Journal of General Virology (2000), **81**, 75–83. Printed in Great Britain

Long-term mutation rates in the hepatitis B virus genome

Charles Hannoun, Peter Horal and Magnus Lindh

Department of Clinical Virology, Göteborg University, Guldhedsgatan 10B, 413 46 Göteborg, Sweden

Mutations in the hepatitis B virus (HBV) genome have so far been investigated in cross-sectional or short-term longitudinal studies. Information about long-term changes is lacking due to the difficulty of sampling over long observation periods. In this study, a retrospective approach was used that allowed the analysis of changes in the viral genome from transmission to late stages of infection without the requirement for sampling early during this period. The entire viral genome was sequenced from serum samples of three mothers and their 10 adult children, who presumably had been infected vertically. The emergence of mutations between birth and sampling (mean 26·5 years) was assessed by comparing the individual sequences with the sequence of the strain assumed to have been transmitted. The mean differences from this sequence were 0·02 and 0·28% in seven asymptomatic and one symptomatic hepatitis B e antigen (HBeAg)-positive carriers, respectively, and 0·62 % in five HBeAg-negative carriers. Mutations occurred throughout the genome and 88% of the mutations caused amino acid substitutions spread over all genes. In HBeAg-negative carriers, the number of nucleotide and amino acid changes was independent of the severity of liver disease and, except the ^{1762}AGG^{1764} → TGA changes, no specific mutation was associated with liver disease. In conclusion, by using a novel method it was found that the entire HBV genome is extremely stable over long periods of time during the HBeAg-positive phase if the immune response (inflammation) is weak, whereas an average of 20 mutations emerged after development of hepatitis and/or loss of HBeAg without association with clinical outcome.

Introduction

Chronic hepatitis B is an infection of major importance, with more than 300 million carriers worldwide, some of whom will develop progressive liver damage. It is generally accepted that liver damage is caused not by hepatitis B virus (HBV) itself, but by the immune response evoked by the virus. Mutations emerge frequently in the viral genome, perhaps selected under pressure from the immune response. Mutations in certain regions, in particular the core promoter (Okamoto *et al.*, 1994; Takahashi *et al.*, 1995; Lindh *et al.*, 1999), the pre-core (Carman *et al.*, 1989; Lok *et al.*, 1994; Lindh *et al.*, 1996) and core (Akarca & Lok, 1995; Carman *et al.*, 1997; Karasawa *et al.*, 1997) regions, have been thoroughly studied. It is, however, not clear to what extent and in what manner these mutations interfere with the course of infection and liver injury i.e. whether they are of pathogenetic importance or merely passive markers of these events.

Mutations in other regions of the HBV genome, in particular the polymerase ORF, have been insufficiently studied. Because studies of mutations in the entire genome are rare, the frequency and importance of mutations in these coding regions are not known. Moreover, long-term longitudinal studies are largely lacking, due to the difficulty of sampling over long periods of time and, thus, the mutation rates in the different stages of HBV infection are not yet well established. So far, mutations have mainly been assessed in cross-sectional studies by comparison with a consensus sequence representing the serotype or genotype of the strain analysed, but it is not known to what extent this strategy reflects changes taking place during infection in a patient.

In the present study, we have examined the entire viral genome obtained from serum samples from three mothers with chronic HBV infection and their 10 adult children, who presumably were infected vertically and are now in different stages of chronic infection. By using these sequences, we have

Author for correspondence: Magnus Lindh.
Fax +46 31 827032. e-mail magnus.lindh@microbio.gu.se

The GenBank accession numbers of the sequences reported in this paper are AF121239–AF121251.

traced the likely 'original' sequence that was probably transmitted at birth. Comparison of this sequence with the sequences now present in the mothers and children allowed a retrospective longitudinal analysis of rates and distribution of mutations that have emerged in the entire viral genome over periods of up to 35 years.

Methods

■ **Patients.** Serum samples were analysed from 13 chronic HBV carriers originating from Vietnam and Turkey. As described in Table 1, the patients were three mothers (Mo) and their adult children (Ch; three, three and four children), who probably contracted the HBV infection vertically. This assumption was based on the fact that the majority of all of the children of these mothers (6/6, 5/5 and 5/6 children) were chronic HBV carriers. The liver damage of all the 13 patients was well characterized, including histological assessment by Knodell scoring in 10 of them. All patients had given informed consent to a liver biopsy, four of which were done on clinical indication and six in a clinical study approved by the local ethics committee. None of the patients had received interferon or antiviral treatment for hepatitis B.

■ **Genotyping and quantitative PCR.** Genotyping was done by phylogenetic comparison of full-length genomes with published sequences of all genotypes (A–F) and by RFLP analysis of S region amplicons (Lindh *et al.*, 1997).

HBV DNA levels in serum were analysed by Amplicor HBV Monitor (Roche Diagnostic Systems) according to the manufacturer's instructions. Samples with HBV DNA above the upper detection level ($10^{7\cdot0}$ copies/ml) were reanalysed after dilution by 1:10000 in HBV-negative serum.

■ **Histopathology and alanine aminotransferase (ALT) activity.** Liver damage was assessed by histology activity index (HAI) scoring in a blinded fashion (Knodell *et al.*, 1981). ALT activities from an observation period of 1–10 years prior to liver biopsy were recorded.

■ **Sequencing.** Overlapping segments covering the entire HBV genome were amplified by PCR with the primers described in Table 2. After gel-purification of PCR bands by using Qiaex (Qiagen), direct sequencing was carried out with the same primers as used for PCR. Each amplicon was analysed in both the sense and antisense directions after cycle sequencing reactions with fluorescent dye (Rhodamine; Applied Biosystems) terminators (ddNTPs). The sequence was read in an ABI Prism 310 automated capillary sequence reader (Applied Biosystems) and processed by using the Sequence Navigator software (Applied Biosystems).

■ **Phylogenetic analysis and assessment of mutations.** The complete DNA sequences of all virus strains were subjected to phylogenetic comparison including sequences obtained from the NCBI database (National Center for Biotechnology Information, Bethesda, MD, USA) of the same genotype (genotype B: D00329, D00330, D00331, D23678, D23679, D50522; and genotype D: J02203, M32138, X02496, Z35716 and X68292) and an outgroup sequence representing genotype A (L13994) with software (DNAdist and Neighbor) included in the PHYLIP package (Felsenstein, 1993; http://evolution.genetics.washington.edu/phylip.html).

Table 1. Clinical data and number of mutations in 13 HBV carriers from three families

HAI_{infl} is the sum of inflammatory scores for interface (piecemeal) inflammation, intralobular inflammation and portal inflammation; HAI_{fibr} is the score for fibrosis; ALT/URV, ALT/upper reference value ratio, given as the median and maximum values of the last five to ten values. NA, Not available.

Patient	Sex	Year of birth	Year of sample	Observed period of divergence (years)	Genotype	HBeAg	HBV DNA (log_{10} copies/ml)	HAI_{infl}	HAI_{fibr}	ALT/URV		Number of nt differences* (%)
										Median	Max.	
Fam1												
Mo	F	1948	1993	16†	D	−	8·2	9	4	1·2	6·2	19 (0·60)
Ch1	F	1963	1994	31	D	+	9·6	2	0	1·2	1·5	0 (0)
Ch2	M	1973	1995	22	D	+	9·5	3	1	1·2	1·4	1 (0·03)
Ch3	M	1977	1998	21	D	+	9·6	8	1	1·4	21	9 (0·28)
Fam2												
Mo	F	1938	1995	20†	B	+	8·5	NA	NA	0·44	0·5	0 (0)
Ch1	M	1970	1993	23	B	+	8·8	1	0	0·36	0·75	2 (0·04)
Ch2	M	1972	1992	20	B	+	9·5	3	1	0·66	1·0	0 (0)
Ch3	F	1975	1998	23	B	+	9·0	3	0	1·6	2·2	1 (0·02)
Fam3												
Mo	F	1937	1994	27†	B	−	5·5	9	4	0·5	1·1	27 (0·84)
Ch1	M	1962	1994	32	B	−	4·0	2	1	0·95	1·2	15 (0·47)
Ch2	F	1963	1998	35	B	+	9·3	2	0	0·39	0·53	0 (0)
Ch3	F	1964	1997	33	B	−	4·2	NA	NA	1·2	2·3	19 (0·59)
Ch4	F	1967	1997	30	B	−	4·6	NA	NA	0·56	0·77	20 (0·62)

* Versus deduced original sequence of each family.
† From birth of the youngest child.

Table 2. Primers used for amplification and sequencing of the entire HBV genome

Primer set	Sense primer(s)		Antisense primer(s)	
	Position*	Sequence	Position*	Sequence
1	2823–2845	TCACCATATTCTTGGGAACAAGA	476–458	GACAAACGGGCAACATACC
	3199–3216†	CATCCTCAGGCCATGCAG		
2	256–278	GTGGTGGACTTCTCTCAATTTTC	986–970	ACTTTCCAATCAATAGG
			796–776†	CGGTA(A/T)AAAGGGACTCA(A/C)GAT
3	635–654	TTCCTATGGGAGTGGGCCTC	1353–1333	GGAGAGCACGACAGAATTGTC
			1285–1267†	CTAGGAGTTCCGCAGTATG
4	1164–1190	GCCAGGTCTGTGCCAAGTGTTTGCTGA	1888–1866	CCCAAGGCACAGCTTGGAGGCTT
			1798–1779†	ACCAATTTATGCCTACAGCC
5	1603–1625	GTTGCATGGAGACCACCGTGAAC	2058–2039	GTATGGTGAGGTGAACAATG
	1680–1699†	ATGTCGACAACCGACCTTGA		
6	1865–1895	CAAGCCTCCAAGCTGTGCCTTGGGTGGCCTT	2488–2469	CCAGTAAAGTTCCCCACCTT
			2467–2448†	TGAGTCCAAGGAATACTAAC
7	2381–2400	AACTCCCTCGCCTCGCAGAC	80–61	TTCCTGAACTGGAGCCACCA
			3098–3079†	ATGCCCTGAGCCTGAGGGCT
8	2823–2845	TCACCATATTCTTGGGAACAAGA	80–61	TTCCTGAACTGGAGCCACCA
			74–54†	AACTGGAGCCACCAGCAGGAA

* Relative to the unique *Eco*RI site.
† Used as the inner primer in semi-nested PCR when required.

The approach for longitudinal sequence analysis was based on three conditions: (i) that transmission took place perinatally, (ii) that the children were infected by the 'original' strain and (iii) that there were no sequence changes during transmission, as indicated by nearly all previous transmission studies (see Günther *et al.*, 1999, pp. 36 and 50). To identify mutations that had emerged since transmission, the virus sequence now observed in each patient was compared with the probable sequence of the 'original' strain of that family. This sequence is a consensus sequence of all sequences of a family and was deduced as follows. If three sequences were identical and one different in a specific position, the original sequence was based on that of the three sequences and the aberrant sequence was thus interpreted as a mutant at this position. If more than one sequence differed from the others at a specific nucleotide position, as was the case for three nucleotide positions in family 1 (Fam1) and for five nucleotide positions in Fam3, deduction of the original sequence was aided by hepatitis B e antigen (HBeAg) status and published sequence data: the nucleotides present in HBeAg-positive carriers were used in the deduced original sequence if present in at least 50% of published sequences, otherwise this position in the original sequence was considered uncertain.

The DNA sequence from each patient was also compared with a consensus sequence representing genotypes B (Fam2 and Fam3) or D (Fam1), as deduced from database sequences obtained from NCBI with the accession numbers D00329, D00330, D00331, D23678, D23679, D50521, D50522, M54923, X97850, X97851 and X98077 (genotype B) and J02203, L27106, M32138, U95551, X02496, Z35716, X85254 and X80925 (genotype D).

Results

Clinical and virological data are presented in Table 1. Eight patients (including one mother) were HBeAg-positive with high virus loads. Seven of these had signs of only minimal or mild liver inflammation and one had severe inflammation with high ALT. Five patients were HBeAg-negative, with HBV DNA levels ranging from $10^{4.0}$ to $10^{8.2}$ copies/ml. Two of these had severe liver damage (cirrhosis) while two had minimal liver damage.

The genetic difference between sequences within each family ranged from 0 to 0·84% of the whole genome. All three families clustered distinctly in phylogenetic tree analysis, as shown in Fig. 1.

Family 1

The mother, 45 years old and HBeAg-negative, showed a high level of virus replication and severe liver damage, and in all 19 mutations within the virus genome. All but one of these caused amino acid substitutions, as described in Table 3. Mutations in the core promoter were seen at nt 1752, 1753, 1762 and 1764. A pre-core TAG stop mutation was present at nt 1896.

All three children were HBeAg-positive with high virus loads. Two of them had minimal inflammation, indicating immunological hyporeactivity or tolerance, and presented 0 and 1 mutations, respectively. The youngest son (Fam1 Ch3) was sampled 4 months after the onset of his first recognized flare of hepatitis, represented by high ALT and histologically chronic active hepatitis (inflammation HAI score = 8). His HBV genome contained nine mutations, including a pre-core TAG stop. Of these nine mutations, three, including the pre-core TAG mutation, were also seen in the mother (Fam1 Mo).

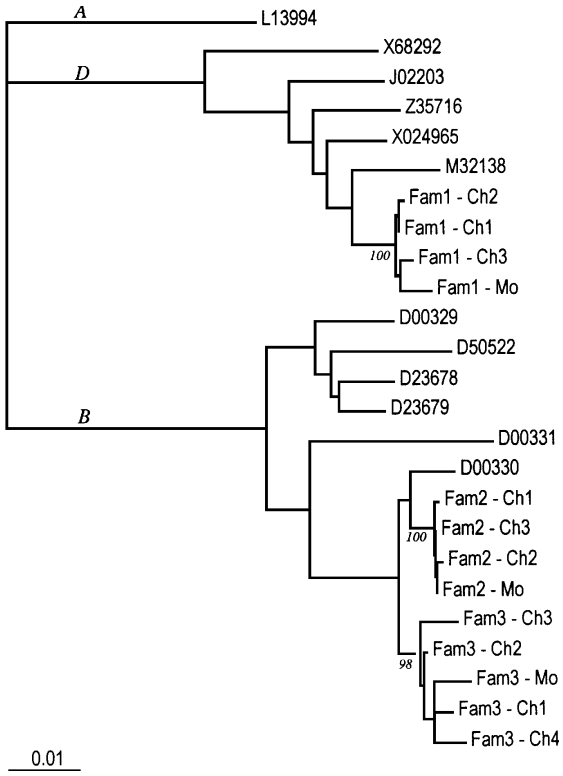

Fig. 1. Rooted phylogenetic tree based on distance matrix analysis using the DNAdist and Neighbor programs of the PHYLIP package. The entire HBV genome from four subjects of one family infected with genotype D (Fam1) and nine subjects from two families infected with genotype B strains (Fam2 and Fam3) were compared with database sequences representing genotypes A (outgroup; L13994), B and D. Relevant bootstrap values are shown (based on 100 replicates created by Seqboot).

Analysis of additional serum samples drawn just before and earlier during the flare (5 and 2·5 months before the sample that was sequenced) showed a pre-core wild-type strain only and mixed (50:50) wild-type/mutant infection, respectively (not shown), indicating that the pre-core TAG mutant had not been transmitted from the mother but had emerged *de novo* in this patient. This recent presence of a pre-core wild-type strain probably explains why this patient was HBeAg-positive at the time of biopsy despite the fact that only a mutant strain was detectable by sequencing.

The deletion in the pre-S region of nt 2854–2886, characteristic for genotype D, was present in all four cases.

Family 2

The mother (Fam2 Mo), aged 57, and all three children were HBeAg-positive with high virus loads and all had minimal or mild liver disease. The genomes in the mother and one child (Fam2 Ch2) showed no mutations, one child (Fam2 Ch1) had two mutations and one child (Fam2 Ch3) had one mutation (0–0·06%). Comparison with the genotype B consensus sequence indicated 59–61 mutations in the entire genome (about 1·8%), a significant overestimate of the mutations that had emerged during infection.

Family 3

At the time of sampling, the mother (Fam3 Mo) was HBeAg-negative and had severe liver damage with active cirrhosis. Four years later, at the age of 61, she developed hepatocellular carcinoma and died. One of the children was a highly viraemic (HBeAg-positive), asymptomatic carrier, one had a relatively mild HBeAg-negative hepatitis with a low level of virus replication and two were low-viraemic, healthy carriers.

In five positions, the virus genome from more than one subject in family 3 had a nucleotide that diverged from the other sequences, and therefore the original sequence was not apparent. In these five positions, sequences from Fam3 Ch2 and Fam3 Ch3 were identical, as were sequences from Fam3 Mo, Fam3 Ch1 and Fam3 Ch4. The probable original sequence in four of these positions was deduced from the sequence in the HBeAg-positive carrier (Fam3 Ch2), in agreement with most published genotype B sequences (92% agreement at nt 167, 2304 and 2525, 100% at nt 1128), and was considered uncertain in one position (33% agreement at nt 2712).

Mutations were not detected in the HBeAg-positive subject (Fam3 Ch2), but were detected in all four HBeAg-negative subjects, with a total number of mutations ranging from 15 to 27, and a similar frequency in the various genomic regions (Table 1; Figs 2 and 3).

The mother had five silent and 22 missense mutations, including an AGG → TGA mutation at nt 1762–1764 and a G → A mutation at nt 1899 in the last pre-core codon, but did not present the pre-core TAG stop mutation (A^{1896}). Three HBeAg-negative children (Fam3 Ch1, Fam3 Ch3 and Fam3 Ch4) had a stop mutation at nt 1896 (in the pre-core region), in one child (Fam3 Ch1) as a mixture (approximately 50:50) with a wild-type strain.

Mutations in relation to clinical stage and disease

The distribution of mutations in the genome is indicated in Fig. 2 and the resulting amino acid substitutions are described in Fig. 3 and Table 3. The eight HBeAg-positive carriers, with a mean observation period of 24·4 years, showed a mean of 1·6 mutations (0·05 %) in the whole genome, compared to a mean of 20 mutations (0·62%) in five HBeAg-negative carriers with a mean observation period of 27·6 years ($P < 0·0001$; Fisher's exact test). Silent mutations were rare and the majority (88%) of the mutations resulted in amino acid substitutions in at least one and in some cases (13%) in two (i.e. overlapping) ORFs.

The individual sequences were also compared with the consensus sequences representing the corresponding genotypes (as obtained from NCBI database sequences), giving nucleotide differences ranging from 1·6 to 2·4%. A large proportion of these differences, especially in HBeAg-positive

Table 3. Amino acid substitutions deduced from sequencing of the entire viral genome from 13 HBV carriers

HAI_{sum} is the sum of HAI scores for inflammation and fibrosis. NA, Not available.

Clinical stage	Patient	HAI_{sum}	Amino acid substitutions						
			Pre-S (174/163 aa*)	HBsAg (226 aa)	X protein (154 aa)	Pre-core (29 aa)	Core (183 aa)	Polymerase (845/834 aa†)	Conservative mutations (nt)
Asymptomatic or mild inflammation, $HBeAg^+$	Fam1 Ch1	2							
	Fam1 Ch2	4						$I^{156} \rightarrow V$	
	Fam2 Mo	NA							
	Fam2 Ch1	1		$M^{213} \rightarrow I$	$V^{127} \rightarrow I$			$A^{568} \rightarrow T$	
	Fam2 Ch2	4							
	Fam2 Ch3	3			$I^{37} \rightarrow V$				
	Fam3 Ch2	2							
Chronic active hepatitis, $HBeAg^+$	Fam1 Ch3	9		$T^{27} \rightarrow K$, $S^{204} \rightarrow R$		$W^{28} \rightarrow$ Stop	$T^{147} \rightarrow C$	$L^{12} \rightarrow V$, $H^{370} \rightarrow Q$, $S^{548} \rightarrow T$, $L^{709} \rightarrow R$	1655, 1757, 2299
Active cirrhosis, $HBeAg^-$	Fam1 Mo	13	$P^{160} \rightarrow L$	$C^{76} \rightarrow Y$	$R^{26} \rightarrow C$, $I^{127} \rightarrow N$, $K^{130} \rightarrow M$, $V^{131} \rightarrow I$	$W^{28} \rightarrow$ Stop	$I^{80} \rightarrow T$, $P^{135} \rightarrow T$, $T^{147} \rightarrow C$, $V^{149} \rightarrow I$	$L^{12} \rightarrow V$, $R^{208} \rightarrow S$, $S^{285} \rightarrow P$, $L^{426} \rightarrow I$, $Y^{592} \rightarrow F$, $M^{644} \rightarrow K$, $L^{709} \rightarrow R$	2017
	Fam3 Mo	13	$D^{54} \rightarrow N$, $S^{109} \rightarrow T$	$A^5 \rightarrow S$, $N^{40} \rightarrow S$, $T^{126} \rightarrow T/A$, $A^{159} \rightarrow A/V$	$I^{44} \rightarrow L$, $H^{86} \rightarrow P$, $G^{87} \rightarrow R$, $K^{130} \rightarrow M$, $V^{131} \rightarrow I$	$G^{29} \rightarrow D$	$T^{114} \rightarrow V$, $P^{135} \rightarrow Q$	$K^{73} \rightarrow N$, $Y^{136} \rightarrow H$, $T^{150} \rightarrow I$, $G^{234} \rightarrow E$, $P^{251} \rightarrow T$, $V^{289} \rightarrow D$, $F^{307} \rightarrow V$, $R^{359} \rightarrow L$, $N^{480} \rightarrow G$, $K^{679} \rightarrow N$	1020, 1032, 1120, 2221, 2233
Low-active healthy carrier, $HBeAg^-$	Fam3 Ch1	3	$Q^{121} \rightarrow K$	$A^5 \rightarrow S$, $T^{126} \rightarrow A$, $L^{175} \rightarrow S$	$V^{127} \rightarrow V/F$	$W^{28} \rightarrow$ Stop	$P^{135} \rightarrow Q$, $Q^{182} \rightarrow K$	$K^{10} \rightarrow R$, $K^{73} \rightarrow N$, $Y^{136} \rightarrow H$, $A^{301} \rightarrow E$, $R^{359} \rightarrow L$, $L^{437} \rightarrow I$, $N^{480} \rightarrow S$, $K^{679} \rightarrow N$	1032, 2188
	Fam3 Ch3	NA	$D^{39} \rightarrow N$, $L^{45} \rightarrow F$, $K^{168} \rightarrow T$	$G^7 \rightarrow R$, $I^{110} \rightarrow L$	$V^{127} \rightarrow I$	$W^{28} \rightarrow$ Stop	$P^5 \rightarrow T$, $Y^{38} \rightarrow H$, $L^{60} \rightarrow V$, $Q^{177} \rightarrow K$	$K^{73} \rightarrow Q$, $K^{93} \rightarrow E$, $N^{118} \rightarrow K$, $R^{219} \rightarrow K$, $P^{225} \rightarrow L$, $E^{347} \rightarrow D$, $R^{361} \rightarrow K$, $N^{464} \rightarrow T$, $N^{480} \rightarrow D$, $K^{679} \rightarrow Q$	2233, 2534
	Fam3 Ch4	NA	$S^{17} \rightarrow A$, $N^{98} \rightarrow S$, $I^{164} \rightarrow T$	$A^5 \rightarrow S$, $G^{44} \rightarrow Q$, $P^{62} \rightarrow L$, $Y^{100} \rightarrow C$, $Y^{161} \rightarrow F$, $Y^{206} \rightarrow C$	$G^{87} \rightarrow W$, $V^{127} \rightarrow F$	$W^{28} \rightarrow$ Stop	$P^{135} \rightarrow Q$, $S^{181} \rightarrow P$	$K^{73} \rightarrow N$, $K^{93} \rightarrow Q$, $Y^{136} \rightarrow H$, $F^{197} \rightarrow C$, $R^{359} \rightarrow L$, $G^{398} \rightarrow A$, $K^{679} \rightarrow N$	1120

* Deletion of aa 0–11 of pre-S in genotype D.

† Deletion of aa 183–193 of polymerase in genotype D.

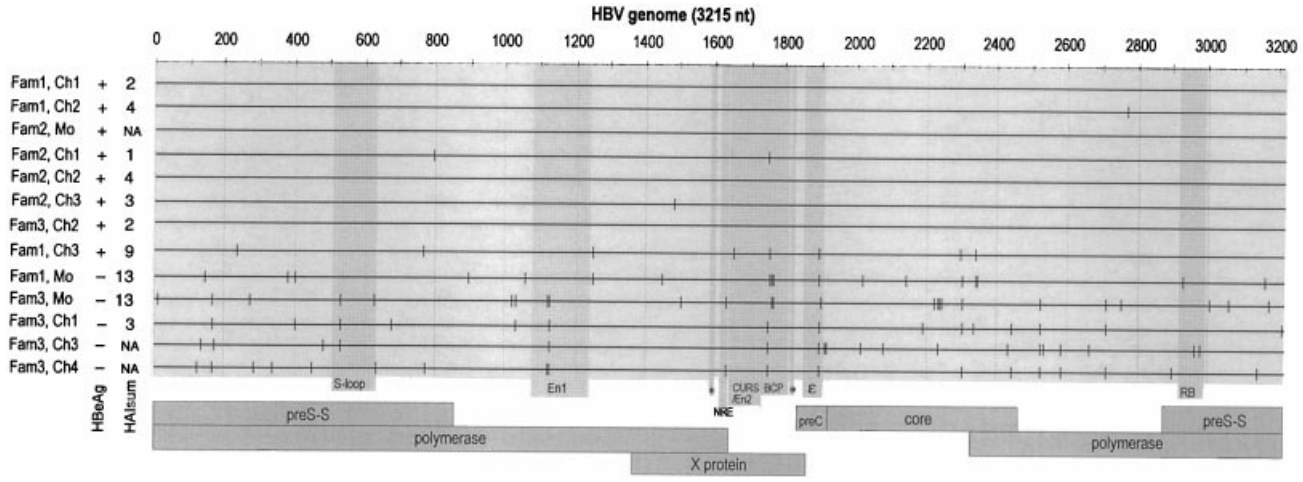

Fig. 2. Distribution of mutations found in the HBV genomes from 13 chronic carriers. The ORFs and some important genomic regions are indicated. HAIsum, Sum of HAI scores for inflammation and fibrosis; NA, not available. S-loop, predicted double loop of HBsAg containing the 'a' determinant; En1, enhancer 1; En2, enhancer 2; CURS, core promoter upstream regulatory sequence; BCP, basal core promoter; RB, putative receptor binding region in pre-S part of large S protein; ε, encapsidation signal; *, direct repeats 1 and 2; preC, pre-core.

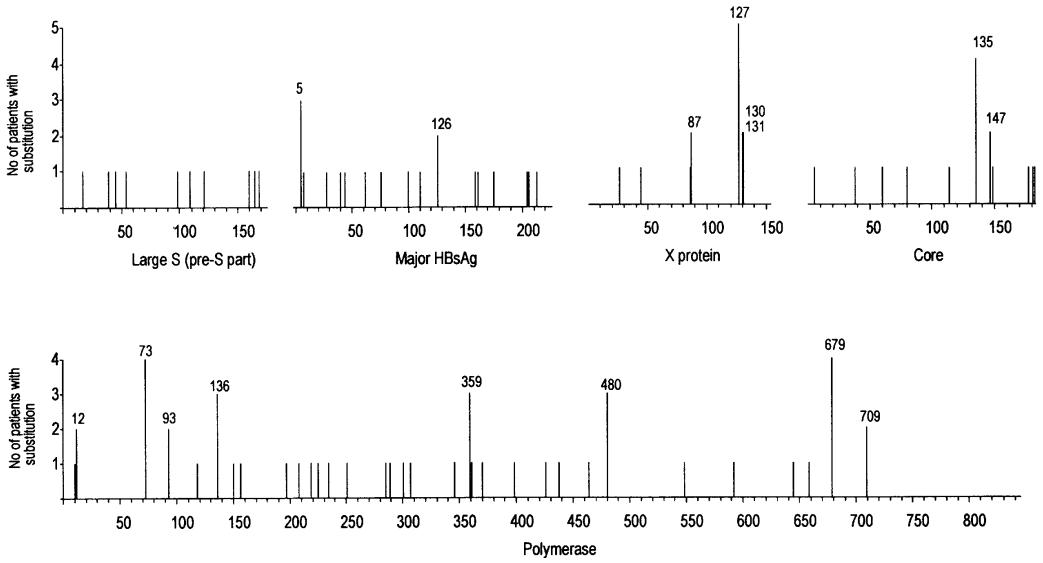

Fig. 3. Amino acid substitutions in the HBV proteins predicted from sequencing of the entire viral genome from all 13 study subjects.

carriers, appeared not to represent mutations that had emerged in the patient during the course of infection.

The total number of mutations ($n = 9$) was higher in the HBeAg-positive patient with chronic active hepatitis compared with the seven HBeAg-positive carriers with mild or minimal inflammation (0–2 mutations, mean = 0·63). In the five HBeAg-negative patients, the total number of mutations was not significantly higher in the two patients with severe liver damage (19 and 27 mutations) compared with the three patients with mild liver disease (15, 19 and 20 mutations). Moreover, the distribution of mutations between the different genomic regions and ORFs seemed not to be associated with the degree of liver damage.

Four of the five HBeAg-negative patients had a G → A mutation at nt 1896, creating a premature stop codon. Only one of these, Fam1 Mo, had severe liver damage. She also had a high level of viraemia ($10^{8·2}$ copies/ml). One further HBeAg-negative patient (Fam3 Mo) had severe liver damage, but showed a pre-core wild-type sequence and moderate viraemia ($10^{5·5}$ copies/ml).

In the pre-S region, no mutations were found in any of the eight HBeAg-positive carriers, while each of the five HBeAg-

negative had at least three mutations. One patient with mild liver damage (Fam3 Ch3) showed two mutations ($D^{39} \rightarrow N$, $L^{45} \rightarrow F$) within the putative receptor-binding region (aa 21–47; Neurath *et al.*, 1986).

In the S region, the four HBeAg-negative patients in family 3 showed mutations at sites linked to *w*/*r* seroreactivity (Okamoto *et al.*, 1986); in two of them within ($T^{126} \rightarrow A$), and in two outside ($I^{110} \rightarrow L$ and $Y^{161} \rightarrow F$), the *a*-determinant loop. Fam3 Mo showed an $N \rightarrow S$ mutation at aa 40, which was recently reported as being frequent in patients with hepatocellular carcinoma and *adw* subtype infection (Tai *et al.*, 1997).

In the X region, missense mutations were found at positions previously described as mutation sites (Uchida *et al.*, 1997), including the $K^{130} \rightarrow M/V^{131} \rightarrow I$ double mutation. This AGG → TGA mutation at nt 1762–1764 in the part of the X ORF that contains the core promoter was observed in both of the HBeAg-negative patients with severe liver damage (active cirrhosis) but in none of the other patients.

There was no apparent association between mutations in the core region and clinical status. The two HBeAg-negative patients with active cirrhosis had six and five mutations, while the three with mild liver damage had six, four and two mutations. A tendency towards clustering in the region from aa 135 to 149 was seen, including four patients with mutations at aa 135. Mutations were relatively more frequent in the putative CTL epitope at aa 141–151 ($P = 0{\cdot}04$, Fisher's exact test) and tended to be more frequent in the B-cell epitope at aa 120–140 ($P = 0{\cdot}06$) than elsewhere in the core. $T^{147} \rightarrow C$ within the CTL epitope was seen only in the two mothers with active cirrhosis.

Mutations were found in all parts of the *pol* ORF, but were rarer carboxy-terminal of aa 490. Substitutions were seen in three or four of five subjects in family 3 at aa 73, 136, 359, 480 and 679.

Discussion

The findings of the present study indicate that the HBV genome is extremely stable unless exposed to host immune responses, as exemplified by the completely conserved nucleotide sequences over a 20–35 year period in HBeAg-positive asymptomatic carriers with very high levels of virus replication. In contrast, mutations were seen in all HBeAg-negative carriers (mean of 20 mutations, 0·62%) and were distributed over all regions of the viral genome.

The stage-dependent mutation rate in chronic hepatitis B, with few mutations in the HBeAg-positive tolerance phase, has been demonstrated in several previous cross-sectional studies (Akarca & Lok, 1995; Karasawa *et al.*, 1997). Longitudinal analyses of the entire genome are lacking, probably because long observation periods are rare. The retrospective approach of the present study provides observation periods equal to the age of the offspring and should be a useful complement for studies of long-term mutation rates in chronic hepatitis B. A similar intrafamilial comparison of the genome has been used in one previous study of mutations in the core region (Bozkaya *et al.*, 1997).

In cross-sectional studies, mutations are sometimes identified by comparing the observed sequence to database sequences of the same genotype. The present study describes the degree of overestimation if this method is used for assessing mutations in individual patients. For example, in four HBeAg-positive carriers with completely unchanged genomes over 20–35 years, a database comparison would indicate 50–59 mutations. Still, comparison with carefully selected reference sequences in cross-sectional studies appears to identify the same mutation hot-spot positions as longitudinal studies, at least in the core region, as was shown in a recent analysis of compiled data (Günther *et al.*, 1999). This indicates that mutations are localized primarily to positions where they are not harmful, not necessarily reflecting the major target antigens or epitopes. This is supported indirectly by the recent observation that the majority of single point mutations, introduced randomly into the HBV genome by PCR, render the virus replication-incompetent (Günther *et al.*, 1998*c*).

In the present study, clustering of mutations could not be analysed adequately due to the relatively small number of patients. However, the observation that mutations were distributed over the entire genome without an apparent concentration to certain regions suggests that the various proteins may be of similar importance as immune targets. In the second half of the Pol region, where the intergenotypic variability is low and where the active sites for reverse transcriptase/polymerase and RNase functions probably reside (Radziwill *et al.*, 1990), the mutation frequency was lower, but this may simply reflect the fact that mutations in this region are likely to be deleterious to the virus and this observation cannot be interpreted as meaning that this region is less immunologically important.

Analysing the pathogenic importance of individual mutations was not the major topic of this work. However, the double mutation at nt 1762–1764 (AGG → TGA) was only seen in the two patients (Fam1 Mo and Fam3 Mo) with active cirrhosis, in agreement with previous reports associating this mutation with more severe liver damage (Takahashi *et al.*, 1995; Lindh *et al.*, 1999). The emergence of this mutation may be due to its location in the core promoter and its effects on transcription of pre-core/core mRNA and/or virus replication (Moriyama *et al.*, 1996; Buckwold *et al.*, 1997; Günther *et al.*, 1998*b*), but also to the $K^{130} \rightarrow M$ and $V^{131} \rightarrow I$ substitutions in the X protein (Li *et al.*, 1999).

Mutations were also seen in other parts of the X region at mutation sites described previously (Uchida *et al.*, 1997). These included mutations at nt 1630 and 1632 in the negative regulatory element, at nt 1655 in enhancer 2 and at nt 1752, 1753 and 1757 in the basal core promoter. More knowledge of the function of the X protein and the regulatory sequences is required in order to interpret these mutations.

A TAG stop mutation at codon 28 in the pre-core region was seen in three patients (Fam3 Ch1, Ch3 and Ch4) with low levels of viraemia and minimal or mild inflammation and in two patients (Fam1 Mo and Ch3) with high virus loads and severe inflammation of the liver. This agrees with the view that the severe liver damage seen in some patients with pre-core mutant infection is not due to the mutation per se, but to the high virus load that may occur despite the absence of the general marker for active infection, HBeAg, in the serum. One further HBeAg-negative patient (Fam3 Mo) had active cirrhosis but had a much lower level of viraemia; she had no pre-core mutation preventing HBeAg synthesis. In the core region, the number of mutations did not differ between HBeAg-negative patients with mild or severe liver disease, in agreement with previous reports (Akarca & Lok, 1995; Karasawa *et al.*, 1997). Mutations were relatively more frequent in the putative CTL epitope at aa 141–151 and in the B-cell epitope at aa 120–140. Mutations at nt 135 and 147, which have also been observed by others (Akarca & Lok, 1995; Carman *et al.*, 1997), were found in four and two patients, respectively (Table 3). Mutations in the S region have mainly been studied previously in the context of serotype characteristics (Okamoto *et al.*, 1987), vaccine escape (Carman *et al.*, 1990) and HBsAg-negative HBV infection (Zhang *et al.*, 1996). In four of five HBeAg-negative patients, we found mutations resulting in substitutions of amino acids that are linked to *w/r* sero-reactivity and genotype; in the *a*-determinant loop in Fam3 Mo and Fam3 Ch1 ($T^{126} \rightarrow A$) and in flanking regions in Fam3 Ch3 ($I^{110} \rightarrow L$) and Fam3 Ch4 ($Y^{161} \rightarrow F$). Furthermore, an $N^{40} \rightarrow S$ mutation, which has been reported as frequent in Japanese *adw* carriers with hepatocellular carcinoma (Tai *et al.*, 1997), was seen in one patient (Fam3 Mo) with active cirrhosis who later developed hepatocellular carcinoma.

To our knowledge, no previous study has analysed natural mutations longitudinally in the entire HBV genome. In a basically cross-sectional study of six patients (including follow-up samples from two patients after 5 months and after 3 years), Horikita *et al.* (1994) found a similar number and distribution of mutations in various parts of the genome in HBeAg-negative patients as in the present study. In a longitudinal study of a patient treated with interferon, Günther *et al.* (1998*a*) found a predominance of mutations in the core region versus other parts of the genome.

In summary, we present the retrospective analysis of the genomic divergence of HBV within families as a feasible means of studying mutations longitudinally in chronic hepatitis B and demonstrate that the entire HBV genome is extremely stable in the early high-replicative phase, when the immune response is considered to be weak. We also describe the degree of overestimation of mutation rates when assessed by comparison with database sequences and show that mutations are distributed fairly evenly in all coding regions. Further longitudinal study is needed to establish mutation rates and to identify clustering in the various genomic regions and to analyse whether natural and interferon-induced genetic variability differ in this respect.

We thank Annkatrin Gusdal for her technical expertise and Stephan Günther for critical review of the manuscript. The project was supported by grants from the Swedish Medical Research Council, the Swedish Medical Society and Göteborg Medical Society.

References

Akarca, U. S. & Lok, A. S. (1995). Naturally occurring hepatitis B virus core gene mutations. *Hepatology* **22**, 50–60.

Bozkaya, H., Akarca, U. S., Ayola, B. & Lok, A. S. (1997). High degree of conservation in the hepatitis B virus core gene during the immune tolerant phase in perinatally acquired chronic hepatitis B virus infection. *Journal of Hepatology* **26**, 508–516.

Buckwold, V. E., Xu, Z., Yen, T. S. B. & Ou, J.-h. (1997). Effects of a frequent double-nucleotide basal core promoter mutation and its putative single-nucleotide precursor mutations on hepatitis B virus gene expression and replication. *Journal of General Virology* **78**, 2055–2065.

Carman, W. F., Jacyna, M. R., Hadziyannis, S., Karayiannis, P., McGarvey, M. J., Makris, A. & Thomas, H. C. (1989). Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* ii, 588–591.

Carman, W. F., Zanetti, A. R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E., Zuckerman, A. J. & Thomas, H. C. (1990). Vaccine-induced escape mutant of hepatitis B virus. *Lancet* **336**, 325–329.

Carman, W. F., Boner, W., Fattovich, G., Colman, K., Dornan, E. S., Thursz, M. & Hadziyannis, S. (1997). Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive disease and in T helper cell epitopes during clinical remission. *Journal of Infectious Diseases* **175**, 1093–1100.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA, USA.

Günther, S., Paulij, W., Meisel, H. & Will, H. (1998*a*). Analysis of hepatitis B virus populations in an interferon-α-treated patient reveals predominant mutations in the C-gene and changing e-antigenicity. *Virology* **244**, 146–160.

Günther, S., Piwon, N. & Will, H. (1998*b*). Wild-type levels of pregenomic RNA and replication but reduced pre-C RNA and e-antigen synthesis of hepatitis B virus with C(1653) → T, A(1762) → T and G(1764) → A mutations in the core promoter. *Journal of General Virology* **79**, 375–380.

Günther, S., Sommer, G., Von Breunig, F., Iwanska, A., Kalinina, T., Sterneck, M. & Will, H. (1998*c*). Amplification of full-length hepatitis B virus genomes from samples from patients with low levels of viremia: frequency and functional consequences of PCR-introduced mutations. *Journal of Clinical Microbiology* **36**, 531–538.

Günther, S., Fischer, L., Pult, I., Sterneck, M. & Will, H. (1999). Naturally occurring variants of hepatitis B virus. *Advances in Virus Research* **52**, 25–137.

Horikita, M., Itoh, S., Yamamoto, K., Shibayama, T., Tsuda, F. & Okamoto, H. (1994). Differences in the entire nucleotide sequence between hepatitis B virus genomes from carriers positive for antibody to hepatitis B e antigen with and without active disease. *Journal of Medical Virology* **44**, 96–103.

Karasawa, T., Shirasawa, T., Okawa, Y., Kuramoto, A., Shimada, N., Aizawa, Y., Zeniya, M. & Toda, G. (1997). Association between

frequency of amino acid changes in core region of hepatitis B virus (HBV) and the presence of precore mutation in Japanese HBV carriers. *Journal of Gastroenterology* **32**, 611–622.

Knodell, R. G., Ishak, K. G., Black, W. C., Chen, T. S., Craig, R., Kaplowitz, N., Kiernan, T. W. & Wollman, J. (1981). Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology* **1**, 431–435.

Li, J., Buckwold, V. E., Hon, M. W. & Ou, J. H. (1999). Mechanism of suppression of hepatitis B virus precore RNA transcription by a frequent double mutation. *Journal of Virology* **73**, 1239–1244.

Lindh, M., Horal, P., Dhillon, A. P., Furuta, Y. & Norkrans, G. (1996). Hepatitis B virus carriers without precore mutations in hepatitis B e antigen-negative stage show more severe liver damage. *Hepatology* **24**, 494–501.

Lindh, M., Andersson, A. S. & Gusdal, A. (1997). Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus – large-scale analysis using a new genotyping method. *Journal of Infectious Diseases* **175**, 1285–1293.

Lindh, M., Hannoun, C., Dhillon, A. P., Norkrans, G. & Horal, P. (1999). Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. *Journal of Infectious Diseases* **179**, 775–782.

Lok, A. S., Akarca, U. & Greene, S. (1994). Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *Proceedings of the National Academy of Sciences, USA* **91**, 4077–4081.

Moriyama, K., Okamoto, H., Tsuda, F. & Mayumi, M. (1996). Reduced precore transcription and enhanced core–pregenome transcription of hepatitis B virus DNA after replacement of the precore–core promoter with sequences associated with e antigen-seronegative persistent infections. *Virology* **226**, 269–280.

Neurath, A. R., Kent, S. B., 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 subtypes. *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.

Okamoto, H., Tsuda, F., Akahane, Y., Sugai, Y., Yoshiba, M., Moriyama, K., Tanaka, T., Miyakawa, Y. & Mayumi, M. (1994). Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. *Journal of Virology* **68**, 8102–8110.

Radziwill, G., Tucker, W. & Schaller, H. (1990). Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. *Journal of Virology* **64**, 613–620.

Tai, P. C., Banik, D., Lin, G. I., Pai, S., Pai, K., Lin, M. H., Yuoh, G., Che, S., Hsu, S. H., Chen, T. C., Kuo, T. T., Lee, C. S., Yang, C. S. & Shih, C. (1997). Novel and frequent mutations of hepatitis B virus coincide with a major histocompatibility complex class I-restricted T-cell epitope of the surface antigen. *Journal of Virology* **71**, 4852–4856.

Takahashi, K., Aoyama, K., Ohno, N., Iwata, K., Akahane, Y., Baba, K., Yoshizawa, H. & Mishiro, S. (1995). The precore/core promoter mutant ($T^{1762}A^{1764}$) of hepatitis B virus: clinical significance and an easy method for detection. *Journal of General Virology* **76**, 3159–3164.

Uchida, T., Saitoh, T. & Shinzawa, H. (1997). Mutations of the X region of hepatitis B virus and their clinical implications. *Pathology International* **47**, 183–193.

Zhang, Y. Y., Nordenfelt, E. & Hansson, B. G. (1996). Increasing heterogeneity of the 'a' determinant of HBsAg found in the presumed late phase of chronic hepatitis B virus infection. *Scandinavian Journal of Infectious Diseases* **28**, 9–15.

Received 26 July 1999; Accepted 21 September 1999