Journal of General Virology (1998), **79**, 591–600. Printed in Great Britain

Extensive analysis of duplicated-inverted hepatitis B virus integrations in human hepatocellular carcinoma

Pascal Pineau,[1] Agnès Marchio,[1] Marie-Geneviève Mattei,[2] Won-Ho Kim,[3] Jung-Koo Youn,[4] Pierre Tiollais[1] and Anne Dejean[1]

[1] Unité de Recombinaison et Expression Génétique, INSERM U163, Institut Pasteur, 75724 Paris Cedex 15, France
[2] Unité de Physiopathologie Chromosomique, INSERM U242, Hôpital d'Enfants, Groupe Hospitalier de la Timone, Marseille, France
[3] Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea
[4] Department of Microbiology, Ajou University School of Medicine, Suwon, South Korea

Hepatitis B virus (HBV) DNA is found chromosomally integrated into the genome of the majority of hepatocellular carcinomas (HCC) arising in chronic HBV carriers suggesting that, in some instances, viral sequences may be directly responsible for oncogenic conversion. In an attempt to clarify the oncogenic potential of integrated HBV sequences, we performed an extensive analysis of two single integrations present in HCC which developed in non-cirrhotic livers from HBsAg-positive Korean patients. In both cases, integrated viral sequences were characterized by a duplicated-inverted configuration involving the flanking cellular sequences, a pattern consistently found in many amplicons isolated from mammalian cells. Integration sites are characterized by an AT-rich content and the presence of topoisomerase I and II cleavage target sequences as well as other recombination-prone motifs. The chromosomal locations of the integration sites were determined as 8q13 and 10q22 in the human genome, two regions known to harbour genes involved in tumorigenesis. The *cis*-activating potential of the integrations in their original configuration was also investigated in a transient transfection assay in HepG2 cells. Integrated sequences, rather than activating heterologous promoters, show either no activity or a weak tendency to inhibit activation of neighbouring reporter genes. The implications of our findings for the understanding of primary liver cancer development are discussed.

Introduction

Chronic hepatitis B virus (HBV) infection is considered to be the aetiological factor responsible for more than 75% of hepatocellular carcinomas (HCC) reported worldwide (Nishioka, 1993). Chromosomal aberrations are one of the prominent characteristics of tumour progression. In virus-associated tumours like HCC or cervical carcinoma, viral sequence integrations represent a specific type of host DNA lesion. HBV DNA has been consistently found integrated in the genomic sequences of the majority of HCC (for a review see Buendia, 1992). Extensive rearrangements undergone by integrated viral sequences lead to deletion of flanking cellular sequences, amplifications, excisions and transpositions of viral inserts, as well as chromosome translocation at the integration sites (Hatada *et al.*, 1988; Tokino *et al.*, 1991; Pineau *et al.*, 1996). Such events render the direct identification of cellular genes affected by HBV integration more difficult and so far the precise role of HBV integrations in liver carcinogenesis remains unclear.

In contrast, in an animal model, activation of the MYC gene family by insertion of woodchuck hepatitis virus (WHV) sequences nearby has been found in more than 50% of woodchuck liver tumours analysed, supporting the hypothesis that WHV, similarly to avian or murine retroviruses, may act as an insertional mutagen. Enhanced expression of oncogenes can be detected following insertion of WHV sequences several hundred kilobases away (Fourel *et al.*, 1994). Furthermore, the N-*myc*2 promoter has been shown to be activated *in vitro* by WHV enhancers (Wei *et al.*, 1992). These findings support the hypothesis that, in human liver oncogenesis, the integration

Author for correspondence: Anne Dejean.
Fax +33 1 45 68 89 43. e-mail adejean@pasteur.fr

The EMBL accession numbers of the sequences reported in this paper are AJ000498 and AJ000499.

process could be accompanied by modified levels of expression of surrounding cellular genes presumably involved in the control of cell proliferation and/or differentiation (Tokino *et al.*, 1987; Zhou *et al.*, 1988; Hino *et al.*, 1989). Only two cases of insertional mutagenesis in human liver tumours have been reported to date (Dejean *et al.*, 1986; Wang *et al.*, 1990).

The characterization of the outcomes of insertional recombination is a major issue for tumour virology and for gene therapy. In this study, we report the cloning and the extensive analysis of two HBV integrations showing an inverted-duplicated structure that involves both virus and cellular flanking sequences, a pattern found in human and animal hepadnavirus-induced liver tumours (Hino *et al.*, 1986; Zhou *et al.*, 1988; Tokino *et al.*, 1991) and frequently associated with gene amplification in mammalian cells (Ford & Fried, 1986; Passananti *et al.*, 1987; Fried *et al.*, 1991). In order to better delineate the potential contribution of HBV-driven *cis*-activating mechanisms during human liver carcinogenesis, we also examined the impact of these integrations on a reporter gene in their 'native' forms in a transient transfection assay in HepG2 cells.

Methods

■ **Tumour samples.** HCC and their non-tumorous liver counterparts were surgically obtained from two patients of the Yonsei University College of Medicine in Seoul (South Korea), quickly frozen in liquid nitrogen and stored at −80 °C. The tumours 5T and 7T were obtained from a 44-year-old man and a 48-year-old man, respectively. Both tumours arose on non-cirrhotic liver parenchyma. The serological markers were positive for HBsAg and negative for anti-hepatitis C virus antibody. Patients were both infected by HBV subtype *adw*.

■ **DNA and RNA analyses.** Genomic DNA was prepared as described previously (Hansen *et al.*, 1993). DNA samples (20 μg) were analysed by Southern blotting after digestion with a restriction enzyme and separation by electrophoresis on a 0·8% agarose gel. Total RNA was isolated from tumour samples or transfected cells by a single-step method (Chomczynski & Sacchi, 1987) using the RNA-plus solution (Bioprobe) according to manufacturer's instructions; it was subsequently treated with DNAse I to remove traces of contaminating DNA. RNA (30 μg) was analysed by Northern blot hybridization.

Southern and Northern blotting were performed according to standard methods (Ausubel *et al.*, 1987). Densitometric analysis of films was performed with Image Quant software (Molecular Dynamics). DNA was sequenced by the chain termination method with the Sequenase version 2.0 sequencing kit (US Biochemical). Nucleotide sequence analyses were performed using software from the University of Wisconsin Genetics Computer Group (GCG) program package (Devereux *et al.*, 1984).

■ **Preparation and screening of genomic libraries.** Genomic DNA (60 μg) from the tumours 5T and 7T was partially digested with *Sau*3AI and fractionated in a 10–40% sucrose gradient. DNA fragments around 20 kb in size were collected and ligated in a *Bam*HI-treated λ-GEM11 vector (Promega). Ligation products were packaged *in vitro* in a Gigapack Gold extract (Stratagene) and used to infect *Escherichia coli* LE392. The libraries (1×10^6 p.f.u.) were screened using total ^{32}P-labelled HBV genome as a probe. Filters were hybridized and washed as previously reported. Phage DNA was prepared from large-scale cultures in LE392 according to standard methods (Maniatis *et al.*, 1982).

■ **Probes.** After purification with a Qiaex kit (Qiagen), DNA restriction fragments were radiolabelled by the random primer method with a Random-prime labelling kit (Amersham). The 645 bp RARβ (retinoic acid receptor β) probe is the fourth *Eco*RI restriction fragment (nt 2347–2992) from cDNA cloned by de Thé *et al.* (1987).

■ **Plasmid construction and activation assay.** The 8 kb *Hin*dIII–*Hin*dIII and 6 kb *Kpn*I–*Kpn*I fragments containing the duplicated HBV inserts and flanking sequences were isolated from clones λ5.116 and λ7.102, respectively, and subcloned into vector pKS(+) (Stratagene). These fragments were subsequently subcloned into pTK–LUC, which carries the herpes simplex virus thymidine kinase (TK) gene promoter upstream of the luciferase (LUC) gene and the SV40 polyadenylation signal (de Thé *et al.*, 1990). Insertion of HBV integrations upstream of the TK promoter yielded constructs p5TK–LUC and p7TK–LUC. Another HBV integration (105T), described previously (Pineau *et al.*, 1996), contained in a *Sac*I–*Bam*HI 5·0 kb fragment was cloned in the TK–LUC vector to obtain the p105TK–LUC construct. Plasmids pEnh1, which contains the HBV enhancer 1 (Enh1, nt 826–1380), and pRSV-βGAL were kindly provided by G. Fourel (Institut Pasteur, Paris, France)

The human hepatoblastoma HepG2 cells were grown in Dulbecco's modified Eagle medium with 10% foetal calf serum, antibiotics and 5% CO_2. Cells were plated 24 h before transfection. The different HBV/LUC constructs (15 μg) and the β-galactosidase expression vector pRSV-βGAL (5 μg), used as an internal control for transfection efficiency, were co-transfected in duplicate into semi-confluent HepG2 cells by the calcium phosphate precipitation method. For direct examination of RNA synthesis, 10 μg HBV plasmids were co-transfected with 5 μg carrier pKS(+) DNA per 10 cm dish. LUC activity was determined from cell extracts prepared 48 h after transfection as described previously (de Thé *et al.*, 1990). For each construct, three independent experiments were performed and results were normalized for transfection efficiency by reference to β-galactosidase activities. Activation is expressed on histograms as the ratio of the level of luciferase activity directed by the indicated construct to that directed by the TK promoter alone.

■ ***In situ* fluorescence hybridization.** Metaphase spreads were prepared from phytohaemagglutinin-stimulated human lymphocytes, cultured at 37 °C for 72 h. 5-Bromodeoxyuridine was added for the final 7 h of growth (60 μg per ml of medium) to ensure a chromosomal R-banding of good quality. The λ5.116 and λ7.105 phages containing an insert of 16–17 kb were biotinylated by nick translation with biotin-16-UTP, as outlined by the Boehringer Mannheim protocol. Hybridization to chromosome spreads was performed by a standard protocol (Pinkel *et al.*, 1986); for each slide, 400 or 500 ng biotinylated DNA was used. Before hybridization, the labelled probe was annealed with a 150-fold excess amount of Cot-1 DNA (GIBCO-BRL) for 45 min at 37 °C in order to compete with the aspecific repetitive sequences. The hybridized probe was detected with fluorescence isothiocyanate-conjugated avidin (Vector laboratories). Chromosomes were counterstained and R-banded with propidium iodide diluted in antifade solution (pH 11) as described previously (Lemieux *et al.*, 1992).

Results

Structural organization of integrated HBV DNA

Genomic DNA was extracted from HCC samples as well as the adjacent non-cirrhotic livers from six male patients living in Korea. Of the samples analysed by Southern blotting, two of

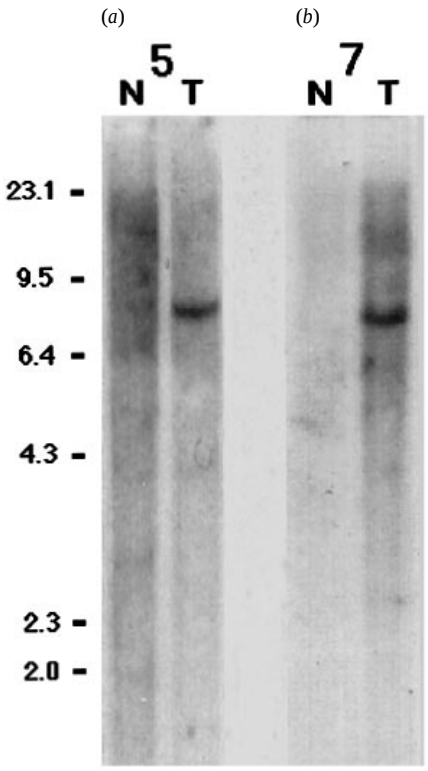

Fig. 1. Southern blot analysis of *Hin*dIII-digested DNA from the tumorous (T) and the corresponding non-tumorous (N) parts of 5T (*a*) and 7T (*b*). Blots were hybridized with an HBV genome probe. The reference sizes are expressed in kilobases.

them, tumours 5T and 7T, exhibited a single 8 kb hybridizing fragment after *Hin*dIII digestion corresponding to a unique HBV integration (Fig. 1). In the two patients from whom these samples came, no free viral genome DNA was detected in the tumorous or non-tumorous samples.

To investigate the possible causal relationship between HBV DNA integration and subsequent tumour development in the two non-cirrhotic livers, we decided to further characterize the structure of integrated viral DNA in tumours 5T and 7T. Four (λ5.82, λ5.83, λ5.101, λ5.116) and two (λ7.102, λ7.105) independent HBV-hybridizing clones were obtained from genomic libraries 5T and 7T, respectively. Restriction maps of the 20 and 20·5 kb genomic loci covered by the four 5T and the two 7T overlapping clones, respectively, are shown in Fig. 2. Based on both the restriction maps and the nucleotide sequence analysis, integrated HBV DNA was found to be of the *adw* subtype in the two patients (Valenzuela *et al*., 1979). Large inverted repeated structures consisting of portions of the HBV genome along with flanking cellular sequences were present in the rearranged alleles of the 5T and 7T genomes (Fig. 2). To more precisely analyse the organization of integrated viral DNA, an 8 kb *Hin*dIII and a 6 kb *Kpn*I fragment encompassing the whole HBV sequence were subcloned from phages λ5.116 and λ7.102, respectively, and subjected to further nucleotide sequence analysis.

In tumour 5T, integrated HBV DNA consisted of two DNA segments of 1·5 and 3 kb linked in an inverted and partially duplicated fashion (Fig. 2*b*). The shorter fragment began in the preS2 domain (nt 60) and ended up in the cohesive-end region at position 1614, a few bp downstream of the second direct repeat (DR2, nt 1592–1602) (Fig. 2*b*) (Dejean *et al*., 1984). The larger fragment covered the virus genome from nt 1825, 1 bp away from the first direct repeat (DR1) motif (nt 1826–1836), to nt 1614 (the same as the shorter fragment). Regarding the orientation of virus transcription, the two segments were found to be tail-to-tail. The inverted duplication extended to the flanking cellular sequences for a distance of at least 5 kb (Fig. 2*a*).

Clones from tumour 7T (Fig. 2*c*) harboured a similar structure containing 5·4 kb of integrated HBV DNA composed of two fragments of 3·2 and 2·2 kb (Fig. 2*d*). The larger fragment consisted of an almost complete virus genome extending from the 5′ coding region of the initiation ATG codon of the X gene (nt 1388) to nt 1366, 10 bp upstream of the X gene ATG. The smaller fragment started within the core region (nt 2390) and ended up at the same virus–cell junction as that observed for the larger segment (nt 1366). The duplicated region also involved a minimum of 3·5 kb of adjacent cellular sequences.

Comparison of HBV DNA sequences present at the virus–virus junctions in each of the two loci revealed short patches of homology (data not shown). At the 5T virus–virus junction, four out of seven nucleotides were identical while at the 7T virus–virus junction a slightly stronger homology was noted with seven out of ten common nucleotides. Such short homologies are thought to have some impact on illegitimate recombination mechanisms by allowing the initial positioning of the two DNA strands.

Overall, HBV integrations in tumours 5T and 7T harbour a general pattern of inverted duplications with an asymmetric non-duplicated portion found at the centre of the duplication (Ford & Fried, 1986; Hyrien *et al*., 1988).

Integration sites

To determine whether the viral genome integrates in close proximity to a putative cellular gene, we looked for cellular DNA devoid of repetitive sequences in the vicinity of the integrations. Hybridization with ^{32}P-labelled human genomic DNA resulted in strong hybridization signals indicating that, in both 5T and 7T, the HBV genome integrated within regions enriched in repetitive elements (shaded boxes in Fig. 2*a*, *c*).

At the tumour 5T integration site, we were able to identify a 2·0 kb *Kpn*I–*Hin*dIII fragment (referred to as KH2.0 in Fig. 2*a*), which showed no hybridization with human repetitive DNA and was located 1·5 kb away from the virus integration site. When used as a probe during Southern blot analysis of *Hin*dIII-digested tumorous and non-tumorous DNA from patient 5, a single hybridizing band was detected in both

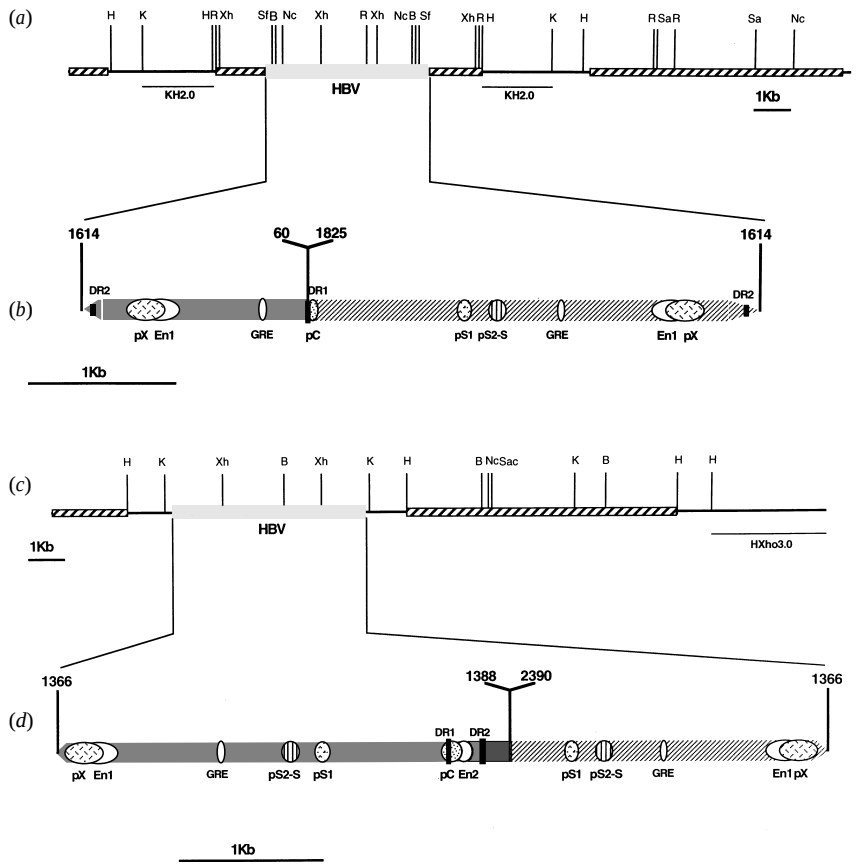

Fig. 2. Restriction maps of HBV integration sites in tumours 5T (*a*, *b*) and 7T (*c*, *d*). Integrated viral sequences are represented by solid boxes and repetitive element-enriched cellular DNA regions as shaded boxes. Probes KH2.0 and HX3.0 (HXho3.0) are depicted by horizontal bars at their corresponding positions on the maps. Abbreviations: H, *Hin*dIII; K, *Kpn*I; R, *Eco*RI; Xh, *Xho*I; Sf, *Sfi*I; B, *Bam*HI; Nc, *Nco*I; Sa, *Sal*I. The viral segment breakpoints with their corresponding nucleotide position on the HBV *adw* genome are indicated above each insert representation. Regulatory elements of HBV are represented schematically by ovals: enhancers 1 and 2 (En1, En2), glucocorticoid responsive element (GRE), promoters for the C, preS2-S, preS1 and X genes (pC, pS2-S, pS1, pX). DR1 and DR2 are depicted as black vertical bars on the map.

samples (Fig. 3*a*). An estimated quantification of the KH2.0 hybridizing band in the T lane compared with that in the N lane with respect to an internal control (RARβ probe) revealed the same value suggesting that duplication of one allele is associated with deletion of the other (data not shown).

Similarly, we analysed the integration site of tumour 7T and isolated a *Hin*dIII–*Xho*I DNA fragment of 3 kb (HX3.0 on Fig. 2*c*) located 10 kb away from the HBV sequences. On a Southern blot from patient 7, the HX3.0 fragment, used as a probe, revealed a unique band with a weak smear (Fig. 3*b*). Densitometric analysis of the HX3.0 hybridizing band in the tumour lane revealed that the signal was 1·5 times that of the non-tumorous sample with respect to the internal RARβ control suggesting the presence of three copies in the 7T tumorous sample.

Northern blot analysis of poly(A) RNA isolated from the tumorous and non-tumorous regions of samples 5T and 7T using probes KH2.0 and HX3.0, respectively, did not reveal any significant signals suggesting that these sequences are either not transcribed or transcribed at a very low level (data not shown).

We determined the sequences of the 1190 and 1335 nt of genomic DNA immediately flanking the integrated viral sequences in 5T and 7T, respectively (EMBL accession numbers of the sequences are AJ000498 and AJ000499, respectively). Screening of databases for each of the two sequences did not reveal any similarity to known genes. We used several simple analyses to investigate the potential role that the sequence structure of the cellular DNA may play in the position of HBV integration. Both cellular integration sites were AT-rich regions with 68 and 73% AT in 5T and 7T, respectively. Such composition is usually considered to be a preferential target for virus or transposable element integrations (Zambryski, 1988; Passananti *et al.*, 1987; Fried *et al.*, 1991; Kahn *et al.*, 1994). It has been shown recently that DNA integration occurs preferentially in unstable regions (Rassool *et al.*, 1991; Merrihew *et al.*, 1996). Using the Findpatterns searching tool of the Wisconsin GCG package, we also looked for short sequence motifs that have been implicated in recombination, including chi-like sequences, immunoglobulin heptamers, alpha protein recognition sites, the consensus topoisomerase I and II recognition sites, three hypervariable minisatellite sequences and consensus common hotspots for gene deletions and translocations (Kahn *et al.*, 1994; Cullen *et al.*, 1995; Wilke *et al.*, 1996). At each integration site we were able to detect one alpha protein recognition sequence, a heptanucleotide motif which should occur only once every 16384 bases. In addition, one topoisomerase II site, a 2- to 3-fold excess of topoisomerase I cleavage sites (5′ C/GTT 3′) and a 1·5- to 2-fold excess of deletion/translocation-related

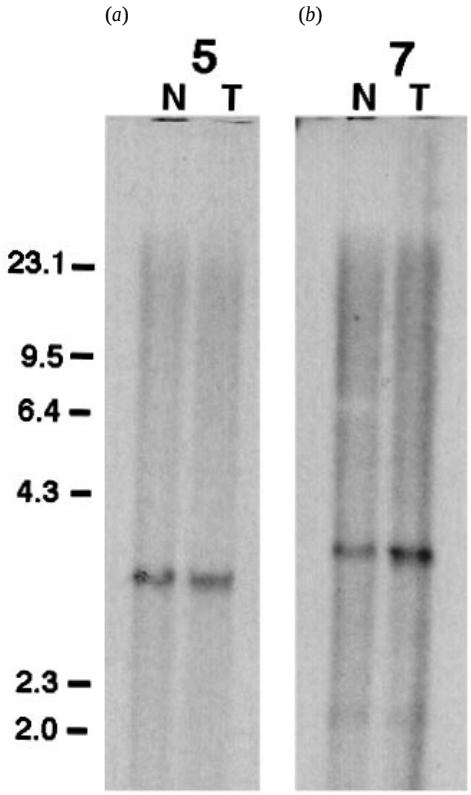

Fig. 3. Southern-blot hybridization of *Hin*dIII-digested DNA from the tumorous (T) and the corresponding non-tumorous (N) parts of liver from 5T (*a*) and 7T (*b*). Blots (*a*) and (*b*) were probed by flanking sequences KH2.0 and HX3.0, respectively.

consensus sequences were present in both flanking sequences. The over-representation of such recombination-prone motifs suggests that both viruses integrated in potentially unstable genomic regions.

We were interested in mapping the two HBV integration sites in the human genome. To do this, phages λ5.116 and λ7.102 were used as probes for *in situ* hybridization. The great majority of the fluorescent spots were concentrated on the 10q22.1 and 8q13 bands of the human genome (Fig. 4). These data indicate that HBV integrated in a region corresponding to the q22.1 band of chromosome 10 in patient 5 and to the q13 band of chromosome 8 in patient 7.

Cis-activity of integrated virus sequences

Cis-activity of integrated HBV sequences has been suspected in some instances to play a part in the tumorigenic process (Tokino *et al.*, 1987; Zhou *et al.*, 1988; Hino *et al.*, 1989). The high prevalence of duplicated-inverted HBV integrated sequences in HCC evokes the situation described for feline leukaemia virus-induced tumours in which it has been shown that repeated integrated sequences behave as strong enhancers (Athas *et al.*, 1995). Mapping studies have demonstrated the existence of two enhancers, Enh1 (nt 965–1235) and Enh2 (nt 1635–1745), within the HBV genome. To determine whether duplicated HBV integrations that contain at least one of the two enhancer regions are endowed with *cis*-activating potential, we inserted integrations cloned from tumours 5T and 7T, as well as from tumour 105T, a previously reported HCC case (Pineau *et al.*, 1996), in front of a TK promoter–LUC reporter gene construct and we tested the different vectors by transient transfection assay in HepG2 cells.

Unexpectedly, none of the three integrations tested (Fig. 5, constructs 3, 4 and 5) exhibited significant enhancer activity when compared to the parental TK–LUC vector alone (construct 1). However, virus sequences in constructs 3 and 5 conferred a 2- and 3-fold repressing activity, respectively. Under the same conditions the 554 bp fragment corresponding to Enh1 (construct 2) induced a 12-fold increase in LUC activity, which is comparable to that obtained in previous work (Wei *et al.*, 1992). Nucleotide sequence analysis of the enhancer elements present in tumour 5T and 7T integrations revealed only minor changes compared to the original Enh1 and Enh2 sequences (Galibert *et al.*, 1979; Valenzuela *et al.*, 1979). Moreover, these mutations have already been described in variant viruses (Norder *et al.*, 1993). It is therefore doubtful that such mutations could abrogate the activating capacity of the enhancer sequences. Another possible explanation for the differences observed between the enhancer properties of HBV integrations reported previously and our data is the use here of full-length 'native' integrations containing enhancer elements in their original context whereas other workers used only minimal subgenomic regulatory sequences (Wei *et al.*, 1992; Fourel *et al.*, 1992, 1996; Ueda *et al.*, 1996).

Discussion

In this report, we describe the structure, chromosomal localization and *cis*-activating potential of HBV integrations in two HCC arising on non-cirrhotic livers. We found that, in both cases, the integrated virus sequences showed an inverted-duplicated organization involving host flanking DNA. Inverted duplications are usually a rare event in the DNA tumour virus integration process (Lawrence *et al.*, 1988; Hurley *et al.*, 1991). They are, however, a common feature in hepadnaviral integrations. A high prevalence of such integration patterns has also been observed in the majority of integrants of *Agrobacterium tumefaciens* T-DNA, a transposon causing neoplastic diseases in dicotyledonous plants (Zambryski, 1988). The fact that both the T-DNA and part of the HBV genome are single-stranded may not be inconsistent with both their ability to 'invade' target genomic sequences and the structure of their integrated forms (Gheysen *et al.*, 1987).

In mammalian cells, head-to-head or tail-to-tail inverted duplication arrays are a common trait of amplified regions and it has been suggested that the formation of such inversions constitutes the initial step leading to genomic instability involving both amplification and deletion processes (Ford & Fried, 1986). These structures form hairpins or stem–loop

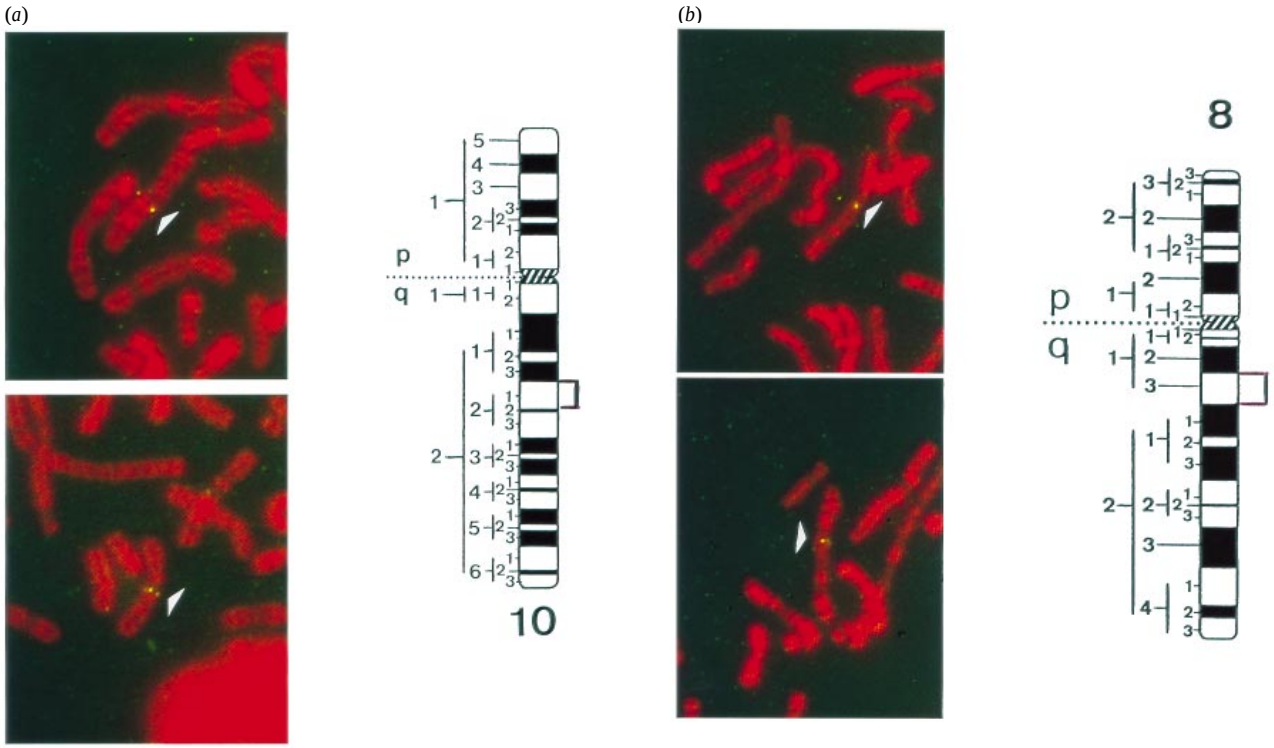

Fig. 4. FISH mapping of the λ5.116 (*a*) and λ7.102 (*b*) phages to human chromosomes 10q22.1 and 8q13, respectively. Partial views are shown of banded metaphase chromosomes stained with propidium iodide; the arrowheads show the FITC–avidin–biotin–λ fluorescent signal. Ideograms of the human G-banded chromosomes 8 and 10 with the sites of hybridization are shown for each case.

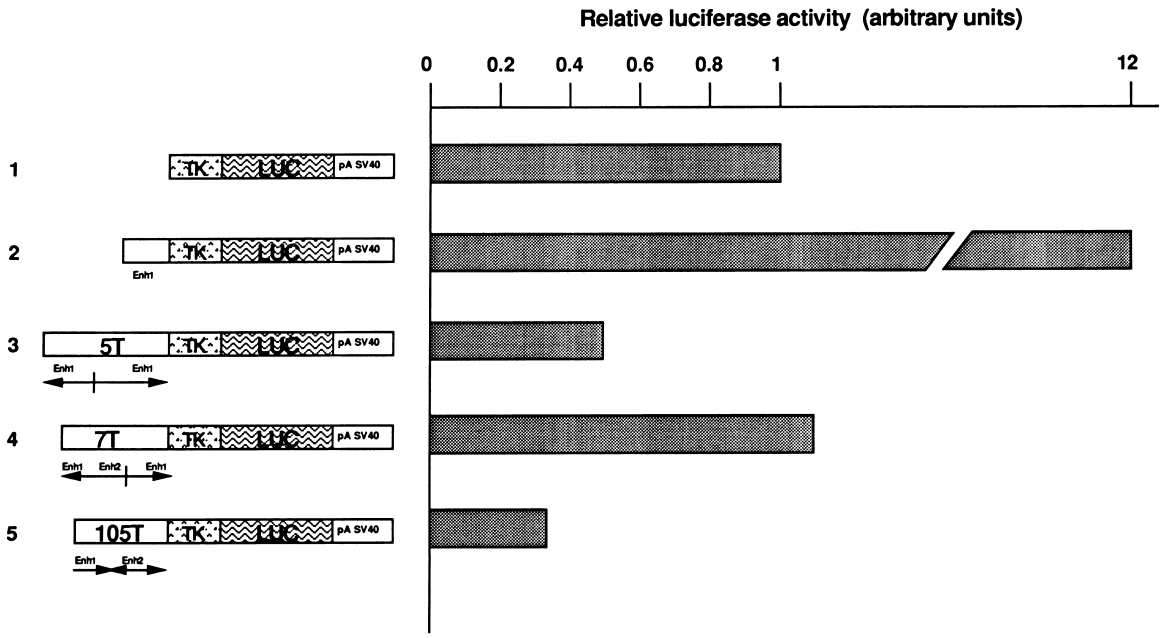

Fig. 5. Enhancer activity of integrated viral sequences in HepG2 cells. A schematic diagram of plasmids used is shown on the left. Constructs contain either Enh1 or HBV integrated sequences in their original configuration (open boxes) fused to a TK promoter combined with a LUC reporter gene and SV40 polyadenylation signal. The orientation of enhancer elements is depicted below each original integrated sequence. The results shown here are the average of three independent transfection experiments, each done in duplicate.

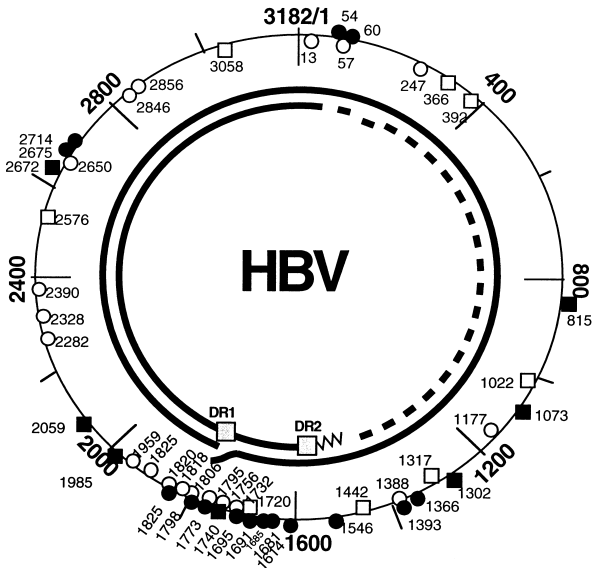

Fig. 6. Schematic representation of HBV virus–virus or virus–cell recombination sites collected from published duplicated-inverted HBV integrations. The HBV genome is represented under its circular configuration with the short strand in the encapsidated form indicated by a dashed line. On the circle surrounding HBV genome, closed circles and squares represent virus–human DNA junctions in tail-to-tail and head-to-head duplicated integrations, respectively; nucleotide positions are indicated. Open circles and squares represent the recombination sites corresponding to virus–virus DNA junctions in the configuration described previously.

structures which are excised from their templates and constitute the precursors for circular inverted dimers often observed during gene amplification (Ruiz & Wahl, 1988; Fried *et al.*, 1991). Alternatively, a study of transgenic mice has recently shown that the inverted repeat region can promote deletions extending into the DNA flanking the transgene (Collick *et al.*, 1996). In the case of tumours 5T and 7T, it is quite clear that the viral duplicated inversions that involve flanking cellular sequences are produced only after integration of HBV DNA into the host chromosome. These oligomeric forms of HBV may possess the capacity to be excised and transposed within the human genome, as suggested previously (Ziemer *et al.*, 1985). However, they are unable to further generate expanded chromosomal regions. This might be due to their lack of a functional DNA replication origin (Nonet *et al.*, 1993) and/or the absence of cellular factors responsible for the amplification process in hepatocytes (Cohen *et al.*, 1994).

An important issue to consider is why HBV sequences are particularly prone to adopt an inverted-duplicated configuration *in vivo* as similar structures are almost never identified with other DNA tumour viruses. One can formulate several hypotheses to explain this unique phenomenon: (i) some specific traits of the primary or secondary structure of the HBV genome permit integrated DNA to evolve toward a duplicated structure endowed with a unique stability; (ii) the repair process at a target site may lead integrated and flanking sequences to adopt such a specific pattern; and (iii) functional properties associated with inverted duplications may confer a selective advantage if either enhancers need to be present in duplicate or duplication is required to maintain a specific chromatin configuration necessary to influence some critical gene expression (Gubbay *et al.*, 1992). A functional role for inverted-duplicated sequences is indicated by the fact that such sequences encoding drug-resistance genes are readily lost in cell culture in the absence of drug selection (Ruiz & Wahl, 1988).

Examination of virus–cell or virus–virus breakpoints in HBV integrations reported so far in the literature reveals that they are not evenly distributed along the virus genome suggesting the existence of a non-random mechanism of integration (Fig. 6). Insertion of HBV sequences into the host DNA is frequently characterized by disruption of the viral genome at the cohesive-end region (nt 1602–1836) and even more consistently in the vicinity of DR1 (Fig. 6) (Nagaya *et al.*, 1987). The region encompassing nt 1600–2000 reaches a recombination site density almost 5-fold higher than the remaining part of the genome. In this respect, it is noteworthy that three out of the four viral junctions present in tumour 5T integration lie within the cohesive-end region. Cellular sequences flanking the two viral integrations contain many illegitimate recombination-prone motifs such as topoisomerase I and II cleavage sites, deletion/translocation-related consensus sequences or alpha protein recognition sequences. Thus, these host DNA sequences represent two particularly suitable cellular targets for virus integration.

Enhancer elements are frequently associated with viral genomes. They function in an orientation-independent manner sometimes over considerable distance (Lazo *et al.*, 1990). The flexibility in their interaction with promoter elements raises the possibility of improper activation of cellular genes by transposed enhancers. Some of them, particularly the simian virus 40 (SV40) enhancer, are able to activate promoters of a variety of genes. The woodchuck model has shown that insertion of WHV enhancers I and II can up-regulate the N-*myc* gene promoter *in vivo* and in transient transfection assays (Wei *et al.*, 1992). Moreover it has been suggested that, in WHV-induced tumours, viral sequence activity is able to stimulate transcription over a distance of several hundred kilobases (Fourel *et al.*, 1994). Our experiments indicate that, in contrast to previous findings, integrated HBV sequences rather than activating heterologous promoters show either no activity or a weak tendency to inhibit activation of neighbouring reporter genes. This discrepancy could be partly explained by the fact that, so far, only selected parts of the viral genome have been used in transient transfection studies, whereas, in the present study, we used full-length integrations in activation assays (Wei *et al.*, 1992; Fourel *et al.*, 1992, 1996; Ueda *et al.*, 1996). A second hypothesis for this lack of function may be that the juxtaposition of several regulatory elements caused by rearranged integrations in tumours 5T, 7T and 105T could result in an enhancer occlusion analogous to that observed with the

SV40 enhancer or the avian leukosis virus long terminal repeat in certain circumstances (Kadesh & Berg, 1986; Cullen *et al.*, 1984). Another possibility may be that the point mutations observed in the enhancer regions of the virus integrations affect the binding of transcriptional activators. Finally, our experiments using an *in vitro* model system may be only distantly related to the situation prevailing *in vivo* where enhancer activity is highly site-dependent (Walters *et al.*, 1996).

We mapped the HBV integration sites from tumours 7T and 5T in chromosomes 8q13 and 10q22.1, respectively. No specific translocation breakpoints have been mapped in these two chromosomal regions. However, fragile sites FRA8F and FRA10D are present in 8q13 and 10q22.1 (Genome Database, 1997). So far, no HBV integration has been described on 8q or 10q but both chromosomal arms are considered as preferential regions of integration for human papillomaviruses (Couturier *et al.*, 1991; Dürst *et al.*, 1987; Kahn *et al.*, 1994). A putative protooncogene, *Lyn*, and a tumour suppressor gene, *PTEN/MMAC*1, have been described in 8q13 and 10q23, respectively (Yamanashi *et al.*, 1987; Li *et al.*, 1997). Other genes that may play a role in cell proliferation are located at these cytogenetic locations: the genes encoding the telomeric repeat binding factor (*TRF*1), the cAMP phosphodiesterase 7A (*PDE7A*) and the GTP-binding mitogen-induced protein (*GEM*) were mapped in 8q13 and the genes encoding the dimerization co-factor of hepatocyte nuclear factor 1 (*DCOH*) and vinculin (Genome Database, 1997) are located in 10q22.

The liver plays a central role in the pathophysiology of many genetic and infectious diseases. It is therefore an important organ for targeted gene therapy (Wilson, 1996). Recent developments suggest that stable integration of virus vectors will provide the necessary long-lasting expression of therapeutic genes in the future (Overturf *et al.*, 1996). Thus, to overcome the risk associated with therapeutic utilization of virus vectors, the characterization of virus integrations is of great importance for a better understanding of the integration mechanism, chromatin organization and stable gene expression in the liver.

We are grateful to F. R. Lehembre and M. Flajolet for help in sequence analysis and transfection assays. We thank G. Fourel for generous gifts of plasmids pEnh1 and pTK–LUC, and HepG2 ATCC cell line. This work was supported by the Ligue Nationale contre le Cancer, the Association contre le Cancer and the Grant 1992 of Yonsei Institute for Cancer Research. P.P. was supported by the Fondation pour la Recherche Médicale and the Institut Electricité-Santé.

References

Athas, G. B., Choi, B., Prabhu, S., Lobelle-Rich, P. & Levy, L. S. (1995). Genetic determinants of feline leukemia virus-induced multicentric lymphomas. *Virology* **214**, 431–438.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987). Current protocol in molecular biology. New York: John Wiley & Sons.

Buendia, M. A. (1992). Hepatitis B viruses and hepatocellular carcinoma. *Advances in Cancer Research* **59**, 167–226.

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analytical Biochemistry* **162**, 156–159.

Cohen, S., Hassin, D., Karby, S. & Lavi, S. (1994). Hairpin structures are the primary amplification products: a novel mechanism for generation of inverted repeats during gene amplification. *Molecular and Cellular Biology* **14**, 7782–7791.

Collick, A., Drew, J., Penberth, J., Bois, P., Luckett, J., Scaerou, F., Jeffreys, A. & Reik, W. (1996). Instability of long inverted repeats within mouse transgenes. *EMBO Journal* **15**, 1163–1171.

Couturier, J., Sastre-Garau, X., Schneider-Maunoury, S., Labib, A. & Orth, G. (1991). Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. *Journal of Virology* **65**, 4534–4538.

Cullen, B. R., Lomedico, P. T. & Ju, G. (1984). Transcriptional interference in avian retroviruses – implications for the promoter insertion model of leukaemogenesis. *Nature* **307**, 241–245.

Cullen, M., Erlich, H., Klitz, W. & Carrington, M. (1995). Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus. *American Journal of Human Genetics* **56**, 1350–1358.

Dejean, A., Sonigo, P., Wain-Hobson, S. & Tiollais, P. (1984). Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11 base pair direct repeat. *Proceedings of the National Academy of Sciences, USA* **81**, 5350–5354.

Dejean, A., Bougueleret, L., Grzeschik, K. H. & Tiollais, P. (1986). Hepatitis B virus DNA integration in a sequence homologous to v-*erbA* and steroid receptor genes in a hepatocellular carcinoma. *Nature* **322**, 70–72.

de Thé, H., Marchio, A., Tiollais, P. & Dejean, A. (1987). A novel thyroid hormone receptor-related gene inappropriately expressed in human hepatocellular carcinoma. *Nature* **330**, 667–669.

de Thé, H., Vivanco-Ruiz, M. M., Tiollais, P., Stunnenberg, H. & Dejean, A. (1990). Identification of a retinoic acid response element in the retinoic acid receptor beta. *Nature* **363**, 177–180.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

Dürst, M., Croce, C. M., Gissmann, L., Schwarz, E. & Huebner, K. (1987). Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. *Proceedings of the National Academy of Sciences, USA* **84**, 1070–1074.

Ford, M. & Fried, M. (1986). Large inverted duplications are associated with gene amplification. *Cell* **45**, 425–430.

Fourel, G., Transy, C., Tennant, B. C. & Buendia, M. A. (1992). Expression of the woodchuck N-*myc*2 retroposon in brain and in liver tumors is driven by a cryptic N-*myc* promoter. *Molecular and Cellular Biology* **12**, 5336–5344.

Fourel, G., Couturier, J., Wei, Y., Apiou, F., Tiollais, P. & Buendia, M. A. (1994). Evidence for long-range oncogene activation by hepadnavirus insertion. *EMBO Journal* **13**, 2526–2534.

Fourel, G., Ringeisen, F., Flajolet, M., Tronche, F., Pontoglio, M., Tiollais, P. & Buendia, M.-A. (1996). The HNF-1/HNF-4-dependent WE2 element of woodchuck hepatitis virus controls viral replication and can activate the N-*myc*2 promoter. *Journal of Virology* **70**, 8571–8583.

Fried, M., Feo, S. & Heard, E. (1991). The role of inverted duplication in the generation of gene amplification in mammalian cells. *Biochimica et Biophysica Acta* **1090**, 143–155.

Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. (1979). Nucleotide sequence of the Hepatitis B virus genome (subtype *ayw*) cloned in *E. coli*. *Nature* **281**, 646–650.

Genome Database (1997). http://gdbwww.org. Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Gheysen, G., Van Montagu, M. & Zambryski, P. (1987). Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences. *Proceedings of the National Academy of Sciences, USA* **84**, 6169–6173.

Gubbay, J., Vivian, N., Economou, A., Jackson, D., Goodfellow, P. & Lovell-Badge, R. (1992). Inverted repeat structure of the *Sry* locus in mice. *Proceedings of the National Academy of Sciences, USA* **89**, 7953–7957.

Hansen, L. J., Tennant, B. C., Seeger, C. & Ganem, D. (1993). Differential activation of myc gene family members in hepatic carcinogenesis by closely related hepatitis B viruses. *Molecular and Cellular Biology* **13**, 659–667.

Hatada, I., Tokino, T., Ochiya, T. & Matsubara, K. (1988). Co-amplification of integrated hepatitis B virus DNA and transforming gene hst-1 in a hepatocellular carcinoma. *Oncogene* **3**, 537–540.

Hino, O., Shows, T. B. & Rogler, C. E. (1986). Hepatitis B virus integration site in hepatocellular carcinoma at chromosome (17;18). translocation. *Proceedings of the National Academy of Sciences, USA* **83**, 8338–8342.

Hino, O., Ohtake, K. & Rogler, C. E. (1989). Features of two hepatitis B virus DNA integrations suggest mechanisms of HBV integration. *Journal of Virology* **63**, 2638–2643.

Hurley, E. A., Klaman, L. D., Agger, S., Lawrence, J. B. & Thorley-Lawson, D. A. (1991). The prototypical Epstein–Barr virus-transformed lymphoblastoid cell line IB4 is an unusual variant containing integrated but no episomal viral DNA. *Journal of Virology* **65**, 3958–3963.

Hyrien, O., Debatisse, M., Buttin, G. & de Saint Vincent, R. (1988). The multicopy appearance of a large inverted duplication and the sequence at the inversion joint suggest a new model for gene amplification. *EMBO Journal* **7**, 407–417.

Kadesh, T. & Berg, P. (1986). Effects of the position of the simian virus 40 enhancer on expression of multiple transcription units in a single plasmid. *Molecular and Cellular Biology* **6**, 2593–2601.

Kahn, T., Turazza, E., Ojeda, R., Bercovich, A., Stremlau, A., Lichter, P., Poutska, A., Grinstein, S. & Zur Hausen, H. (1994). Integration of human papillomavirus type 6a DNA in tonsillar carcinoma: chromosomal localization and nucleotide sequence of the genomic target region. *Cancer Research* **54**, 1305–1312.

Lawrence, J. B., Villnave, C. A. & Singer, R. H. (1988). Sensitive, high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a lymphoma line. *Cell* **52**, 51–61.

Lazo, P. A., Lee, J. S. & Tsichlis, P. N. (1990). Long-distance activation of the *Myc* protooncogene by provirus insertion in *Mlvi-1* or *Mlvi-4* in rat T-cell lymphomas. *Proceedings of the National Academy of Sciences, USA* **87**, 170–173.

Lemieux, N., Dutrillaux, B. & Viegas-Péquignot, E. (1992). A simple method for simultaneous R- and G-banding and fluorescence in situ hybridization of small single-copy genes. *Cytogenetics and Cell Genetics* **59**, 311–312.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S., Giovanella, B. C., Ittman, M., Tycko, B., Hishboosh, H., Wigler, M. H. & Parsons, R. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast and prostate cancer. *Science* **275**, 1943–1947.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Merrihew, R. V., Marburger, K., Pennington, S. L., Roth, D. B. & Wilson, J. H. (1996). High-frequency illegitimate integration of transfected DNA preintegrated target sites in a mammalian genome. *Molecular and Cellular Biology* **16**, 10–18.

Nagaya, T., Nakamura, T., Tokino, T., Tsurimoto, T., Imai, M., Mayumi, T., Kamino, K., Yamamura, K. & Matsubara, K. (1987). The mode of integration of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. *Genes & Development* **1**, 773–782.

Nishioka, K. (1993). Hepatitis C virus infection and hepatocellular carcinoma. In *Primary Liver Cancer: Etiological and Progression Factors*, pp. 41–48. Edited by C. Bréchot. Boca Raton, Fla.: CRC Press.

Nonet, G. H., Carroll, S. M., DeRose, M. L. & Wahl, G. M. (1993). Molecular dissection of an extrachromosomal amplicon reveals a circular structure consisting of an imperfect inverted duplication. *Genomics* **15**, 543–558.

Norder, H., Courouce, A. M. & Magnius, L. O. (1993). Complete nucleotide sequences of six hepatitis B viral genomes encoding the surface antigen subtypes ayw4, adw4q-, and adrq- and their phylogenetic classification. *Archives of Virology* **8** (Suppl.), 189–199.

Overturf, K., Al-Dhalimy, M., Tanguay, R., Brantly, M., Ou, C.-N., Finegold, M. & Grompe, M. (1996). Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinemia. *Nature Genetics* **12**, 266–273.

Passananti, C., Davies, B., Ford, M. & Fried, M. (1987). Structure of an inverted duplication formed as a first step in a gene amplification event: implications for a model of gene amplification. *EMBO Journal* **6**, 1697–1703.

Pineau, P., Marchio, A., Terris, B., Mattei, M. G., Tu, Z.-X., Tiollais, P. & Dejean, A. (1996). A t(3;8) chromosomal translocation associated with hepatitis B virus integration involves the carboxypeptidase N locus. *Journal of Virology* **70**, 7280–7284.

Pinkel, D., Straume, T. & Gray, J. W. (1986). Cytogenetic analysis using quantitative high sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences, USA* **83**, 2934–2938.

Rassool, F. V., McKeithan, T. W., Neilly, M. E., Van Melle, E., Espinosa, R., III & Le Beau, M. M. (1991). Preferential integration of marker DNA into chromosomal fragile site at 3p14: an approach to cloning fragile sites. *Proceedings of the National Academy of Sciences, USA* **88**, 6657–6661.

Ruiz, J. C. & Wahl, G. M. (1988). Formation of an inverted duplication can be an initial step in gene amplification. *Molecular and Cellular Biology* **8**, 4302–4313.

Tokino, T., Fukushige, S., Nakamura, T., Nagaya, T., Murotsu, T., Shiga, K., Aoki, N. & Matsubara, K. (1987). Chromosomal translocation and inverted duplication associated with integrated hepatitis B virus in hepatocellular carcinomas. *Journal of Virology* **61**, 3848–3854.

Tokino, T., Tamura, H., Hori, N. & Matsubara, K. (1991). Chromosome deletions associated with hepatitis B virus integration. *Virology* **185**, 879–882.

Ueda, K., Wei, Y. & Ganem, D. (1996). Activation of N-*myc2* gene expression by *cis*-acting elements of oncogenic hepadnaviral genomes: key role of enhancer II. *Virology* **217**, 413–417.

Valenzuela, P., Grey, P., Quiroga, M., Zaldivar, J., Goodman, H. M. & Rutter, W. J. (1979). Nucleotide-sequence of the gene coding for the major protein of hepatitis B virus surface-antigen. *Nature* **280**, 815–819.

Walters, M. C., Magis, W., Fiering, S., Eidmiller, J., Scamzo, D., Groudine, M. & Martin, D. I. K. (1996). Transcriptional enhancers act in

cis to suppress position-effect variegation. *Genes & Development* **10**, 185–195.

Wang, J., Chenivesse, X., Henglein, B. & Bréchot, C. (1990). Hepatitis B virus integration in a cyclin A gene in a human hepatocellular carcinoma. *Nature* **343**, 555–557.

Wei, Y., Fourel, G., Ponzetto, A., Silvestro, M., Tiollais, P. & Buendia, M. A. (1992). Hepadnavirus integration: mechanisms of activation of the N-*myc*2 retrotransposon in woodchuck liver tumors. *Journal of Virology* **66**, 5265–5276.

Wilke, C. M., Hall, B. K., Hoge, A., Paradee, W., Smith, D. I. & Glover, T. W. (1996). FRA3B extends over a broad region and contains a spontaneous HPV-16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites. *Human Molecular Genetics* **5**, 187–195.

Wilson, J. M. (1996). Round two for liver gene therapy. *Nature Genetics* **12**, 232–233.

Yamanashi, Y., Fukushuge, S.-I., Semba, K., Sukegawa, J., Miyajama, N., Matsubara, K.-I., Yamamoto, T. & Toyoshima, K. (1987). The yes-related cellular gene lyn encodes a possible tyrosine kinase similar to $p56^{lck}$. *Molecular and Cellular Biology* **7**, 237–243.

Zambryski, P. (1988). Basic process underlying *Agrobacterium*-mediated DNA transfer to plant cells. *Annual Review of Genetics* **22**, 1–30.

Zhou, Y. Z., Slagle, B. L., Donehower, L. A., van Tuinen, P., Ledbetter, D. H. & Butel, J. S. (1988). Structural analysis of a hepatitis B virus genome integrated into chromosome 17p of a human hepatocellular carcinoma. *Journal of Virology* **62**, 4224–4231.

Ziemer, M., Garcia, P., Shaul, Y. & Rutter, W. J. (1985). Sequence of hepatitis B virus DNA incorporated into the genome of a human hepatoma cell line. *Journal of Virology* **53**, 885–892.

Received 30 July 1997; Accepted 16 October 1997