J. gen. Virol. (1988), **69**, 1383–1389. *Printed in Great Britain*

Key words: *hepatitis B virus/nucleotide sequence/divergence*

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 *Eco*RI site common to most strains of HBV and a convenient single-cut enzyme was not readily identifiable, it was cloned as two separate *Pst*I fragments of 1758 and 1424 nucleotides (nt) into pBR328. The *Pst*I 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 *Pst*I 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.

```
 M  D  I  D  P  Y  K  E  F  G  A  T  V  E  L  L  S  F  L  P  S  D  F  F  P  S  V  R  D  L  L  D  T  A  S  A  L  Y  R  E
ATGGACATTGACCCTTATAAAGAATTTGGAGCTACAGTGGAGTTACTCTCTTTTTTGCCTTCTGATTTCTTTCCGTCGGTCCGTGATCTCCTCGACACCGCCTCAGCTCTTTACCGGGAA
        10        20        30        40        50        60        70        80        90       100       110       120

 A  L  E  S  P  E  H  C  S  P  N  H  T  A  L  R  Q  A  I  L  C  W  G  E  L  M  T  L  A  S  W  V  G  N  N  L  E  D  P  A
GCCTTAGAGTCTCCAGAGCATTGTTCACCTAACCATACAGCACTTAGGCAAGCTATACTGTGCTGGGGTGAGTTAATGACTCTGGCCTCCTGGGTGGGCAATAATTGGAAGATCCAGCA
       130       140       150       160       170       180       190       200       210       220       230       240

 S  R  E  Q  V  V  N  Y  V  N  T  N  M  G  L  K  I  R  Q  L  L  W  F  H  I  S  C  L  T  F  G  R  E  T  V  L  E  Y  L  V
TCCAGGGAACAAGTAGTTAATTATGTCAATACCAATATGGGTTAAAGATCAGACAATTATTGTGGTTTCATATTTCCTGTCTTACTTTTGGAAGAGAAACTGTCCTTGAGTATTTGGTG
       250       260       270       280       290       300       310       320       330       340       350       360

                          Polymerase      M  P  L  S  Y  Q  H  F  R  K  L  L  L  L  D  D  E  A  G  P  L  E  E  E  L
 S  F  G  V  W  I  R  T  P  P  A  Y  R  P  P  N  A  P  I  L  S  T  L  P  E  T  T  V  V  R  R  R  G  R  S  P  R  R  R  T
TCTTTTGGAGTGTGGATTCGCACTCCCCCCCCTTATAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACT
       370       380       390       400       410       420       430       440       450       460       470       480

 P  R  L  A  D  E  G  L  N  R  R  V  A  E  D  L  N  L  Q  L  P  N  V  S  I  P  W  T  H  K  V  G  N  F  T  G  L  Y  S  S
 P  S  P  R  R  R  R  S  Q  S  P  R  R  R  R  S  Q  S  P  A  S  Q  C  *      Core
CCCTCGCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCCAGCTTCCCAATGTTAGTATTCCTTGGACTCATAAGGTGGGAAATTTTACTGGGCTTTATTCTTC
       490       500       510       520       530       540       550       560       570       580       590       600

 T  L  P  V  F  N  P  N  W  Q  T  P  S  F  P  D  I  H  L  H  Q  D  I  I  N  K  C  E  Q  F  V  G  P  L  T  V  N  E  K  R
TACTCTACCTGTCTTTAACCCTAACTGGCAAACTCCTTCTTTCCCTGATATTCATTGCACCAAGATATTATTAACAAGTGTGAACAATTTGTGGGCCCTCTTACGGTTAATGAAAAAAG
       610       620       630       640       650       660       670       680       690       700       710       720

 R  L  K  L  S  M  P  A  R  F  Y  P  N  S  T  K  Y  L  P  L  E  K  G  I  K  P  Y  Y  P  D  N  V  V  N  H  Y  F  Q  T  R
AAGATTGAAGTTAAGTATGCCTGCCAGATTTTATCCAAATTCTACCAAGTATTTGCCTCTAGAGAAAGGTATAAAACCCTATTATCCAGATAATGTAGTTAATCATTACTTCCAAACTAG
       730       740       750       760       770       780       790       800       810       820       830       840

 H  Y  L  H  T  L  W  Q  A  G  I  L  Y  K  R  Q  T  T  R  S  A  S  F  C  G  S  P  Y  S  W  E  Q  E  L  Q  H  G  A  E  S
                                                                                    Surface          M  G  Q  N
ACACTATTTACATACTCTATGGCAGGCAGGCATTTTATATAAGAGAGAAACCACACGTAGCGCTTCATTCTGTGGGTCACCATATTCTTGGGAACAAGAGCTACAGCATGGGGCAGAATC
       850       860       870       880       890       900       910       920       930       940      950       960

 F  H  Q  Q  S  A  G  I  F  S  R  A  P  V  G  S  S  I  Q  S  K  H  Q  Q  S  R  L  G  L  Q  P  Q  K  Q  L  L  A  R  G  N
 L  S  T  S  N  P  L  G  F  F  P  E  H  Q  L  D  P  A  F  K  A  N  T  N  N  P  D  W  D  F  N  P  K  K  D  Y  W  P  E  A
TTTCCACCAGCAATCCGCTGGGATTTTTTCCCGAGCACCAGTTGGATCCAGCATTCAAAGCAAACACCAACAATCCAGATTGGGACTTCAACCCCAAAAAGGACTACTGGCCAGAGGCAA
       970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080

 E  G  R  S  W  S  V  R  S  R  V  H  P  T  T  W  R  S  F  G  V  E  P  S  S  S  G  H  T  N  N  F  A  S  K  S  A  S  C  L
 T  K  V  G  A  G  A  F  G  P  G  F  T  P  P  H  G  G  L  L  G  L  S  P  Q  A  Q  G  I  L  T  T  L  P  A  N  P  P  P  A
CGAAGGTAGGAGCTGGAGCGTTCGGTCCAGGGTTCACCCCACCACATGGAGGTCTTTTGGGGTTGAGCCCTCAAGCTCAGGGCATACTAACAACTTTGCCAGCAAATCCGCCTCCTGCCT
       1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200

 H  Q  S  V  V  R  K  A  A  Y  P  T  F  S  T  T  K  R  H  S  S  S  G  H  A  V  E  L  H  N  I  S  S  S  S  A  G  S  Q  S
 S  T  N  R  Q  S  G  R  Q  P  T  P  L  S  P  P  L  R  D  T  H  P  Q  A  M  Q  W  N  S  T  T  F  H  Q  A  L  Q  D  P  R
CCACCAATCGGCAGTCAGGAAGGCAGCCTACCCCACTTCTCCACCACTAAGAGACACTCATCCTCAGGCCATGCAGTGGAACTCCACAACATTTCATCAAGCTCTGCAGGATCCCAGAG
       1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

 K  G  P  V  F  S  C  W  W  L  Q  F  R  N  I  E  P  C  S  E  Y  C  L  S  H  L  V  S  L  L  D  D  W  G  P  C  T  E  H  G
 V  R  G  L  Y  F  P  A  G  G  S  S  S  G  T  L  N  P  V  P  N  T  A  S  H  I  S  S  V  F  S  T  T  G  D  P  A  P  N  M
TAAGGGGCCTGTATTTTCCTGCTGGTGGCTCCAGTTCAGGAACATTGAACCCTGTTCCGAATACTGCCTCTCACATCTCGTCAGTCTTCTCGACGACTGGGGACCCTGCACCGAACATGG
       1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430     1440

 E  H  H  I  R  I  P  R  T  P  A  R  V  T  G  G  V  F  L  V  D  K  N  P  H  N  T  A  E  S  R  L  V  V  D  F  S  Q  F  S
 E  N  I  T  S  G  F  L  G  P  L  L  V  L  Q  A  G  F  F  L  L  T  K  I  L  T  I  P  Q  S  L  D  S  W  W  T  S  L  N  F
AGAACATCACATCAGGATTCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAAAATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTC
       1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

 R  G  S  T  R  V  P  W  P  K  F  A  V  P  N  L  Q  P  L  T  N  L  L  S  S  N  L  S  W  L  S  L  D  V  S  A  A  F  Y  H
 L  G  G  A  P  V  C  L  G  Q  N  S  Q  S  P  T  S  N  H  S  P  T  S  C  P  P  I  C  P  G  Y  R  W  M  C  L  R  R  F  I
TAGGGGGAGCACCCGTGTGCCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCCTCCAATTTGTCCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATCA
       1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680

 L  P  L  H  P  A  A  M  P  H  L  L  V  G  S  S  G  L  S  R  Y  V  A  R  L  S  S  N  S  R  I  L  D  H  Q  H  G  T  M  Q
 I  F  L  F  I  L  L  L  C  L  I  F  L  L  V  L  L  D  Y  Q  G  M  L  P  V  C  P  L  I  P  G  S  S  T  T  S  T  G  P  C
TCTTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTTGGTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCCTCTAATTCCAGGATCCTCGACCACCAGCACGGGACCATGCA
       1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800

 N  L  H  D  S  C  S  R  N  L  F  D  S  L  M  L  L  Y  K  T  F  G  R  K  L  H  L  Y  S  H  P  I  I  M  G  F  R  K  I  P
 K  T  C  T  T  P  A  Q  G  T  S  L  I  P  S  C  C  C  T  K  P  S  D  G  N  C  T  C  I  P  I  P  S  S  W  A  F  A  K  F
AAACTTGCACGACTCCTGCTCAAGGAACCTCTTTGATTCCCTCATGTTGCTGTACAAAACCTTCGGACGGAAATTGCACCTGTATTCCCATCCATCATGGGCTTTCGCAAAATTCC
       1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920

 M  G  V  G  L  S  P  F  L  L  A  Q  F  Y  S  A  I  C  S  V  V  R  R  A  F  P  H  C  L  A  F  S  Y  M  D  D  V  V  L  G
 L  W  Q  W  A  S  V  R  F  S  W  L  S  L  L  A  P  F  V  Q  W  F  A  G  L  S  P  T  V  W  L  L  A  I  W  M  M  W  Y  W
TATGGGAGTGGGCCTCAGTCCGTTTCTCCTGGCTCAGTTTACTAGCGCCATTTGTTCAGTGGTTCGCAGGGCTTTCCCCCACTGTTTGGCTTTTAGCTATATGGATGATGTGGTATTGGG
       1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040

 A  K  S  V  Q  H  H  E  S  L  Y  T  A  V  T  N  F  L  L  S  L  G  I  H  L  N  P  N  K  T  K  R  W  G  Y  S  L  H  F  M
 G  P  N  L  Y  N  I  L  S  P  F  I  P  L  L  P  I  F  F  C  L  W  V  Y  I  *      Surface
GGCCAAATCTGTACAACATCTTGAGTCCCTTTATACCGCTGTTACCAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAACGCTGGGGTTATTCCCTACATTTCAT
       2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160

 G  Y  V  I  G  S  W  G  T  L  P  Q  E  H  I  V  Q  K  T  K  N  C  F  R  K  L  P  V  N  R  P  I  D  W  K  V  C  Q  R  I
GGGTTATGTAATTGGCAGTTGGGGGACATTACCACAAGAACATATTGTACAAAAATCAAAAATTGTTTCAGAAAACTTCCTGTAACAGACCTATAGATTGGAAAGTATGTCAAAGAAT
       2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

 V  G  L  L  G  F  A  A  P  F  T  Q  C  G  Y  P  A  L  M  P  L  Y  A  C  I  Q  A  K  Q  A  F  T  F  S  P  T  Y  K  A  F
TGTGGGTCTTTTGGGATTTGCTGCCCCTTTTACGCAATGTGGTTATCCTGCGTTAATGCCATTGTATGCATGTATACAAGCAAAACAGGCTTTCACTTTCTCGCCAACTTATAAGGCCTT
       2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

 L  S  Q  Q  Y  S  T  L  Y  P  V  A  R  Q  R  S  G  L  C  Q  V  F  A  D  A  T  P  T  G  W  G  L  V  M  G  H  Q  R  M  R
TCTAAGTCAACAATATTCGACCCTTTACCCCGTTGCCCGGCAACGGTCCGGTCTGTGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCATGCG
       2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
```

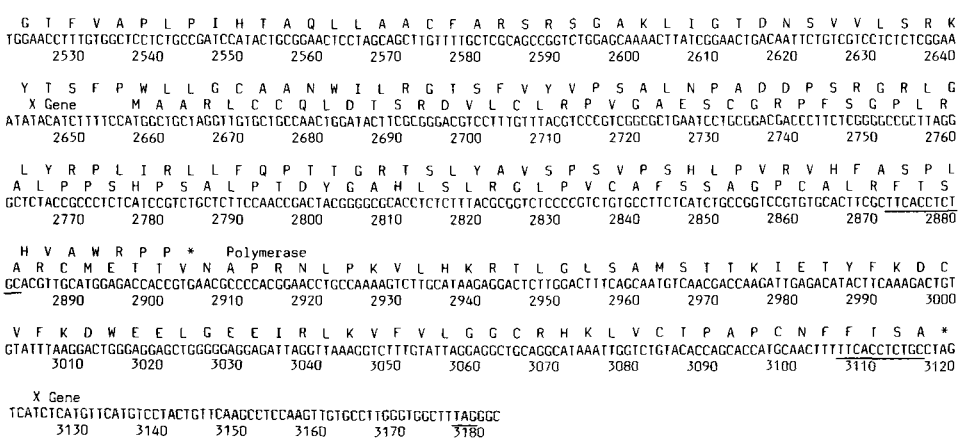

Fig. 1. The complete nucleotide sequence of the LSH strain of HBV. Position 1 is the start of the core region. The predicted amino acid sequences of the four major open reading frames are shown above the nucleotide sequence. The following sequences are underlined: the stop codon in the pre-core region (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 sequence includes most of the expected features of the HBV genome. The 11 bp direct repeats, which are believed to be central to the replication strategy of the virus (for review, see Ganem & Varmus, 1987), are located at positions 2872 and 3106. The enhancer element (TGTTTGCT; Shaul & Ben-Levy, 1987) is located at position 2463. A consensus sequence which may function as a specific glucocorticoid receptor site (CAANNTGTYCT; Tur-Kaspa *et al.*, 1986) is located at nucleotide position 1636. The extended TATA box (TATATAA) which is presumed to be the pre-S1 promoter (Cattaneo *et al.*, 1987) is located at position 876. A motif with homology to the simian virus 40 late promoter, which is presumed to be the promoter for the major 2·1 kb surface antigen transcript (Cattaneo *et al.*, 1983), is located around position 1225. A signal believed to be involved in the termination of all HBV transcripts (TATAAA) is located at position 16 and a second signal which may be involved in the cleavage/polyadenylation reaction (Hart *et al.*, 1985; Renan, 1987) is located at position 43 to 56.

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)

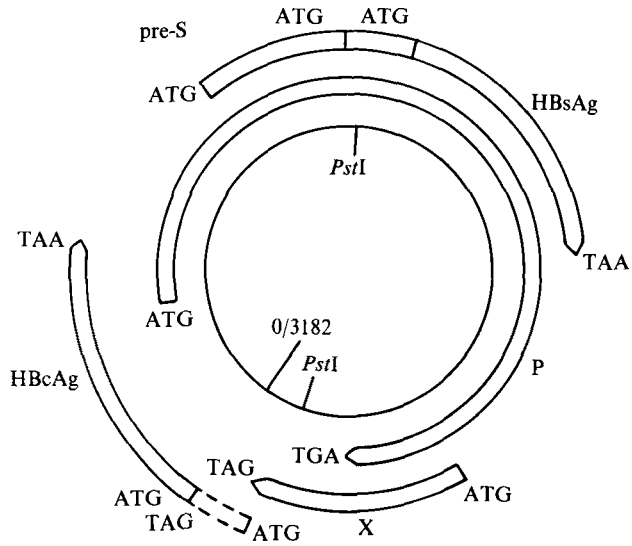

Fig. 2. A diagrammatic representation of the genome of the LSH strain of HBV. The inner circle shows the numbering system (clockwise, with the start of the HBcAg open reading frame as position 1) and the two *Pst*I sites used in the original cloning. The locations of the four major open reading frames 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	TACAGCATG----------GGGCAGAATCTT
ayw	TACAGCATG----------GGGCAGAATCTT
adw-2	TACAGCATG----------TCGCAAAGGCATGGGGACGAATCTT
adw-1	TACAGCATGGGAGGTTGGTCATCAAAACCTCGCAAAGGCATGGGGACGAATCTT
adr and *ayr*	TACAGCATGGGAGGTTGGTCTTCCAAACCTCGACAAGGCATGGGGACGAATCTT

Fig. 3. Comparison of the LSH sequence nt 942 to 962 with other HBV strains. Sequences are taken from Galibert *et al.* (1979; *ayw*), Ono *et al.* (1983; *adw*-2), Valenzuela *et al.* (1980; *adw*-1), Fujiyama *et al.* (1983; *adr*) and Okamoto *et al.* (1986; *ayr*).

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

	adw-1	*adw*-2	*ayw*	*ayw*-2	*adyw*	*adr*	*adr*-4	*ayr*
adw-LSH	9·8	9·5	9·5	9·4	9·7	9·6	10·0	10·0
ayr	9·7	9·6	10·2	10·3	10·9	2·4	2·2	
adr-4	9·4	10·0	10·5	10·7	11·4	2·0		
adr	8·9	9·9	10·6	11·5	12·0			
adyw	11·0	10·6	3·6	4·4				
ayw-2	11·0	10·1	2·7					
ayw	10·8	9·9						
adw-2	1·8							

* Comparisons are with the sequences of Valenzuela *et al.* (1980; *adw*-1), Ono *et al.* (1983; *adw*-2 and *adr*), Galibert *et al.* (1979, *ayw*), Bichko *et al.* (1985, *ayw*-2), Fujiyama *et al.* (1983, *adr*-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 & Schloemer, 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 *Pst*I fragment from a different plasma sample from the same animal (taken over 12 months later) and sequenced it from the *Pst*I 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

BICHKO, V., PUSHKO, P., DREILINA, D., PUMPEN, P. & GREN, E. (1985). Subtype *ayw* variant of hepatitis B virus. DNA primary structure analysis. *FEBS Letters* **185**, 208–212.

BURRELL, C. J., MACKAY, P., GREENAWAY, P. J., HOFSCHNEIDER, P. H. & MURRAY, K. (1979). Expression in *E. coli* of hepatitis virus DNA sequence cloned in plasmid pBR322. *Nature, London* **279**, 43–47.

CATTANEO, R., WILL, H., HERNANDEZ, N. & SCHALLER, H. (1983). Signals regulating hepatitis B surface antigen transcription. *Nature, London* **305**, 336–338.

CATTANEO, R., SPRENGEL, R., WILL, H. & SCHALLER, H. (1987). Molecular biology of hepatitis B. In *Primers in Developmental Biology*, vol. 2: *Molecular Genetics of Mammalian Cells*, pp. 271–295. Edited by G. M. Malacinski. London: Collier Macmillan.

CHANG, C., ENDERS, G., SPRENGEL, R., PETERS, N., VARMUS, H. E. & GANEM, D. (1987). Expression of the precore region of an avian hepatitis B is not required for viral replication. *Journal of Virology* **61**, 3322–3325.

FUJIYAMA, A., MIYANCHARA, A., NOZAKI, C., YONEYAMA, J., OHTOMO, N. & MATSUBARA, K. (1983). Cloning and structural analysis of hepatitis B virus DNAs, subtype *adr*. *Nucleic Acids Research* **11**, 4601–4610.

GALIBERT, F., MANDART, E., FITOUSSI, F., TIOLLAIS, P. & CHARNAY, P. (1979). Nucleotide sequence of the hepatitis B virus genome (subtype *ayw*) cloned in *E. coli*. *Nature, London* **281**, 646–650.

GANEM, D. & VARMUS, H. E. (1987). The molecular biology of hepatitis B viruses. *Annual Review of Biochemistry* **56**, 651–693.

HART, R. P., MCDEVITT, M. A., ALI, H. & NEVINS, J. R. (1985). Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and the early simian virus 40 polyadenylation sites. *Molecular and Cellular Biology* **5**, 2975–2983.

HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. *Gene* **28**, 351–359.

KOS, A., DIJKEMA, R., ARNBERG, A. C., VAN DER MEIDE, P. H. & SCHELLEKENS, H. (1986). The hepatitis delta (δ) virus possesses a circular RNA. *Nature, London* **323**, 558–560.

MAKINO, S., CHANG, M-F., SHIEH, C-K., KAMAHORA, T., VANNIER, D. M., GOVINDARAJAN, S. & LAI, M. M. C. (1987). Molecular cloning and sequencing of a human hepatitis delta (δ) virus RNA. *Nature, London* **329**, 343–346.

MILLER, R. H. (1987). Proteolytic self-cleavage of hepatitis B virus core protein may generate serum e antigen. *Science* **236**, 722–725.

NEURATH, A. R., KENT, S. B. H., STRICK, N. & PARKER, K. (1986). Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. *Cell* **46**, 429–436.

OKAMOTO, H., IMAI, M., SHIMOZAKI, M., HOSHI, Y., IIZUKA, H., GOTANDA, T., TSUDA, F., MIYAKAWA, Y. & MAYUMI, M. (1986). Nucleotide sequence of a cloned hepatitis B virus genome, subtype *ayr*: comparison with genomes of the other three serotypes. *Journal of General Virology* **67**, 2305–2314.

OKAMOTO, H., IMAI, M., TSUDA, F., TANAKA, T., MIYAKAWA, Y. & MAYUMI, M. (1987). Point mutation in the S gene of hepatitis B virus for a d/y or w/r subtypic change in two blood donors carrying a surface antigen of compound subtype *adyr* or *adwr*. *Journal of Virology* **61**, 3030–3034.

ONO, Y., ONDA, H., SASADA, R., IGARISHI, K., SUGINO, Y. & NISHIOKA, K. (1983). The complete nucleotide sequences of the cloned hepatitis B virus DNA: subtype *adr* and *adw*. *Nucleic Acids Research* **11**, 1747–1757.

OU, J. H., LAUB, O. & RUTTER, W. J. (1986). Hepatitis B virus gene function: the precore targets the core antigen to cellular membranes and causes the secretion of the E antigen. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 1578–1582.

RENAN, M. J. (1987). Conserved 12-bp element downstream from mRNA polyadenylation sites. *Gene* **60**, 245–254.

ROOSSINCK, M. J., JAMEEL, S., LOUKIN, S. H. & SIDDIQUI, A. (1986). Expression of hepatitis B viral core region in mammalian cells. *Molecular and Cellular Biology* **6**, 1393–1400.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SHAUL, Y. & BEN-LEVY, R. (1987). Multiple nuclear proteins in liver cells are bound to hepatitis B virus enhancer element and its upstream sequences. *EMBO Journal* **6**, 1913–1920.

SUMMERS, J. & MASON, W. S. (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**, 403–415.

TOH, H., HAYASHIDA, H. & MIYATA, T. (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature, London* **305**, 827–829.

TUR-KASPA, R., BURK, R. D., SHAUL, Y. & SHAFRITZ, D. A. (1986). Hepatitis B virus DNA contains a glucocorticoid-responsive element. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 1627–1631.

TWU, R-J. & SCHLOEMER, R. H. (1987). Transcriptional *trans*-activating function of hepatitis B virus. *Journal of Virology* **61**, 3448–3453.

UY, A., BRUSS, V., GERLICH, W. H., KOCHEL, H. G. & THOMSSEN, R. (1986). Precore sequence of hepatitis B virus inducing e antigen and membrane association of the viral core protein. *Virology* **155**, 89–96.

VALENZUELA, P., QUIROGA, M., ZALDIVAR, J., GRAY, P. & RUTTER, W. J. (1980). The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. In *Animal Virus Genetics*, pp. 57–70. Edited by B. Fields, R. Jaenisch & C. F. Fox. New York & London: Academic Press.

WANG, K-S., CHOO, Q-L., WEINER, A. J., OU, J-H., NAJARIAN, R. C., THAYER, R. M., MULLENBACH, G. T., DENNISTON, K. J., GERIN, J. L. & HOUGHTON, M. (1986). Structure, sequence and expression of the hepatitis delta (δ) viral genome. *Nature, London* **323**, 508–514.

WEIMER, T., SALFIELD, J. & WILL, H. (1987). Expression of the hepatitis B virus core gene *in vitro* and *in vivo*. *Journal of Virology* **61**, 3109–3113.

WILL, H., KUHN, C., CATTANEO, R. & SCHALLER, H. (1982). Structure and function of the hepatitis B virus genome. In *Primary and Tertiary Structure of Nucleic Acids and Cancer Research* , vol. 12, pp. 237–247. Edited by M. Miwa, S. Nishimura, A. Rich, D. G. Soell & T. Sugimura. Tokyo: Japan Scientific Society Press.

WILL, H., CATTANEO, R., DARAI, G., DEINHARDT, F., SCHELLEKENS, H. & SCHALLER, H. (1985). Infectious hepatitis B virus from cloned DNA of known nucleotide sequence. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 891–895.

ZUCKERMAN, A. J., THORNTON, A., HOWARD, C. R., TSIQUAYE, K. N., JONES, D. M. & BRAMBELL, M. R. (1978). Hepatitis B outbreak among chimpanzees at the London Zoo. *Lancet* ii, 652–654.

(Received 10 November 1987)