A novel recombinant of *Hepatitis B virus* genotypes G and C isolated from a Thai patient with hepatocellular carcinoma

Kamol Suwannakarn,1 Pisit Tangkijvanich,2 Apiradee Theamboonlers,1 Kenji Abe3 and Yong Poovorawan1

1Center of Excellence in Viral Hepatitis Research, Department of Pediatrics, Chulalongkorn Hospital, Rama IV Road, Patumwan, Bangkok 10330, Thailand
2Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand
3Department of Pathology, National Institute of Infectious Diseases, Tokyo 162-8640, Japan

Genomic recombination between different genotypes of *Hepatitis B virus* (HBV) resulting in hybrid strains has been increasingly documented. In this study, a novel recombinant of HBV genotypes G and C isolated from a Thai patient with hepatocellular carcinoma is reported. Based on phylogenetic analyses of the S, P and X genes and the entire genome, the HBV isolate clustered on a branch within genotype G, but clustered with genotype C on analysis of the C gene. Using the program SIMPLOT and bootscanning analysis, the recombination breakpoints were located at nt 1860 and 2460 of the precore/core region. The hallmarks of the original genotype G, including a 36 bp insertion in the core region and dual stop codons in the precore region, were not identified in this isolate. These data should encourage further investigations on the epidemiological and virological characteristics of HBV genotype G involved in recombination with other genotypes.

*Hepatitis B virus* (HBV) is one of the major causes of chronic liver diseases, including chronic hepatitis, cirrhosis and hepatocellular carcinoma (HCC), which affect more than 350 million people worldwide (Ganem & Prince, 2004). HBV, a member of the family *Hepadnaviridae*, is a circular double-stranded DNA virus of approximately 3200 bp that encodes four overlapping open reading frames (ORFs). Based on sequence divergence in the entire genome of > 8%, HBV is currently classified into eight genotypes, designated A to H (Weber, 2005). The genotypes of HBV have distinct geographical distributions and may influence the clinical outcomes of patients with chronic infection (Schaef er, 2005). Genotypes A and D are common in Europe and North America, whereas genotypes B and C are highly prevalent in Asia. Genotypes F and H are restricted to Central and South America. Genotype E is found predominantly in West Africa and genotype G is found in the USA and Europe.

Hybrid HBV strains resulting from genomic recombination between different genotypes have been increasingly documented (Bollyky et al., 1996; Bowyer & Sim, 2000; Fares & Holmes, 2002; Morozov et al., 2000). For instance, phylogenetic analysis revealed that hybrids between HBV genotypes B and C, which have sites of recombination over the precore/core region, were found ubiquitously in Asian countries, except for Japan (Luo et al., 2004; Sugauchi et al., 2002). Similarly, hybrids between HBV genotypes A and D have been reported in Italy and South Africa (Morozov et al., 2000; Owiredu et al., 2001) and an aberrant recombinant between genotype C and a subgroup of genotype A was isolated in Vietnam (Hannoun et al., 2000). Recently, hybrids of HBV genotypes C and D have been identified in Tibet and China (Cui et al., 2002; Wang et al., 2005).

In a previous study, the distribution of HBV genotypes in 332 Thai patients with chronic HBV infection was investigated by using PCR–restriction fragment-length polymorphism analysis (Tangkijvanich et al., 2005). Our data showed that the most common HBV genotypes in this group were C, B and A, accounting for 73, 21 and 3%, respectively. However, the genotypes of the remaining isolates could not be specified by using this technique. As a result, samples that could not be genotyped were sent for direct sequencing of the preS1 gene. Among these samples, an aberrant HBV strain belonging to genotype G was recovered from a Thai patient with HCC. Through a more extensive analysis, it was revealed that the isolate represented a novel HBV hybrid with genotype C in the precore/core region.
Here, the complete genomic sequence of this HBV isolate is reported. The isolate was from a patient (male, 47 years old) who was a resident in southern Thailand and had undergone follow-up at Chulalongkorn Memorial Hospital (Bangkok, Thailand) between February and April 1998. The patient was diagnosed with HCC by the presence of mass lesions in the liver on hepatic imaging and serum $\alpha$-fetoprotein levels above 400 ng ml$^{-1}$. The patient was seropositive for HBsAg (ELISA; Abbott Laboratories), but negative for HBeAg (ELISA; Abbott Laboratories) and anti-HCV (ELISA; Ortho Diagnostic Systems). To determine the complete nucleotide sequence of the HBV isolate, DNA was extracted from 100 $\mu$l stored serum ($-70^\circ$C) by using proteinase K/SDS in Tris buffer, followed by phenol/chloroform extraction and ethanol precipitation. DNA pellets were dissolved in 30 $\mu$l sterile water and subjected directly to PCR-based amplification. PCR was performed by using a set of primers to amplify six overlapping fragments of the HBV genome as follows. In fragment 1, sense and antisense primers were CORE1 (5'-GAGTGTGGATTCCACCTCCTCC-3'; nt 2268–2289) and R1 (5'-TGTAAAACAGGAGGGGTCCTA-3'; nt 201–180), respectively. In fragment 2, sense and antisense primers were F2 (5'-CATCCTCAGGCCATGTGGA-3'; nt 3193–3214) and R4 (5'-ATGGCACTAGAAACTGACCC-3'; nt 689–669), respectively. In fragment 3, sense and antisense primers were F4 (5'-GTCCTCTCAATTGTTGCGG-3'; nt 348–366) and R6 (5'-GGCGAGAAATGGAAGCCTG-3'; nt 1103–1084), respectively. In fragment 4, sense and antisense primers were F6 (5'-ATATGGATGATGTGATTTGGG-3'; nt 737–758) and MD26 (5'-GTTACGGTGTTGCTCCAT-3'; nt 1625–1608), respectively. In fragment 5, sense and antisense primers were Xi1 (5'-AGCTTGTGTTTGCTCGCAGC-3'; nt 1287–1305) and Xi3 (5'-GCAGCATTTGGAAGGTTG-3'; nt 1883–1865), respectively. In fragment 6, sense and antisense primers were X101 (5'-TCTGGCTCTCCTCATCTG-3'; nt 1552–1569) and CORE2 (5'-CCTTATGCTCAAGG-3'; nt 2476–2457), respectively. Amplicons were purified with a gel-extraction kit (Perfectprep Gel Cleanup; Eppendorf). DNA sequencing analysis of the PCR products was performed with a Perkin-Elmer 310 sequencer.

For phylogenetic analysis, nucleotide sequences were multiply aligned by using the program CLUSTAL_X (version 1.83). Alignments were then fed into the software program PHYLIP (version 3.5c). SEQBOOT, DNADIST and NEIGHBOR were used for bootstrapping of a 1000-replicate dataset; CONSENS was used to compute a consensus tree. The phylogenetic trees were visualized by using TREEVIEW software (version 1.6.6) (http://taxonomy.zoology.gla.ac.uk/rod/rod.html).

An entire HBV genome of 3213 nt was obtained, which was designated CU400. Analysis of the whole genome sequence after running through the BLAST program and comparing with sequences in GenBank showed that the isolate belonged to genotype G. Phylogenetic analysis of the S gene sequences showed that the isolate clustered on a genotype G branch (Fig. 1a); this was also observed on analysis of the P and X genes and the full genome sequences (data not shown). However, in trees based on the C gene, the sequence clustered with genotype C, instead of genotype G (Fig. 1b). Phylogenetic analyses of the S, P and X genes of this hybrid exhibited a close sequence similarity (96.8–97.3%) to the original isolates of genotype G, but <88.7% similarity to any isolates of genotypes A to F. Taken together, the clustering of sequences in discordant positions following phylogenetic analysis provided evidence of a novel recombinant between genotypes G and C.

To locate the breakpoints of genomic recombination more accurately, the program SIMPLOT (version 2.5) (Lole et al., 1999) and bootscanning analysis (Salminen et al., 1995) were used. The SIMPLOT program was applied to identify phylogenetically informative sites supporting alternative tree topologies. This method was performed by considering four sequences at a time: one putative recombinant sequence, two reference sequences of original G (GenBank accession no. AB056513) and C (GenBank accession no. X04615) genotypes, and one sequence of a known outgroup (genotype E; GenBank accession no. X75664). Each informative site supports one of three possible phylogenetic relationships among the four taxa. Bootscanning and cluster analysis maximizing the $\chi^2$ parameter were used to identify breakpoints in the intergenotypic recombinants; the $P$ value for the subsequent division of the sequence into genotypes was calculated by using Fisher’s exact test. Following these methods, the similarity plots of the complete sequence showed regions of higher similarity to genotype G alternating with regions of higher similarity to genotype C within the precore/core region (Fig. 2a). The recombination breakpoints were estimated at positions nt 1860 and 2460, which are located in the terminal areas of the precore and core regions, respectively (Fig. 2b).

Unlike other HBV genotypes, the authentic genotype G would not be able to encode HBeAg because of stop codons at positions 2 and 28 (Stuyver et al., 2000) in the precore region, which can abort translation of the HBeAg precursor made of 10 aa encoded by the 3'-terminal part of the precore region and 149 aa encoded by the 5'-terminal part of the C gene (Okamoto et al., 1990). When the nucleotide sequence of the new isolate was aligned, a translational codon was detected at codon 2 (C1817T) in the precore region, but the stop codon at codon 28 could not be identified. Thus, the presence of a translational stop mutation at codon 2 indicates that this strain would not express HBeAg in the serum. In contrast to the original strain of HBV genotype G (Stuyver et al., 2000), the core region did not possess an insertion of 36 bp located at the 5’ end. However, a 3 bp deletion at position 3 in the preS1 region, which is a typical characteristic of genotype G, was found. In the HBsAg region, amino acids at position 122, 127 and 160 were lysine (K), proline (P) and lysine (K), respectively. Hence, the serological subtype of this HBV strain was expected to be adw2.

The most interesting result of the present study is that a
A novel hybrid HBV strain resulting from recombination between genotypes G and C has been described at length for the first time. This mosaic strain was recognized because it produced an unusual genotyping pattern by nucleotide sequencing and phylogenetic analysis. Indeed, strong evidence suggests that genotype G is frequently co-infected with other genotypes. For instance, all of the isolates of genotype G recovered in San Francisco were co-infected with genotype A (Kato et al., 2002b). Similarly, all of the genotype G isolates from Canada were co-infected with either genotype A or genotypes A and C (Osiowy & Giles, 2003). Given evidence for the high frequency of co-infection with other genotypes, it is uncertain whether genotype G is competent to replicate by itself or mainly depends on other genotypes for replication (Kato et al., 2002a). Moreover, whether the recombinant strain of genotypes G and C would facilitate the evasion of immune surveillance and have an advantage for persistence in hosts over the authentic genotype G is unclear.

It should be mentioned that the recombination breakpoints identified in this study occurred in the vicinity of the DR1 region and encapsidation signal of the HBV pregenome. In an in vitro recombination assay, fragments containing the region spanning DR1, which is believed to be the origin of virus replication and a preferred site in the viral genome for integration, enhanced recombination in the presence of extracts from actively dividing cells (Hino et al., 1991). Of relevance to this evidence, it was demonstrated that the region encompassing nt 1600–2000 achieved a recombination-site density that was almost fivefold higher than that of the remaining part of the genome (Pineau et al., 1998). Thus, it is possible that the genomic region covering DR1 would be responsible for intergenotypic recombination of genotype G, as well as for integration between HBV and the host genome. However, additional data on the recombination of HBV genotype G are still required before any conclusions can be drawn. Furthermore, the possibility that such a recombination might contribute to the development of HCC in this case needs further investigation.

**Acknowledgements**

This study was supported by the Thailand Research Fund and Center of Excellence, Viral Hepatitis Research Unit, Chulalongkorn University and the International Joint Research Program, Research Fund Project on Health Sciences focusing on Drug Innovation of Japan.
Fig. 2. (a) Nucleotide similarity comparison of the complete genome of CU400 with those of genotype G (GenBank accession no. AB056513) and genotype C (GenBank accession no. X04615), showing a window size of 200 bp and a step size of 20 bp. (b) Bootscan plots showing the likelihood of clustering of the putative G/C recombinant sequence (CU400) with the reference sequences of genotypes G and C, as well as genotype E as the outgroup (GenBank accession no. X75664).

References


