Naturally Occurring Hepatitis B Virus Genomes Bearing the Hallmarks of Retroviral G → A Hypermutation

Stephan Günther,* Gunhild Sommer,* Uwe Plika† Alicja Iwanska,* Simon Wain-Hobson‡ Hans Will,* and Andreas Meyerhans† 1

*Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie, D-20251 Hamburg, Germany; †Abteilung Virologie, Institut für Medizinische Mikrobiologie und Hygiene, Universität Freiburg, D-79104 Freiburg, Germany; and ‡Unité de Rétrovirologie Moléculaire, Institut Pasteur, F-75724 Paris Cedex 15, France

Received March 27, 1997; returned to author for revision May 7, 1997; accepted June 9, 1997

Two hypermutated genomes of hepatitis B virus (HBV) were cloned from sera of chronic virus carriers. Twelve percent and 26% of guanosine residues were replaced by adenosine, with the transitions being erratically distributed along the genome. G→A substitutions showed a strong dinucleotide preference, decreasing in the order GpA > GpG > GpC > GpT. Such traits are typical of retroviral G→A hypermutation which results from cDNA synthesis coinciding with fluctuations in the intracellular [dTTP]/[dCTP] ratio. The observations offer an explanation for the high prevalence of HBV variants bearing a tryptophan 28→stop codon in the pre-core region of carriers with chronic active or fulminant hepatitis. The HBV hypermutants indicate that a small proportion of hepatocytes have distorted dNTP pools, which might have implications for the fidelity of hepatocyte DNA replication or repair.

INTRODUCTION

RNA virus and retroviral mutation rates are the highest for any replicon, being on the order of 10^{-4} to 10^{-5} per base per cycle (Drake, 1993; Preston and Dougherty, 1996; Williams and Loeb, 1992). Although certainly incompatible with survival, higher rates do exist, with retroviral G→A hypermutation being a case in point (Pathak and Temin, 1990; Vartanian et al., 1991). In some cases, many hundreds of Gs may be replaced by As throughout the entire 10-kb retroviral genome, while for others it may be erratic or confined to a small isolated segment (Borman et al., 1995; Gao et al., 1992; Johnson et al., 1991; Vartanian et al., 1991; Pathak and Temin, 1990; Perry et al., 1992; Vartanian et al., 1994; Wain-Hobson et al., 1995). Locally the mutation rate can attain values as high as 0.6 per G per cycle, although the usual range is more on the order of 0.1 - 0.3 per G. Transitions are strongly coupled to the local dinucleotide context, decreasing in the order GpA > GpG > GpT > GpC. G→A hypermutation arises when reverse transcription of single-stranded genomic RNA into the complementary DNA strand occurs at the same time as fluctuations in the intracellular [dTTP]/[dCTP] ratio (Vartanian et al., 1994, 1997). Such biases are conducive to the formation of rG:dT mismatches, which are not only the most stable of mismatches but also the most readily accommodated by reverse transcriptases (Sala et al., 1995; Mendelman et al., 1990). Such large numbers of transitions may indeed be produced in a single cycle of replication in vivo and reproduced in vitro (Martinez et al., 1994; Pathak and Temin, 1990; Vartanian et al., 1997).

The extent of G→A hypermutation is most pronounced for the lentiviral subset of retroviruses which includes the human and simian immunodeficiency viruses. Of the few examples described for the oncoretroviruses, the mutation rates are an order of magnitude lower, suggesting that the lentiviral reverse transcriptase influences the mutation rate, perhaps by efficient elongation beyond repetitive rG:dT mismatches (Martinez et al., 1995; Sala et al., 1995; Mendelman et al., 1990). The hepadnaviruses and badnaviruses are distinct mammalian and plant viruses, respectively, sharing a unique trait with the classical retroviruses (Covey, 1991; Nassal and Schaller, 1993). Although virion-associated nucleic acid is DNA, a more than full-length RNA transcript (RNA pregenome) is reverse transcribed into a partially double-stranded DNA genome. Because reverse transcription in hepadnavirus and badnavirus, like that in their retroviral counterpart, is presumed not to be subject to 3′ exonucleolytic editing or proofreading, it is possible that their replication fidelity might also be sensitive to fluctuations in intracellular hepatocyte dNTP concentrations. Although there is a considerable sequence da-
tabase for human hepatitis B virus (HBV), the prototype hepadnavirus, hypermutants have not been hitherto
reported.

MATERIALS AND METHODS

HBV genomes were amplified by polymerase chain reaction (PCR) from sera of eight chronic HBV carriers. Briefly, 300 µl serum was incubated at 65°C for 4 hr in 20 mM Tris–HCl (pH 8.0), 10 mM EDTA, 0.1% SDS, and 0.8 µg/ml proteinase K. The DNA was extracted with phenol/chloroform and precipitated with ethanol using 20 µg tRNA as carrier. HBV DNA was amplified by PCR with primers 5'-CCGGAAAGCTTGGCCTTCATCCCTGACCTCTGAGCAGCCACGACCTTGAGG-3' (Fig. 1A). The overall G>A transition frequencies (fG>A) were 0.12 (55/1978 bp) for 7720 and 0.26 (96/1622 bp) for 7648 when compared to non-hypermutated sequences from the same patient (Fig. 2). The two sequences revealed multiple stop codons in most reading frames, and in the case of 7720 a deletion of a single G in a run of six Gs, with the most frequently substituted site being TGG GGC, with the most frequently substituted site being underlined. In light of the link between HBV replication fidelity and fluctuations in the intracellular [dTTP]/[dCTP] ratio, this finding may be rationalized as follows: reverse strand synthesis, both G → A and C → T transitions would have been anticipated.

The overall G → A transition frequencies (fG→A) were 0.12 (55/1978 bp) for 7720 and 0.26 (96/1622 bp) for 7648 when compared to non-hypermutated sequences from the same patient (Fig. 2). The two sequences revealed multiple stop codons in most reading frames, and in the case of 7720 a deletion of a single G in a run of six Gs (Figs. 2B and 2C). The distribution of transitions was highly erratic along the genome (Figs. 2D and 2E). In the case of 7648 the YMDD amino acid motif, typical of the catalytic site of most retroviral reverse transcriptases, was mutated to YINN. Analysis of the context surrounding G → A transitions revealed a strong dinucleotide preference, namely GpA→GpG→GpC→GpT (Fig. 3A). No significant dinucleotide context was apparent for the 5' nucleotide, i.e., NpG (Fig. 3B). Given the sense of reverse transcription of RNA, this 5' nucleotide is downstream to the site of rG:dT misincorporation.

There are two forms of the core (C) protein, resulting from two sites of translation initiation. The pre-C product encodes an additional 29 residues and is involved in the synthesis and secretion of HBeAg. The second ATG (codon 30) is used when there is a G → A transition (Trp → stop) at codon 28. Such a Trp28stop mutant is frequently found after seroconversion to anti-HBe. The second ATG (codon 30) is used when there is a G → A transition (Trp → stop) at codon 28. Such a Trp28stop mutant is frequently found after seroconversion to anti-HBe. The second ATG (codon 30) is used when there is a G → A transition (Trp → stop) at codon 28. Such a Trp28stop mutant is frequently found after seroconversion to anti-HBe ( Günther et al., 1992; Hadzynannis, 1995; Okamoto et al., 1990). Interestingly the transition occurs in a run of Gs, TGG GGC, with the most frequently substituted site being underlined. In light of the link between HBV replication fidelity and fluctuations in the intracellular [dTTP]/[dCTP] ratio, this finding may be rationalized as follows: reverse
FIG. 2. Genomic organization of two G → A hypermutated deletion mutants of HBV and distribution of the G → A transitions. (A) Hypermutated clones 7648 and 7720 are shown in comparison to a full-length HBV genome. The four open reading frames are denoted C, P, S, and X. For simplicity the pre-C, pre-S1, and pre-S2 are omitted. (B and C) The deletions of 1583 and 1223 nucleotides in clones 7648 and 7720, respectively, derive from splicing of the RNA pregenome and are marked with lines. Clone 7648 bore G → A transitions across the splice junction, indicating that the mutations occurred at the level of reverse transcription of a spliced RNA pregenome (C). Vertical bars within the four open reading frames denote stop codons. (D and E) Distribution of G → A substitutions along the hypermutated HBV genomes. Individual intrapatient consensus sequences were used as reference sequences. The %G → A/G was scored by sliding a 100-bp window at 50-bp intervals. Average substitution frequencies were 26% and 12% for clones 7648 and 7720 as noted. In addition to the 98 G → A transitions for clone 7648 there were 5 other mutations with respect to the reference sequence (3 A → G, 1 A → T, and 1 T → C). For 7720 the numbers are: 55 G → A, 1 G → T, 2 A → G, 2 A → T, 2 T → C, 1 C → T, and 1 C → A. It has previously been shown in vitro that a small number (usually ≤5%) of other substitutions may accompany G → A hypermutation (Martinez et al., 1994). The EcoRI restriction site used as reference origin is indicated by an arrowhead.
transcription of the three preceding Gs would consume dCTPs, leading to a local depletion. As G:T mismatches are the most stable of base mismatches, dTTP will be misincorporated with the greatest probability, leading to a G → A transition.

The characteristics of these hypermutated HBV subgenomes, namely the high frequency of G → A transitions, erratic distribution, and strong dinucleotide preference, are precisely the hallmarks of retroviral G → A hypermutations, particularly those of the lentiviruses (Borman et al., 1995; Vartanian et al., 1991, 1994; Wain-Hobson et al., 1995). Their occurrence suggests that strong intracellular dNTP biases exist in a small proportion of hepatocytes, just as they do for leukocytes. An impact of such biases on the fidelity of DNA replication and repair, as well as the long-term evolution of the host cell genome, is suspected, given the excess of G → A and C → T transitions among pseudogenes and cellular genes such as TP53 or those reported as the cause of inherited disease (Gojobori et al., 1982; Krawczak et al., 1992, 1995; Li et al., 1984).

ACKNOWLEDGMENTS

We thank Otto Haller for continuous interest in the work. S.G., G.S., A.J., and H.W. were supported by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (01K19560) and from the Wilhelm-Sander Stiftung. U.P. and A.M. were supported by the Deutsche Forschungsgemeinschaft. The Heinrich-Pette Institut für Experimentelle Virologie und Immunologie is supported by the Bundesministerium für Gesundheit and the Hansestadt Hamburg.

REFERENCES


fected hosts along with seroconversion to the antibody against e antigen. J. Virol. 64, 1298 - 1303.


