 bp), the X gene (462 bp) and the P gene (2529 bp). The three HBV DNA clones were therefore considered to represent the complete genomes of these subtype adw isolates. When the nucleotide sequences were compared with each other, differences were found at only 125 (3.9%) nucleotide positions between pJDW233 and pODW282, at 144 (4.5%) positions between pODW282 and pIDW420, and at 170 (5.6%) positions between pIDW420 and pJDW233.

A total of 18 HBV genomes, including the three adw clones described here as well as 10 previously reported and five unpublished clones of various subtypes (six adw, four adr, four ayw and one ayr), were subjected to pairwise analysis with the results listed in Table 1. For the sake of visual clarity, the divergence, rather than the homology, in nucleotide sequence is shown. On the basis of an inter-group difference in the nucleotide sequence of 8% or greater they were classified into four groups.

Group A comprised two subtype adw genomes, both derived from American carriers (Ono et al., 1983; Valenzuela et al., 1980), group B consisted of four subtype adw genomes including the three reported here, and group D contained four subtype ayw genomes. These three groups were homogeneous in terms of HBsAg subtypes. Group C was heterogeneous, including three adw, four adr and one ayr genomes. Accordingly, by the criterion of sequence homology of more than
The complete nucleotide sequence of HBV DNA clones of subtype adw. The sequence is shown for pJDW233 originated in the mainland Japan (a), pODW282 in Okinawa Prefecture (b) and pIDW420 in Indonesia (c). Since the clones did not possess an EcoRI site, nucleotides were numbered from a hypothetical EcoRI site as in the HBV DNA clones already reported (Galibert et al., 1979; Ono et al., 1983; Okamoto et al., 1986).

Fig. 1. The complete nucleotide sequence of HBV DNA clones of subtype adw. The sequence is shown for pJDW233 originated in the mainland Japan (a), pODW282 in Okinawa Prefecture (b) and pIDW420 in Indonesia (c). Since the clones did not possess an EcoRI site, nucleotides were numbered from a hypothetical EcoRI site as in the HBV DNA clones already reported (Galibert et al., 1979; Ono et al., 1983; Okamoto et al., 1986).
Table 1. Pairwise analysis of divergence in the complete nucleotide sequences of 18 clones of HBV DNA of various subtypes*

<table>
<thead>
<tr>
<th>Group and clone†</th>
<th>Subtype</th>
<th>Length (bp)</th>
<th>Origin</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>adw</td>
<td>3221</td>
<td>U.S.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 2</td>
<td>adw</td>
<td>3200</td>
<td>U.S.A.</td>
<td>1-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B 3</td>
<td>adw</td>
<td>3215</td>
<td>Japan</td>
<td>8-3</td>
<td>8-4</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B 4</td>
<td>adw</td>
<td>3215</td>
<td>Japan</td>
<td>9-3</td>
<td>9-2</td>
<td>3-9</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B 5</td>
<td>adw</td>
<td>3215</td>
<td>Indonesia</td>
<td>9-2</td>
<td>9-3</td>
<td>5-6</td>
<td>4-5</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B 6</td>
<td>adw</td>
<td>3215</td>
<td>Indonesia</td>
<td>9-2</td>
<td>9-3</td>
<td>5-4</td>
<td>4-4</td>
<td>0-6</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 7</td>
<td>adw</td>
<td>3215</td>
<td>Japan</td>
<td>8-9</td>
<td>8-7</td>
<td>9-0</td>
<td>8-3</td>
<td>8-1</td>
<td>8-2</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 8</td>
<td>adw</td>
<td>3215</td>
<td>Japan</td>
<td>8-9</td>
<td>8-7</td>
<td>9-0</td>
<td>8-3</td>
<td>8-1</td>
<td>8-2</td>
<td>0-2</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 9</td>
<td>adw</td>
<td>3215</td>
<td>Indonesia</td>
<td>8-5</td>
<td>8-4</td>
<td>9-2</td>
<td>8-5</td>
<td>8-1</td>
<td>8-1</td>
<td>2-8</td>
<td>2-8</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 10</td>
<td>adr</td>
<td>3215</td>
<td>Japan</td>
<td>9-1</td>
<td>8-9</td>
<td>9-5</td>
<td>8-8</td>
<td>8-3</td>
<td>8-5</td>
<td>1-8</td>
<td>1-7</td>
<td>3-3</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 11</td>
<td>adr</td>
<td>3188</td>
<td>Japan</td>
<td>8-7</td>
<td>8-5</td>
<td>9-0</td>
<td>8-4</td>
<td>8-1</td>
<td>8-2</td>
<td>1-7</td>
<td>1-7</td>
<td>2-8</td>
<td>2-1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 12</td>
<td>adr</td>
<td>3214</td>
<td>Japan</td>
<td>9-3</td>
<td>9-1</td>
<td>9-6</td>
<td>8-9</td>
<td>8-4</td>
<td>8-6</td>
<td>2-3</td>
<td>2-3</td>
<td>3-1</td>
<td>2-7</td>
<td>2-9</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 13</td>
<td>adr</td>
<td>3215</td>
<td>Japan</td>
<td>8-7</td>
<td>8-5</td>
<td>9-1</td>
<td>8-5</td>
<td>8-0</td>
<td>8-1</td>
<td>1-3</td>
<td>1-3</td>
<td>2-9</td>
<td>2-6</td>
<td>2-2</td>
<td>2-2</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 14</td>
<td>ayr</td>
<td>3215</td>
<td>Japan</td>
<td>8-8</td>
<td>8-7</td>
<td>9-4</td>
<td>8-6</td>
<td>8-3</td>
<td>8-4</td>
<td>1-9</td>
<td>1-8</td>
<td>2-9</td>
<td>2-2</td>
<td>2-2</td>
<td>2-2</td>
<td>1-9</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 15</td>
<td>ayw</td>
<td>3182</td>
<td>France</td>
<td>9-6</td>
<td>9-5</td>
<td>9-8</td>
<td>10-4</td>
<td>9-9</td>
<td>9-9</td>
<td>9-4</td>
<td>9-3</td>
<td>9-8</td>
<td>9-6</td>
<td>9-8</td>
<td>9-5</td>
<td>9-5</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 16</td>
<td>ayw</td>
<td>3182</td>
<td>U.S.S.R.</td>
<td>9-7</td>
<td>9-7</td>
<td>9-7</td>
<td>10-2</td>
<td>9-8</td>
<td>9-9</td>
<td>9-6</td>
<td>9-4</td>
<td>9-4</td>
<td>10-0</td>
<td>9-9</td>
<td>9-9</td>
<td>9-6</td>
<td>9-5</td>
<td>9-5</td>
<td>2-7</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D 17</td>
<td>ayw</td>
<td>3182</td>
<td>Japan</td>
<td>10-1</td>
<td>10-1</td>
<td>10-3</td>
<td>11-0</td>
<td>10-6</td>
<td>10-6</td>
<td>9-8</td>
<td>9-9</td>
<td>9-6</td>
<td>10-3</td>
<td>9-8</td>
<td>10-3</td>
<td>9-9</td>
<td>10-0</td>
<td>4-6</td>
<td>5-1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D 18</td>
<td>ayw</td>
<td>3182</td>
<td>New Guinea</td>
<td>9-5</td>
<td>9-7</td>
<td>9-8</td>
<td>10-3</td>
<td>10-0</td>
<td>10-0</td>
<td>9-3</td>
<td>9-3</td>
<td>9-1</td>
<td>9-6</td>
<td>9-4</td>
<td>9-8</td>
<td>9-3</td>
<td>9-3</td>
<td>9-2</td>
<td>3-7</td>
<td>3-9</td>
<td>4-8</td>
</tr>
</tbody>
</table>

* A total of 18 clones of HBV DNA were subjected to pairwise analysis of differences in nucleotide sequence. When the total number of nucleotides was different for two clones the comparison included only those nucleotides in corresponding positions in each clone; regions of deleted or inserted nucleotides were excluded from the calculation of divergence. A difference of 8.0% or greater is shown in bold type.

† Clone 1, pHBV3200 (Valenzuela et al., 1980); clone 2, pHBV933 (Ono et al., 1983); clone 3, pJDW233 (Fig. 1); clone 4, pODW282 (Fig. 1); clone 5, pIDW420 (Fig. 1); clone 6, pRTB299 (Sastrosoewignjo et al., 1985); clone 7, pSK619 (Okamoto et al., 1987); clone 8, pAK66; clone 9, pJK6146; clone 10, pHBV1-1 (Kobayashi & Koike, 1984); clone 11, pHBR330 (Ono et al., 1983); clone 12, pBRHBrdrd (Fujiyama et al., 1983); clone 13, pNR260; clone 14, pYRB259 (Okamoto et al., 1986); clone 15, Eco HBV DNA (Galibert et al., 1979); clone 16, pHBV320 (Bichko et al., 1985); clone 17, pYWB796; clone 18, pPYW310. The sequences for clones 8, 9, 13, 17 and 18, as well as their proposed genetic organizations (unpublished data) are available upon request from the authors.
92%, ayw, adr and ayr genomes were homogeneous, but adw genomes were not. Nine adw genomes distributed into groups A, B and C, and those in group C were closely related to adr and ayr genomes.

Comparisons were made of amino acid sequences, as well as nucleotide sequences, of the four ORFs of the HBV clones (Table 2). Again, the intra-group divergence was much smaller than the inter-group divergence. The intra-group variation in translation products was roughly in parallel with that in the nucleotide sequences for the pre-S region/S gene, X gene and P gene. For the pre-C region/C gene, the variation in translation products was small, out of proportion to the variation in nucleotide sequences.

DISCUSSION

Variants of viruses are most conveniently identified by means of distinct antigenic determinants carried on their envelopes. This also applies to HBV where the four major subtypes of HBsAg can be discriminated by a variety of immunological methods (Le Bouvier, 1971; Holland et al., 1972; Hollinger et al., 1973; Imai et al., 1974; Miyakawa et al., 1975; Hoofnagle et al., 1977; Usuda et al., 1986). The four subtypes of HBsAg arise from the combination of two pairs of mutually exclusive determinants, d and y, and w and r (Le Bouvier, 1971; Bancroft et al., 1972). These differences are attributed to the substitution of amino acid residues in the 226 residue translation product of the S gene (Gerin et al., 1983; Peterson et al., 1984). Without any other means of classifying HBV strains, and on the premise that the four subtypes of HBsAg represent the phenotypic expression of the four major genotypes of HBV (Le Bouvier et al., 1972), these subtypes have been widely employed in clinical, virological and epidemiological studies (Mayumi & Nakajima, 1973; Mazzur et al., 1974; Yamashita et al., 1975; Ishimaru et al., 1976; Okada et al., 1976; Courouce-Pauty et al., 1983).

The genomic variation recognizable by the antigenicity of the envelope polypeptides might, however, be the tip of the iceberg. There could be considerable sequence divergence throughout the entire HBV genome, most of which would not affect the epitopes carried by envelope polypeptides. The HBV genome is subject to variation, and has an estimated nucleotide substitution rate per site per year of $1.4 \times 10^{-5}$ to $3.2 \times 10^{-5}$ (Okamoto et al., 1987a). This is of the same order as the substitution rate of some RNA retroviruses for which the value is in the order of $10^{-5}$ (Holland et al., 1982). The estimated rate of HBV substitution is 102-fold lower than that of the env gene of human immunodeficiency virus (Hahn et al., 1986), and 104-fold higher than most DNA genomes (Holland et al., 1982).

The similarity of HBV to retroviruses, its replication by reverse transcription of an RNA intermediate in particular, has been pointed out (Miller & Robinson, 1986; Summers & Mason, 1982; Toh et al., 1983). RNA viruses including retroviruses are susceptible to mutations, probably because they do not have proof-reading enzymes for correcting errors during duplication (Steinhauer & Holland, 1986). Indeed, the high mutability may be taken as another characteristic of HBV that resembles the properties of retroviruses.

We determined the complete nucleotide sequences of the three HBV DNA clones of subtype adw, from Japan and Indonesia. They showed sequence variation of 3.9 to 5.6%. A total of 18 HBV DNA clones, including the three reported here, 10 previously reported and five unpublished clones of various subtypes, were evaluated for sequence homology. HBV genomes were tentatively classified into four groups, which we called A, B, C and D, with divergence between groups of 8.0% or greater and variation within groups of 5.6% or less.

Aside from the sequence homology, some characteristics intrinsic to each group have emerged. For example, both of the two HBV DNA clones of subtype adw in group A have a 6 bp insertion in the C gene (nucleotides 2355 to 2356 counting from the EcoRI site) that is not shared by any of the other 16 clones. All four HBV DNA clones of subtype ayw in group D have a genomic length of 3182 bp, shorter than the other 14 clones. This is due to a 33 bp deletion spanning nucleotides 2850 to 2882 that does not occur in the other clones.

The geographical distribution of HBV DNA sequences was evident in groups A, B and C. Group A consisted of two clones, pHBV3200 reported from the U.S.A. by Valenzuela et al. (1980) and pHBV933 reported from Japan by Ono et al. (1983). The latter clone, however, also
Table 2. Percentage differences in the amino acid sequences of translation products of, and in the nucleotide sequences of, the four ORFs of HBV DNA clones within and among groups A, B, C and D*

<table>
<thead>
<tr>
<th>ORFs</th>
<th>Intra-group divergence</th>
<th>Inter-group divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within A</td>
<td>Within B</td>
</tr>
<tr>
<td>Pre-S region and S gene</td>
<td>2.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Pre-C region and C gene</td>
<td>1.9</td>
<td>3.0</td>
</tr>
<tr>
<td>X gene</td>
<td>1.9</td>
<td>2.7</td>
</tr>
<tr>
<td>P gene</td>
<td>†</td>
<td>4.8</td>
</tr>
</tbody>
</table>

* Comparison was made only for the amino acid residues or nucleotides in positions that were commonly occupied in both of the HBV DNA clones in question. All possible combinations of two members belonging to the same group, or belonging to distinct groups, were examined for sequence differences and the results were expressed as the mean percentage difference. Differences in nucleotide sequences, similarly calculated, are indicated in parentheses.

† Comparison was not feasible, because one of the two members (pHBV933; Ono et al., 1983) had a stop codon in the P gene that prevented the translation of the putative DNA polymerase.
had a U.S. origin; it was derived from a chimpanzee experimentally infected with HBV and supplied by Dr D. W. Bradley of the Center for Disease Control, then in Phoenix, Arizona. Groups B and C were composed of four and eight clones, respectively, all of which originated in Asia. All of the four clones in group B were of subtype adw. Group C was heterogeneous as regards subtypes, including three adw, four adr and one ayr clones. The three adw clones in group C, therefore, were closer in nucleotide sequence to adr and ayr clones than they were to the adw clones in groups A and B.

Group D comprised four ayw clones from different countries. It is not certain why a geographical grouping is not obvious in group D. The spatial distribution of HBsAg subtypes would have been maintained, at least before the introduction of medical procedures, by the transmission of HBV through close body contacts, typified by mother-to-baby transmission. The spread of subtype ayw might have been affected by a specific factor not acting on the other three subtypes. Of possible relevance to this is an epidemic spread of HBV subtype ayw among drug abusers (Magnius et al., 1973; Nielsen et al., 1973).

The results obtained highlighted a marked genomic heterogeneity of HBV subtype adw. There are at least three subgroups of HBV subtype adw differing in nucleotide sequence to an extent comparable with inter-subtypic variations. Courouce-Pauty et al. (1983) have reported four subgroups of HBsAg subtype ayw, distinguishable by monospecific antibody, which they designated ayw1, ayw2, ayw3, and ayw4. In addition, they recognized two subgroups of HBsAg subtype adw that they called adw2 and adw4. It remains to be seen how their w subspecificities correlate with the HBV subtype adw genomes belonging to groups A, B or C.

Recently, the nucleotide sequence has been reported for HBV isolated from a naturally infected chimpanzee (Vaudin et al., 1988). The sequence shows approximately 10% divergence from any of the other 18 HBV DNA clones, and therefore is not classified into group A, B, C or D. Whether the chimpanzee HBV represents a unique strain, having evolved to adapt to a host other than human beings, or represents a fifth group of HBV that may be prevalent in Africa, is not clear at the moment. Determination of complete nucleotide sequences of HBV strains derived from human carriers in Africa will help in answering this question.

When the translation products of the four ORFs were examined the divergence was smaller between members of the same group than between members of different groups (Table 2). Inter-group variation in amino acid sequences of translation products paralleled that in nucleotide sequences for the pre-S region/S gene, X gene and P gene. For the pre-C region/C gene, however, amino acid variation was small, out of proportion to nucleotide variation.

This suggests that conservation of the amino acid sequence encoded by the pre-C region/C gene is essential for HBV. The amino acid sequence of the translation product of the retrovirus gag gene (coding for the internal structural proteins of the virion and considered to correspond to the pre-C region/C gene of HBV) is also conserved (Coffin, 1984; Alizon et al., 1986; Hahn et al., 1986). The env gene (coding for the proteins found on the surface of the virion envelope, corresponding to the pre-S region/S gene of HBV), in contrast, appears to be more liable to nucleotide substitutions, enabling escape from the immune defences of the host.

On the basis of our results, the HBsAg subtypes do not truly reflect the genotypic variation of HBV isolates. This may not come as a surprise, knowing that the d or y determinant is specified by lysine or arginine at amino acid position 122 in the S gene product, and that of w or r is specified by lysine or arginine at position 160 (Okamoto et al., 1987b).

It would be worthwhile to re-group HBV genotypes by the criterion of sequence homology, in an attempt to seek their possible significance in virological, clinical and epidemiological terms. Although the determination of entire genomic sequences is obviously not practical in extended studies, HBV genomes belonging to a given group may have distinct cleavage site(s) for particular endonucleases, thus allowing them to be distinguished by restriction mapping. A first step in this direction is provided by the BamHI site at nucleotide 2937 which is common to group D genomes but is not present in the genomes of members of the other three groups (A, B and C).

REFERENCES

Genotypes of hepatitis B virus


(Received 29 April 1988)