Genomic variability in the preS1 region and determination of routes of transmission of hepatitis B virus

Angela Uy,1 Gerhard Wunderlich,1 David B. Olsen,2† Klaus-Hinrich Heermann,1 Wolfram H. Gerlich1*; and Reiner Thomssen1

1Department of Medical Microbiology, Georg August University, Göttingen and 2Max Planck Institute of Experimental Medicine, Department of Chemistry, Göttingen, Germany

On the basis of published sequence data the preS1 attachment region of hepatitis B virus (HBV) appears to be highly variable. Using a novel method for rapid DNA sequencing by the polymerase chain reaction we screened 34 HBV DNA-positive sera for mutations in a variable part of the preS1 region of the HBV genome. The sequence data were used to analyse potential chains of infection, and strongly supported the expected routes of HBV transmission among patient groups. Furthermore, sequence comparisons permitted sub-genotyping of the viruses. In the 22 cases of subtype adw, we found a very low number of point mutations. This shows that the attachment site of HBV is more highly conserved than that of other blood-transmissible viruses such as human immunodeficiency virus or hepatitis C virus.

Variability of surface epitopes seems to be an important mechanism used by persistent viruses to escape immune surveillance. The small surface protein of hepatitis B virus (HBV) (SHBs) occurs in several allelic forms, d or y and w or r being the best known. It has been shown that the subtype of SHBs remains constant throughout an 'infection chain', and analysis of these determinants may help to elucidate transmission routes in presumably related cases. However, owing to the small number of serological SHBs subtypes, many unrelated cases of HBV infection show an identical subtype. The most accurate and sensitive marker for the identity of, or differences between, two or more HBV isolates is the DNA sequence of the genome itself. With the advent of the polymerase chain reaction (PCR; Saiki et al., 1988) for the amplification and rapid sequencing of minute DNA samples, it has become possible to investigate the suitability of this analytical approach.

We selected 22 sera from 20 patients with HBV of SHBs subtype adw, for whom seven different potential chains of infection had been identified by the use of anamnestic data, and determined the DNA sequence in the preS1 region of the HBV genome. This gene region was selected because it is highly variable between HBV isolates encoding different SHBs subtypes. The preS1 domain is the N-terminal part of the largest surface protein of HBV (LHBs; for a review see Gerlich & Heermann, 1991). The amplified gene region partially overlapped with a site of attachment of HBV to hepatocyte membranes (Neurath et al., 1986; Pontisso et al., 1989). Comparison of our sequence data with those published and those from 10 further unrelated HBV DNA isolates showed a surprisingly high degree of sequence conservation within a given HBV subtype.

Hepatitis B surface antigen (HBsAg)-positive sera from individuals with acute or chronic HBV infection were collected by the diagnostic laboratory of the Department of Medical Microbiology, Göttingen, Germany. The sera contained between 105 and 4 × 108 genome equivalents/ml of HBV DNA as determined by dot-blot hybridization (Zyzik et al., 1986). The HBsAg subtype of the sera examined (allele d or y) was determined by immunodiffusion using a monoclonal antibody against d (C14/02), or an adsorbed guinea-pig antiserum against ay, both made in our laboratory. Twenty-two sera containing HBsAg subtype ad were obtained from individuals who were probably infected from one common source or had infected each other, e.g. homosexual or heterosexual couples, or patients on a dialysis ward. Twelve further sera had various subtypes; the sources of infection in these cases were most likely not related to each other, especially as many of these serum donors came from foreign countries, e.g. Vietnam and India. Sera were stored at -20 °C for as long as several years.
Isolation of HBV DNA from serum was performed as described by Kaneko et al. (1989). Briefly, a 200 µl sample of serum was digested in 500 µl of a solution containing 10 mM-Tris–HCl pH 8, 5 mM-EDTA pH 8, 100 mM-NaCl and 2.5 µg/mL proteinase K (from *Trichrachiurn album*, Sigma) for 2 h at 56 °C. After two phenol–chloroform and one chloroform–isoamyl alcohol extractions, HBV DNA was ethanol-precipitated (1 min in liquid nitrogen, followed by centrifugation for 30 min, 0 °C, 12000 r.p.m. in a bench top centrifuge) and washed with ethanol. The DNA pellet was recovered in 100 µl of double-distilled sterile water.

DNA amplification was done in a 25 µl or 50 µl final volume containing 10 mM-Tris–HCl pH 8.5, 2.5 mM-MgCl₂, 0.025% NP40, 0.025% Tween 20, 250 µM-dNTPs (with 25% or 50% of one α-thio-dNTP and three ‘normal’ dNTPs, e.g. concentrations in the 50% d-α-thio-ATP mix were 125 µM-d-α-thio-ATP, 125 µM-dATP and 250 µM each of dCTP, dGTP and dTTP). Concentrations of labelled and unlabelled primers were 0.01 A₂₆₀ units/ml (approx. 50 pm), and 30 units/ml *Taq* polymerase (Perkin-Elmer Cetus or Amersham) was used in each cup. All samples containing primers, DNA template and reaction buffer were preheated to 95 °C for 5 min to denature the DNA template and inactivate possible DNases or proteases. After addition of dNTPs and *Taq* polymerase, the samples were overlaid with 50 µl paraffin oil to prevent evaporation. The following programme was used for amplification: 40 cycles of 72 °C for 1 min polymerization, 96 °C for 15 s denaturation, and 55 °C for 15 s annealing. All experimental steps included positive controls containing plasmid pHBV991 DNA containing the complete HBV genome (Schüler, 1990; sequence available from EMBL, accession no. X51970) and negative serum samples.

Sequencing was performed by the method of Olsen & Eckstein (1989). A set of DNA fragments with non-specific base termini is generated by incomplete primer extension reactions during PCR and incorporation of one α-thio-dNTP/base. Without further purification, subsequent exonuclease III digestion of the amplified product yields DNA fragments specifically shortened by one base because incorporated thionucleotides are hydrolysed 100-fold more slowly than the normal nucleotides. The trimmed fragments can be detected by radioactive 5’-end labelling of one PCR primer. The advantage of this method is that fewer pipetting and preparative steps than in conventional dideoxynucleotide sequencing are required.

For sequencing, 5 µl ³²P-labelled PCR product from each of the four base-specific incorporation reactions was used. These samples were digested without further purification with 50 units of exonuclease III (New England Biolabs) for 30 min at 16 °C. The reaction was terminated by adding 2 µl of stop mix containing 96% deionized formamide, 10 mM-EDTA pH 8, and 0.5% bromophenol blue and 0.5% xylene cyanol FF. Samples were incubated for 5 min at 96 °C and stored on ice. Analysis was performed using a TBE buffer gradient sequencing gel prepared as described by Biggin et al. (1983). Gels were 43 x 21 x 0.04 cm in size and were run at 40 W (approx. 1500 V) for 2.5 h, vacuum-dried for 2 h without previous washing and exposed on Kodak XAR films. All DNA samples were sequenced on both strands to compensate for reading gaps and to confirm the sequencing results.

Direct sequencing of PCR amplification products by the α-thio-dNTP method requires the amplified product to be no longer than 300 bp. Although we obtained some readable sequencing results from fragments of up to 2000 bp by this method, the most convenient length is 200 to 300 bp. To obtain readable sequences without prior purification of the amplification product, annealing of the primer has to be very accurate and requires stringent conditions, especially for the sequence-generating labelled primer. Accordingly, primers of 24 bases from highly conserved regions were used. On the other hand, the amplification product should contain a highly variable region to enable detection of subtypic differences in the DNA sequence. A highly variable region of 120 bp flanked by two conserved stretches for priming was found in the preS1 region of the HBV genome. The sequence of the most abundant genotype in our serum bank was homologous to that of subtype adw (Ono et al., 1983). The DNA of this subtype showed 26 base changes compared to the sequence published by Okamoto et al. (1988) for SHBs subtype adw³ (see Fig. 1). The oligonucleotide primers had the sequence 5’ ATATCA-TGGAGATTTTCAACC 3’ (positions 2971 to 2992, numbering from the *EcoRI* site of the HBV genome), and 5’ CGGATTGGTGGAGGAGGA 3’ (positions 3150 to 3129) or 5’ CGGATTGGTGGAGGAGGA 3’ (positions 3149 to 3130). The latter was used to amplify the ay subtype HBV strains because the other antisense primer has a 3’-terminal mismatch with several of these strains. Using these primers, nearly all 34 sera giving a positive result in the dot-blot hybridization assay yielded an amplification product of the expected size as detected by ethidium bromide staining after gel electrophoresis (data not shown). In 32 cases a readable sequencing autoradiograph was obtained; two samples did not contain enough DNA to give sufficiently clear bands.

The samples analysed can be grouped according to their epidemiological relatedness, as shown in Fig. 2. One group of nine sera consisted of three samples from an HBsAg-positive surgeon collected over a period of 4 years, and one serum from each of six persons who
Fig. 1. (a) Amino acid sequence alignment of the preS1 region of published HBV strains and patients examined. The sequences of samples 1, 4, 5, 7, 8, 9, 10, 11, 12, 14, 21, 23, 28, 31, 32, 33, 34, 35, 38 and 39 are not shown because they were the same as the published sequence (adw) (Ono et al., 1983) and identical to the sequence of sample X. A hyphen indicates an amino acid residue identical to that in the sequence of sample X; # indicates codon change that does not change the amino acid residue; ? indicates that the amino acid is not known owing to ambiguous sequencing gel results. The published sequences were obtained from the EMBL databank. adw, adr2 (Ono et al., 1983); adw1, adw2, adw3 (Okamoto et al., 1988); adr1289 (Rho et al., 1989); adr1288 (Fujiyama et al., 1988); adr4, EMBL X01587; pHBV991L, EMBL X51970 (Schüler, 1990); ayr (Okamoto et al., 1986); aywl (Bichko et al., 1985); ayw6 (Galibert et al., 1979); adw2val (Valenzuela et al., 1981); chimp (Vaudin et al., 1988). (b) Phylogenetic tree of different HBV strains based on the preS1 DNA sequence. Phylogenetic grouping was performed by calculating numbers of mutations. Alignment of sequences was performed using the UNIX programs GAP and ALIGN (Devereux et al., 1984). For the comparison of the sequences the PC program MicroGenie (Queen & Korn, 1984; Beckman) was used. For the phylogenetic tree construction, the average distance method introduced by Nei (1987) was used. Gaps in the experimentally determined sequences were ignored in the calculation of the number of mutations, e.g., if only 123 of 126 bp were known, the missing three bases were counted neither as mismatches nor as matching bases. n = 19 indicates that 19 samples were identical to the adw sequence.

became HBsAg-positive several months after they had been operated on by this surgeon. These six persons did not report any other significant risk for HBV infection. The time course of seroconversion to HBsAg suggested strongly that transmission occurred from the surgeon to the patient. The blood of the surgeon contained on average $2 \times 10^8$ HBV DNA molecules/ml during the observation period. Many surgeons occasionally experience minor injuries of the hands while operating and thus, infectious blood may have entered the patient's wound.

Three patients from one dialysis ward acquired HBV infection within several months. No new HBV infection had been observed in this ward for several years, but before the cases occurred a new HBsAg-positive patient had been inadvertently treated in the 'clean' division of the ward. Five HBV DNA-positive couples formed five groups in which one partner was presumed to be the infectious source.

All 22 sera had the serological SHBs subtype ad. As controls, 10 sera were included for which no data on the source of infection were available, but the different ethnic and geographical origins of the HBV carriers...
suggested that these and the sera under examination were unrelated.

Our study shows that all samples within one chain of infection have an identical sequence (see Fig. 2). Notably, the three samples collected over a period of 4 years from the surgeon (no. 8, 9 and 10) were identical. In four of the seven chains of infection an identical HBV sequence was found in each patient. This sequence has been reported by Ono et al. (1983) for the HBs serotype adw, and therefore our study shows that this serotype is a prototype in northern Germany. In the dialysis group (no. 16, 17 and 21), an exchange at base 3013 was observed. The DNA sequence determined from one couple (no. 2 and 40) showed two exchanges; one of these resulted in a mutation changing Thr 76 to Lys. The sequence determined from the DNA from another couple (no. 29 and 37) showed one nucleotide exchange at base 3051 which did not result in an amino acid exchange. The high degree of identity of the DNA sequences within HBs serotype adw was not caused by contamination with an HBV isolate in our laboratory because every HBV DNA isolation and amplification step was checked by using control reactions. Moreover, the HBV isolate used mainly in our laboratory (encoded by pHBV991L) has a sequence different to that predominantly identified in this study.

The 10 unrelated serum samples included three ayw isolates (no. 6, 13 and 18), the sequence of which clearly correlated with those published by Bichko et al. (1985) (aywl) and Galibert et al. (1979) (ayw6). HBV DNA isolate no. 6 had the same sequence as that published by Galibert et al. (1979). An HBs ad isolate (no. 24) from an East Asian carrier had a sequence very similar to that published by Ono et al. (1983) for the adr subtype. HBV from another carrier, from India (no. 25), had a sequence which showed greatest similarity to that reported by Okamoto et al. (1988) for the HBs subtype adw2. Thus, each of the sequences determined were completely or almost completely identical to a previously published sequence, and can be assigned to the four HBV genotypes described by Okamoto et al. (1988; see Fig. 1b).

The high conservation of the amplified preS1 sequence in this study was surprising because the published sequences compiled in Fig. 1(a) suggested that much more variation would be observed. It was also surprising that all the samples analysed generated a DNA amplification product of the expected length. Recently, a variant with a deletion of 61 amino acids in the preS1 region has been reported by Gerken et al. (1991). However, the design of our study would not allow the identification of minor HBV variants. Nevertheless, the data suggest that a complete preS1 domain is essential for the virus life cycle. The preS1 sequence (amino acids 21 to 47) has been found to be involved in attachment of the virus particle to cells (Neurath et al., 1986; Pontisso et al., 1989) and the LHBs protein, which contains the preS1 domain, is essential for virion morphogenesis (Bruss & Ganem, 1991). High sequence variability is frequent in enveloped virus surface proteins. The most notable examples are human immunodeficiency virus (HIV) (Hahn et al., 1986) and hepatitis C virus (HCV) (Weiner et al., 1991). HBV, as another blood-transmitted virus, would also be a candidate for such high variability, particularly due to its retrovirus-like replication strategy (Girones & Miller, 1989).

The data in Fig 1(6) show that the preS1 domain is highly variable, with up to 20% sequence variability and hot spots at amino acids 43 to 45, 63, and 74 to 78. HBV is known to generate spontaneous mutations which are selected by the persistently infected host. An example is the frequent occurrence of HBe-negative preC mutants (Carman et al., 1989; Brunetto et al., 1991). However, the results of this study suggest that HBV forms, in contrast to the potential attachment sites of HIV and HCV, very stable subtypes of the surface-exposed preS1 domain; no variation was found in one carrier over 4 years of persistent infection. No more than two mutations within the 126 bp of the preS1 sequence of subtype adw, which is predominant in northern Germany, were found. These data suggest that mutation of an attachment site is not the major mechanism by which HBV escapes the immune response in a persistently infected individual. In fact, many viraemic HBV carriers have circulating immune complexes enriched in preS1 antigen, as compared to free antigen (Madalinski et al., 1991), and most likely also in anti-preS1 antibody (B. Burczynska & W. H. Gerlich, unpublished results). Anti-preS1 antibody is frequently found in patients convalescing from hepatitis B virus infection (Deepen et al., 1990). In view of the frequent mutations of the HBV genome reported by others (Carman et al., 1989, 1990; Brunetto et al., 1991; Santantonio et al., 1992), we suggest that the selection of certain preS1 sequences is favoured, irrespective of how many mutations may occur.

The low genome variability within a subtype made genome typing by sequence less informative than we had originally expected. Nonetheless, some useful observations were made. The suspected transmission of HBV from a surgeon to six of his patients was at least compatible with the finding of a constant DNA sequence. In the three dialysis cases, one mutation not found in other cases made a common source of infection very likely. Sequencing also showed that two unrelated groups of subtype ad-positive patients were infected with the same HBV genotype; however, the sequences from the members of three groups showing HBsAg subtype adw were slightly different. Thus, in the majority of these
cases the virus had not been imported from foreign regions where genotypes with the HBsAg d allele prevail. Finally, analysis of the preS1 region to examine relationships produced a phylogenetic tree similar to that produced on the basis of the entire HBV genome (Fig. 1b).

References


(Received 17 March 1992; Accepted 10 July 1992)