Integration of hepatitis B virus DNA through a mutational hot spot within the cohesive region in a case of hepatocellular carcinoma

Kristina Quade, John Saldanha, Howard Thomas and John Monjardino*

Department of Medicine, Queen Elizabeth the Queen Mother Wing, St Mary's Hospital Medical School, South Wharf Street, London W2 1PG, U.K.

Integrated hepatitis B virus (HBV) DNA previously cloned from a hepatocellular carcinoma genomic library derived from a Japanese patient was characterized further. Sequence analysis of restriction fragments bearing the virus–host junctions defined 3125 nucleotides of essentially un-rearranged HBV DNA of the adr subtype with the right junction mapping within the cohesive region at position 1970 and the left within the pre-core at position 1886. The right viral–host junction contains a 7 bp repeat (TGTAGGC) and a possible 2 bp inversion. The integrated HBV DNA includes the complete open reading frames for core, pre-S, S and polymerase and a 3' truncated X gene, and lacks most of the pre-core. Integration has occurred at a mutational hot spot of the viral genome and appears to be located in a region of semi-repetitive genomic DNA 3' to the β-globin gene cluster.

In spite of strong epidemiological evidence which links chronic hepatitis B virus (HBV) infection with hepatocellular carcinoma the molecular basis for a possible role of the virus in hepatocarcinogenesis has not been established. Integrated viral DNA sequences have now been identified in carcinoma-derived genomic DNA in a significant proportion of cases but no pattern of integration has emerged and their involvement in tumour development remains unclear. In the tumours so far analysed integration appears to occur preferentially through the 5' terminal overlap and to be targeted randomly or semi-randomly throughout the human genome (Hino et al., 1989; Shih et al., 1987; Tokino et al., 1987; Yaginuma et al., 1985, 1987; Zhou et al., 1988).

In the present report we have analysed further the HBV DNA sequence which we had cloned previously and partially characterized from a hepatocellular carcinoma derived from a Japanese patient (Fowler et al., 1986).

A 10 kb HindIII fragment of tumour-derived genomic DNA containing integrated HBV DNA (L1C1) previously described (Fowler et al., 1986) was amplified and the DNA purified using conventional techniques (Maniatis et al., 1982). The host–virus junction-containing DNA fragments were prepared by restriction with appropriate enzymes (Fig. 1), and ligated to restricted plasmid pUC9 DNA. After transformation of Escherichia coli HB101, insert-containing colonies were identified, amplified and DNA was extracted for sequence analysis (Zagursky et al., 1985).

The sequencing strategy which also made use of chemically synthesized oligodeoxynucleotide primers is shown in Fig. 1. Sequencing of the left junction was initially carried out on a subcloned HindIII/BglII fragment (HB12) shown in Fig. 1 which we predicted to contain the host–viral junction from previous restriction site mapping (Fowler et al., 1986). The sequence obtained was however entirely non-viral implying that the viral DNA ended precisely at the BglII site (position 1987 on the HBV genome) or that an additional BglII/BglII fragment previously undetected contained the host–viral junction. To investigate the latter, BglII-digested fragments were cloned into the BamHI site of pUC9 and two HBV DNA-positive clones (BB2 and BB3) were isolated by colony hybridization using a full 3.2 kb HBV DNA probe. BB2 was found to be about 500

![Fig. 1. Map of partial clones from L1C1 and synthetic oligodeoxynucleotides used in sequencing virus–host junctions. Area in black corresponds to previously established minimum viral sequence from restriction analysis (Fowler et al., 1986). A, AseI; B, BamHI; Bg, BglII; E, EcoRI; X, XhoI. , Minus polarity synthetic primers, BB2 extension and BB3 extension; the latter was used for confirmation of the host flanking region. , Plus polarity primers 1585 and BE5EXT. Map positions of primers are indicated in the text.](image-url)
shown in Fig. 2. AvaI/EcoRI fragments based on previous restriction site assignments which placed the AvaI site at position 1475, from the enzymes tested, closest to the right boundary (Fig. 1). Sequencing of the BamHI/EcoRI clone (BE5) both from the pUC9 sequencing primers and from two synthetic oligonucleotides (1586 to 1609 and 1762 to 1777, plus-strand polarity) established the position of the right viral-host junction at position 1790 followed by a 7 bp repeat (TGTAGGC) and a possible 2 bp inversion (TA) as shown in Fig. 2 and 3. Further sequencing into the viral DNA from both flanking sequences excluded any additional rearrangements of the integrated viral DNA within about 200 nucleotides downstream of 1886 (left junction) or nearly 400 nucleotides upstream of 1790 (right junction) as shown in Fig. 2.

Integrated HBV DNA thus comprises 3125 bp of essentially un-rearranged DNA of the adr subtype (Fowler et al., 1986) with a 7 bp duplication and 2 bp inversion at the right junction. The sequence includes the core (but not pre-core), surface (including pre-S1 and pre-S2) and polymerase genes, and an X gene truncated by 47 nucleotides at the 3' end (Fig. 2 and 3) which would encode an X gene product 16 amino acids short at the carboxyl end, followed by an additional leucine, and terminating within the 7 bp repeat. Both the pre-core AUG and most of the pre-core coding region are absent. The enhancer sequence downstream from S and the pre-core inhibitory sequence (1987 to 2431), whereas BB3 was just over 100 bp long and was presumed to be the viral BglII/BglII sequence (1987 to 2431), whereas BB3 was just over 100 bp and was expected to contain the junction (Fig. 1). Sequencing of the two clones confirmed these predictions and placed the left junction at HBV DNA position 1886 (or 1861 of the adr subtype which contains a 27 bp deletion) within the pre-core region (Fig. 2 and 3). Further confirmation of the sequence at the virus-host boundary was also obtained by sequencing from a synthetic primer from within BB2 (2021 to 2040, minus-strand polarity) using DNA from clone 5e (HindIII left of EcoRI) as template (Fig. 1). Analysis of the right viral-host junction was carried out by subcloning both AvaI/EcoRI and BamHI/EcoRI fragments based on previous restriction site assignments which placed the AvaI site at position 1475, from the enzymes tested, closest to the right boundary (Fig. 1). Sequencing of the BamHI/EcoRI clone (BE5) both from the pUC9 sequencing primers and from two synthetic oligonucleotides (1586 to 1609 and 1762 to 1777, plus-strand polarity) established the position of the right viral-host junction at position 1790 followed by a 7 bp repeat (TGTAGGC) and a possible 2 bp inversion (TA) as shown in Fig. 2 and 3. Further sequencing into the viral DNA from both flanking sequences excluded any additional rearrangements of the integrated viral DNA within about 200 nucleotides downstream of 1886 (left junction) or nearly 400 nucleotides upstream of 1790 (right junction) as shown in Fig. 2.

Integrated HBV DNA thus comprises 3125 bp of essentially un-rearranged DNA of the adr subtype (Fowler et al., 1986) with a 7 bp duplication and 2 bp inversion at the right junction. The sequence includes the core (but not pre-core), surface (including pre-S1 and pre-S2) and polymerase genes, and an X gene truncated by 47 nucleotides at the 3' end (Fig. 2 and 3) which would encode an X gene product 16 amino acids short at the carboxyl end, followed by an additional leucine, and terminating within the 7 bp repeat. Both the pre-core AUG and most of the pre-core coding region are absent. The enhancer sequence downstream from S and the pre-core inhibitory sequence (1987 to 2431), whereas BB3 was just over 100 bp long and was presumed to be the viral BglII/BglII sequence (1987 to 2431), whereas BB3 was just over 100 bp and was expected to contain the junction (Fig. 1). Sequencing of the two clones confirmed these predictions and placed the left junction at HBV DNA position 1886 (or 1861 of the adr subtype which contains a 27 bp deletion) within the pre-core region (Fig. 2 and 3). Further confirmation of the sequence at the virus-host boundary was also obtained by sequencing from a synthetic primer from within BB2 (2021 to 2040, minus-strand polarity) using DNA from clone 5e (HindIII left of EcoRI) as template (Fig. 1). Analysis of the right viral-host junction was carried out by subcloning both AvaI/EcoRI and BamHI/EcoRI fragments based on previous restriction site assignments which placed the AvaI site at position 1475, from the enzymes tested, closest to the right boundary (Fig. 1). Sequencing of the BamHI/EcoRI clone (BE5) both from the pUC9 sequencing primers and from two synthetic oligonucleotides (1586 to 1609 and 1762 to 1777, plus-strand polarity) established the position of the right viral-host junction at position 1790 followed by a 7 bp repeat (TGTAGGC) and a possible 2 bp inversion (TA) as shown in Fig. 2 and 3. Further sequencing into the viral DNA from both flanking sequences excluded any additional rearrangements of the integrated viral DNA within about 200 nucleotides downstream of 1886 (left junction) or nearly 400 nucleotides upstream of 1790 (right junction) as shown in Fig. 2.

Integrated HBV DNA thus comprises 3125 bp of essentially un-rearranged DNA of the adr subtype (Fowler et al., 1986) with a 7 bp duplication and 2 bp inversion at the right junction. The sequence includes the core (but not pre-core), surface (including pre-S1 and pre-S2) and polymerase genes, and an X gene truncated by 47 nucleotides at the 3' end (Fig. 2 and 3) which would encode an X gene product 16 amino acids short at the carboxyl end, followed by an additional leucine, and terminating within the 7 bp repeat. Both the pre-core AUG and most of the pre-core coding region are absent. The enhancer sequence downstream from S and the pre-core inhibitory sequence (1987 to 2431), whereas BB3 was just over 100 bp long and was presumed to be the viral BglII/BglII sequence (1987 to 2431), whereas BB3 was just over 100 bp and was expected to contain the junction (Fig. 1). Sequencing of the two clones confirmed these predictions and placed the left junction at HBV DNA position 1886 (or 1861 of the adr subtype which contains a 27 bp deletion) within the pre-core region (Fig. 2 and 3). Further confirmation of the sequence at the virus-host boundary was also obtained by sequencing from a synthetic primer from within BB2 (2021 to 2040, minus-strand polarity) using DNA from clone 5e (HindIII left of EcoRI) as template (Fig. 1). Analysis of the right viral-host junction was carried out by subcloning both AvaI/EcoRI and BamHI/EcoRI fragments based on previous restriction site assignments which placed the AvaI site at position 1475, from the enzymes tested, closest to the right boundary (Fig. 1). Sequencing of the BamHI/EcoRI clone (BE5) both from the pUC9 sequencing primers and from two synthetic oligonucleotides (1586 to 1609 and 1762 to 1777, plus-strand polarity) established the position of the right viral-host junction at position 1790 followed by a 7 bp repeat (TGTAGGC) and a possible 2 bp inversion (TA) as shown in Fig. 2 and 3. Further sequencing into the viral DNA from both flanking sequences excluded any additional rearrangements of the integrated viral DNA within about 200 nucleotides downstream of 1886 (left junction) or nearly 400 nucleotides upstream of 1790 (right junction) as shown in Fig. 2.

We next analysed the flanking host sequences for homology with known human genomic sequences in both the GenBank and EMBL databanks using the DNASIS program. This revealed significant similarity of both host flanking sequences with repetitive sequences which map 3' to the β-globin cluster on the short arm of chromosome 11 as shown in Fig. 4 for the left flanking sequence. Aligning the host flanking sequences with the β-globin gene cluster sequence predicts a size for HB12 of un-rearranged genomic DNA in good agree-
ment with our estimation of 758 bp, determined from the mobility of the fragment on an agarose gel (not shown).

Integration of HBV DNA in this human hepatocellular carcinoma confirms the preference for insertion through the cohesive region of the viral genome with one of the junctions mapped at nucleotide 1790, close to the 5' end of the minus strand. We have found a duplication of the last 7 bp (1784 to 1790) possibly followed by a 2 bp inversion which would place the ‘breaking point’ at position 1792, where the 27 bp deletion in the adr subtype was mapped by Ono et al. (1983). This location is within a region identified for preferential viral integration in tumours analysed by Yaginuma et al. (1987) and Zhou et al. (1988), and compatible with the model of integration proposed by the former. In the report by Yaginuma et al. (1987) the integrated sequences also terminated within the X gene which was truncated at the carboxyl end. In view of the trans-activating function of the X gene, recently reported in the case of the c-myc proto-oncogene (Balsano et al., 1991), the potential role of truncated X gene products in hepatocarcinogenesis will require further investigation.

Our study also suggests that integration appears to have occurred in a region of semi-repetitive genomic DNA designated TBG41, located 3 kb downstream from the fl-globin gene (Hattori et al., 1985). This sequence is part of the KpnI family, made up of long interspersed repetitive sequences found in primate genomes. The

![Fig. 3. Map of viral-host junctions.](image)

![Fig. 4. Sequence homology between left L1CI host flanking sequences (clone HB12) and human fl-globin gene cluster (HUMHBB.DNA). Sequences in (a) and (b) are from both forward (HB12FOR.DNA) and reverse (HB12REV.SEQ) pUC9 sequencing primers.](image)
Short communication

sequence is about 6 kb in length and repeated about $10^4$ times per haploid genome. Our finding would corroborate a previous report that HBV integration may occur in association with repetitive sequences which may contain recombination signals and a number of direct and indirect repeats thought to facilitate rearrangements (Hattori et al., 1985). A significant amino acid similarity between human transferrins and amino acid sequences in TBG41 has been reported, transferrins being members of a protein family which includes melanoma antigen p97, ChBlym-1 and probably HuBlym-1. The potential significance of this association will be investigated further by evaluating the transforming potential of the 10 kb LIC1 cloned DNA.

We wish to thank Mrs Champa Patel for her invaluable technical assistance. This work was supported by the Cancer Research Campaign (J.M. and H.C.T.) and by a Gunnar Nilsson Research Award to J.M.

References


(Received 10 June 1991; Accepted 10 September 1991)