JOURNAL OF VIROLOGY, Sept. 1986, p. 731–734
0022-538X/86/090731-04$02.00/0
Copyright © 1986, American Society for Microbiology
Vol. 59, No. 3

Integration of Hepatitis B Virus DNA in Chromosome-Specific Satellite Sequences

YOSEF SHAUL,[1]* PABLO D. GARCIA,[2] STEVEN SCHONBERG,[3] AND WILLIAM J. RUTTER[2]

Department of Virology, The Weizmann Institute of Science, Rehovot 76100, Israel,[1] and Hormone Research Institute[2] and Department of Pediatrics,[3] University of California, San Francisco, California 94143

Received 13 February 1986/Accepted 27 May 1986

We previously reported the cloning and detailed analysis of the integrated hepatitis B virus sequences in a human hepatoma cell line. We report here the integration of at least one of hepatitis B virus at human satellite DNA sequences. The majority of the cellular sequences identified by this satellite DNA were organized as a multimeric composition of a 0.6-kilobase *Eco*RI fragment. This clone hybridized in situ almost exclusively to the centromeric heterochromatin of chromosomes 1 and 16 and to a lower extent to chromosome 2 and to the heterochromatic region of the Y chromosome. The immediate flanking host sequence appeared as a hierarchy of repeating units which were almost identical to a previously reported human satellite III DNA sequence.

On the basis of epidemiological studies, human hepatitis B virus (HBV) is considered to be a major cause of liver cancer (hepatocellular carcinoma; 28). Integration of HBV DNA was detected in primary liver tumors (4, 8, 22), in cell lines derived from human hepatomas (5, 7, 12), and in carriers (3, 4, 14). The most decisive evidence to date on the nature of the integrated HBV genomes was obtained from molecular cloning experiments that focused mainly on the PLC/PRF/5 (Alexander) cell line (1). This cell line contains about seven copies of the HBV genome, none of which are intact (24, 31). Although these clones do not display the extensive rearrangements documented previously for the integrated woodchuck hepatitis B virus genome (19), they nevertheless show a catalog of deletions, duplications, and inversions. On the basis of extensive mapping data and nucleotide sequence analysis of the virus host junction, integration was shown to be essentially random. However, junctions mapped within either DR1 or DR2 were shown in two independent liver tumors (11) and in some other cases (28). Thus, integration may occur with a statistical bias at the viral direct repeats. To date, the nature of host sequences neighboring the HBV integration site has not been studied. In a number of cases, cellular sequences next to integration sites or normal counterpart sites were reported (18, 30) but were not characterized in detail.

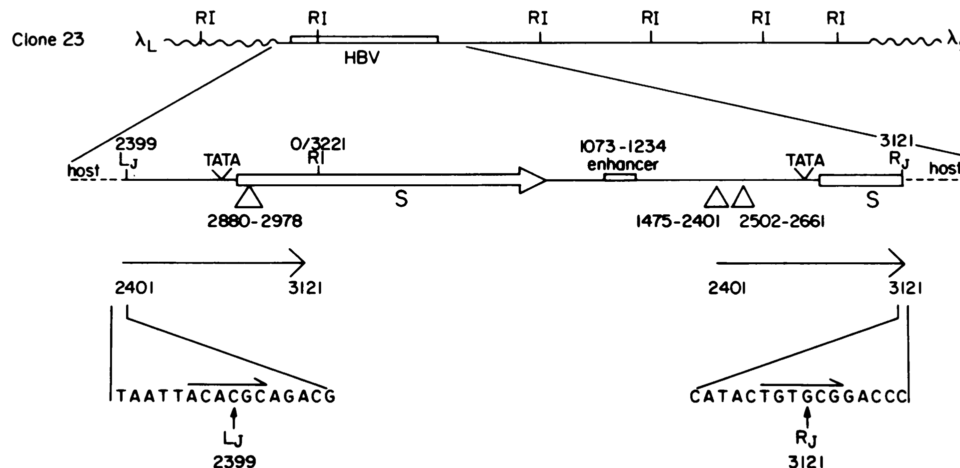

FIG. 1. Structure of clone A123 and its HBV DNA. The *Eco*RI map of the insert of A123 is shown at the top of the figure. The HBV sequence is boxed at the top of the figure. The positions of the viral TATA box and the enhancer element are shown. The sequences of the host viral function of the left (Lj) and the right (Rj) ends of the virus are shown at the bottom of the figure. The base pair numbers are from the HBV sequence of Valenzuela et al. (29). Symbols: λ_L〰, long arm of the phage; 〰λ_S, short arm of the phage; ----, flanking host sequences; △, HBV deleted region; ⇒ and ▭, surface and presurface antigen genes (S); →, region and direction of duplicate HBV sequences; ⇀, complementary repeated 6 nucleotides.

Clone A123 was isolated from a genomic library of the Alexander cell line. A detailed study of this clone, including electron microscopy analysis, restriction mapping, and sequence analysis, was published previously (24) and is summarized in Fig. 1. The unique feature of this integrated HBV sequence is the duplication of a portion of the viral sequences. The integrated viral DNA contains about 1 kilobase (from nucleotides 2401 to 3121) of direct repeats at its ends.

* Corresponding author.

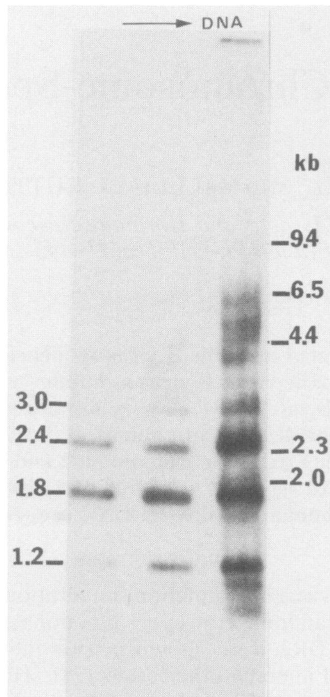

FIG. 2. Organization of genomic sequences homologous to A123 host DNA. DNA from blood cells of a normal human male was prepared as described previously (2) and was digested with *Eco*RI. Samples (4, 10, and 20 μg) were subjected to gel electrophoresis, blotting, and hybridization with A123 [^{32}P]DNA, as previously described (24). The numbers on the right represent kilobases of DNA molecular size. The sizes of the major bands are indicated in kilobases on the left of the figure.

A123 contains an intact S gene, including its promoter, next to the unique viral *Eco*RI site (6, 27). The pre-S region is incomplete and contains a deletion of about 150 base pairs (bp) (Fig. 1). The viral core gene is mostly deleted. The promoter region of the core gene is missing, but the viral enhancer element (23) is present. The sequence of the viral host junction is shown at both ends of the integrated virus DNA in Fig. 1. The left and right junctions share an exact complementary direct repeat of 6 bp.

Our analysis of seven different HBV fragments present in Alexander cells revealed that the HBV DNAs were integrated into the host genome at regions that contained highly repeated sequences (24). Most of the clones contained few copies of repeated sequences of the *Alu* family. Our preliminary analysis of the A123 host sequences resulted in a misinterpretation and led us to conclude that this clone contained *Alu* family repeats (24). By using a specific *Alu* repeat probe, along with a total human DNA probe, we found that the repetitive sequences of A123 were not of an *Alu* repeat type. To define the nature of the A123 repetitive sequences, we labeled the DNA of A123 by nick translation and hybridized it to a Southern blot of total human DNA digested with *Eco*RI enzyme. A number of distinct bands were observed, of which the 1.2-, 1.8-, and 2.4-kilobase fragments were dominant (Fig. 2). The exact determination of the sizes of the bands revealed that the bands were of a multimetric composition, 0.6 kilobases in size. These ladder pattern band are characteristic of satellite DNA in mammals (13, 17).

It is well documented that the majority of satellite DNAs are located near the centromeric region of chromosomes. We therefore performed in situ hybridization with chromosomal spreads from normal human cells and nick-translated A123 as a probe. A significant hybridization to the centromeric region of a number of chromosomes was detected after a 2-day exposure (Fig. 3). Clusters of grains were observed at the centromeric region of chromosomes 1 and 16 and to a lesser extent at that of chromosome 2 and at the heterochromatic region of the Y chromosome, indicating specific hybridization of A123 in a chromosome-specific manner to regions known to be rich in satellite DNA sequences. Of 82 chromosomal spreads that were analyzed, 44% of the clustered grains (two or more grains per site)

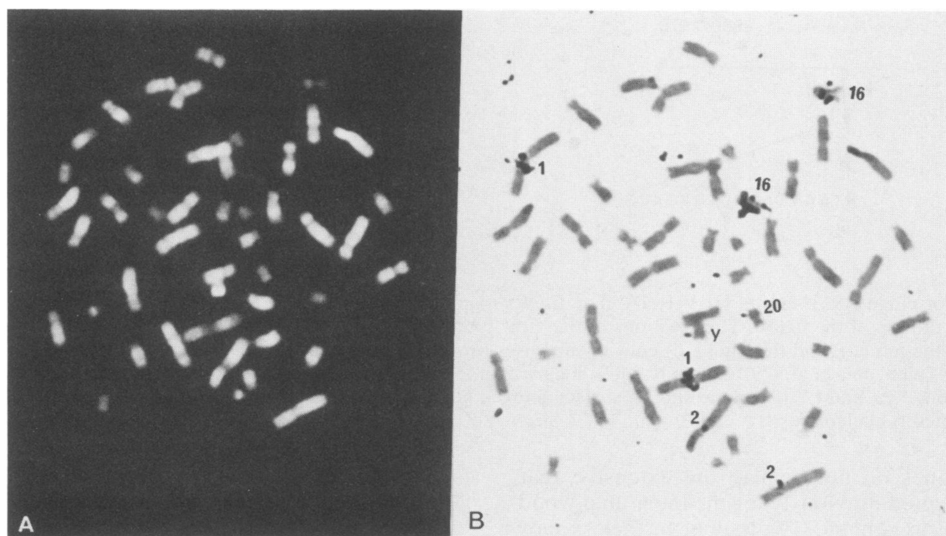

FIG. 3. In situ hybridization of ^{3}H-labeled A123 DNA to normal human chromosomes. ^{3}H-labeled nick-translated A123 DNA was hybridized in situ to metaphase chromosomes obtained from human lymphocytes of normal males. Chromosomes were identified before hybridization by quinacrine banding (A). The hybridization reaction was carried out by the method of Kirsch et al. (15). Slides were covered with nuclear track emulsion (NTB2; Kodak), incubated at 4°C for 24 to 48 h, and stained with Giemsa (B). The numbers of the relevant chromosomes are indicated.

were mapped on chromosome 1 at C(p11q12) region, 26% at 16C(p11q11) region, 5% at 2C(p11q11) region, and 5% at Yqh(q12) region.

In mammals, the sequences comprising the satellite DNA show common tandem short repeats. A series of these short units constitutes a longer repeating unit that is itself repeated in tandem with some variations. Thus, mammalian satellite DNAs are constructed from a hierarchy of repeating units (13, 25, 26). We analyzed about 800 bp of the host sequence at the right junction of A123 and observed a repetitive unit of 26 bases (Fig. 4). Within these 26 bases there is a duplication of CAT and CGAATGGAAT. The latter can be easily constructed from a duplication of GGAAT CGAAT, with subsequent mutations of C to G and of G to C, respectively. The 26 bases are almost identical to the clone pPD17 sequence derived from human satellite III DNA (10). However, a completely different pattern of host sequence is present at the left side of the junction (Fig. 4). We do not know whether this sequence is also satellite DNA; however, a repetition of TAAA or CAAA can be seen. To compare individuals for a possible polymorphic appearance of these satellite DNA sequences, we tested the DNA of eight normal persons. The presence of extra bands and differences in the levels of intensity of a few bands were noticed (data not shown). To analyze the possible evolutionary conservation of this repeat, we tested mouse, rat, Chinese hamster, and woodchuck DNA and were not able to detect any hybridization signals (data not shown).

The data emerging from several sequence analyses of integrated HBV virus and immediate host DNAs or counterpart sites in normal cells in a number of hepatoma cell lines are not compatible with the notion of specific sequences required for HBV integration. Whereas in one instance cellular sequences are inversely repeated (18), in other instances the host sequences remain intact with a duplication of 12 bp at the viral-host junction (30) or are deleted (16; Y. Shaul, unpublished data). In A123 there is a complementary, direct repeat of 6 bp at the site of integration, the significance of which is not clear. The HBV sequence from 2,404 to 3,121 bp is duplicated directly at the ends of the viral fragment, in a fashion similar to that of polyomavirus and simian virus 40. It was postulated that a head-to-tail recombination of the viral DNAs before integration would lead to such a situation (9, 20).

We conclude that the host sequence of A123 is of a satellite type for the following reasons. First, the sequence is pericentromeric. Second, the sequence appears as a hierarchy of repeating units from 3 to 5 bp to units of 26 bp and probably to units of 600 bp. The latter figure was deduced from Southern blot analysis. Thus, the sequence is almost identical to previously reported subtypes of the human satellite III DNA sequence (10). However, these repeating units are characteristic of the right-hand host DNA and not of the left-hand host, indicating possible cellular DNA rearrangements at the integration site. Alternatively, HBV was integrated at the border line of two different types of satellitelike sequences. It is also of interest to note that the host sequence of another clone, A18, is identical to that of A123 at the left-hand junction (31). Unfortunately, the right-hand junction of this clone was not included in the DNA isolated from the Alexander genomic library, so we do not know its relationship to the right-hand junction of A123. DNA restriction mapping and sequence analysis of HBV revealed that A123 and A18 are not identical (24, 31). However, it is possible that both clones were derived from duplication of an integrated HBV fragment with its immediate host sequence with subsequent mutations. Alternatively, although the possibility seems unlikely to us, this host sequence may provide a preferred target site for HBV integration. To the best of our knowledge, integration of viruses at satellite DNA has not previously been reported. Upon serial passage of simian virus 40 in permissive monkey cells in tissue cultures, a recombinant defective virus genome that contains a variety of cellular DNAs including satellite sequences is accumulated (21). However, it is not known whether this recombinant genome is the result of integration of the simian virus 40 genome into the host DNA with subsequent excision or whether it reflects a random recombination between viral genomes and cellular extrachromosomal DNA fragments.

```
_____ _____ ___________________ ___________________
C A T C A T C G A A T G G A A T C G A A T G G A A T
            _________ _________ _________ _________


C A T C A T C G A A T G G A A A T G A A A G G A G T
C A T C A T C T A A T G G A A T T G C A T G G A A T
C A T C A T A A A A T G G A A T C G A A T G G A A T
C A - C A T C A A A T G G A A T C A A A T G G A A T
C A T - - T G - A A C G G A A T T G A A T G G A A T
C G T C A T C G A A T - G A A T T G A A T G C A A T
C A T - - - C G A A T G G T C T C G A A T G G A A T
C A T C T T C A A A T G G A A T - G A A T - G A A T
C A C C A T A G A A T C G A A T C G A A T G G A A T
     G
T A T C A T C G A A T T G A A T C A A A T G G A A T
C A A C A T C A A A C G G A A A A A A A A C G G A A T
T A T - - - C G A A T G G A A T C G A A G A G A A T
C A T - - - C G A A T G G A C C C G A A T - G A A T
C A T - - - C T A A C G G A A T G G A A T G A A A T
                                          AT
C A T C A T C G A A T G G A A T C G A A T A G A A T
T A T - - - G G A A T G A A A T C C A G T G T G A T
C A T C A T C G A A T G G A C C C G A A T G G A A T
C A T - - - C T A A T G G A A T G G A A T G G A A T
A A T C - - C - - A T G G A C T C G A A T G C A A T
C A T C A T C G A A T G G A A T C G A A T G G A A T
C A T C A T G G A C T C G A A T G G A A T - - A A T
     G A
C A T - - T - G A A C G G A A T C G A A T G G A A T
C A T C A T C G G A T G G A A A T G A A T G G A A T
C A T C A T C G G A T G G A A A C G A A A G G A A T
A A T C A T T G A A C G G A A T C G A A T G G A A T
C A T - - - C G A A T G G A C T C G A A T G G A A T
C A T C A T C G A A T G G A A T C G A A T G G A A T
A A T A A T C C A C - G G A C C C G A A T G C A A T
A A T T A T C T A A T G G A A T G G A A T G G - - -
. . . . . . . .Rj - C G G A C C C G A A T G G - - -

A C - Lj. . . . . . . . . HBV . . . . . . . . . . .

A A A G A T C A A A T G C A T A A A G T A A T T A C
G C T A T G G A G G T T A A A G A A T T A A A C A T
A C A T C T T A T A T C A C A A T A A A A C A A A T
                    C T G T T C C G G A T C T
```

FIG. 4. The A123 host sequence of the immediate virus host junctions. The host sequences next to the left (Lj) and right (Rj) junctions are shown. The dotted region represents the integrated HBV sequences. The arrangement of the host cellular sequence next to the right junction of the virus is depicted in units of 26-base repeats. The bases between the lines are extra bases that do not match the repeat unit. The most abundant base in each column is written at the top of the figure to give the sequence of the representative repeat units.

LITERATURE CITED

1. Alexander, J. J., E. M. Bey, E. W. Geddes, and G. Lecatsas. 1976. Establishment of a continuously growing cell line from

primary carcinoma of the liver. S. Afr. Med. J. **50**:2124–2128.
2. **Bell, G. I., J. H. Karam, and W. J. Rutter.** 1981. Polymorphic DNA region adjacent to the 5' end of the human insulin gene. Proc. Natl. Acad. Sci. USA **78**:5759–5763.
3. **Brechot, C., M. Hadchouel, J. Scotto, F. Degos, P. Chatnay, C. Trepo, and P. Tiollais.** 1981. Detection of hepatitis B virus DNA in liver and serum: a direct appraisal of the chronic carrier state. Lancet **ii**:765–768.
4. **Brechot, C., M. Hadchouel, J. Scotto, M. Fonck, F. Potet, G. N. Vyas, and P. Tiollais.** 1981. State of hepatitis B virus DNA in hepatocytes of patients with hepatitis B surface antigen-positive and negative liver disease. Proc. Natl. Acad. Sci. USA **78**:3906–3910.
5. **Brechot, C., C. Pourcel, A. Louise, B. Rain, and P. Tiollais.** 1980. Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. Nature (London) **286**:533–535.
6. **Cattaneo, R., H. Will, N. Hernandez, and H. Schaller.** 1983. Signals regulating hepatitis B surface antigen transcription. Nature (London) **305**:336–338.
7. **Chaktaborty, P. R., N. Ruiz-Opazo, D. Shouval, and D. A. Shafritz.** 1980. Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular carcinoma cell line. Nature (London) **286**:531–533.
8. **Chen, D. S., B. H. Hoyer, J. Nelson, R. H. Purcell, and J. L. Getin.** 1982. Detection and properties of hepatitis B viral DNA in liver tissues from patients with hepatocellular carcinoma. Hepatology **2**:42S–46S.
9. **Clayton, C. E., and W. S. Rigby.** 1981. Cloning and characterization of the integrated viral DNA from three lines of SV40-transformed mouse cells. Cell **25**:547–559.
10. **Deininger, P. L., D. J. Jolly, C. M. Rubin, T. Friedman, and C. W. Schmidt.** 1981. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J. Mol. Biol. **151**:17–33.
11. **Dejean, A., P. Sonigo, S. Wain-Hobson, and P. Tiollais.** 1984. Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11 base-pair direct repeat. Proc. Natl. Acad. Sci. USA **81**:5340–5354.
12. **Edman, J. C., P. Gray, P. Valenzuela, L. R. Rall, and W. J. Rutter.** 1980. Integration of hepatitis B virus sequences and their expression in a human hepatoma cell. Nature **286**:535–538.
13. **Frommet, M., J. Prosser, D. Tkachuk, A. H. Reisner, and P. C. Vincent.** 1982. Simple repeated sequences in human satellite DNA. Nucleic Acids Res. **10**:545–563.
14. **Kam, W., L. B. Rall, E. A. Smuckler, R. Schmid, and W. J. Rutter.** 1982. Hepatitis B viral DNA in liver and serum of asymptomatic carriers. Proc. Natl. Acad. Sci. USA **79**:7522–7526.
15. **Kirsch, J. R., C. C. Morton, K. Nakahara, and P. Leder.** 1982. Human immunoglobin heavy chain gene map to a region of translocation in malignant B lymphocytes. Science **216**:301–303.
16. **Koshy, R., S. Koch, A. F. von Loringham, R. Kahmann, K. Murray, and P. H. Hefschneider.** 1983. Integration of hepatitis B virus DNA: evidence for integration in the single-stranded gap. Cell **34**:215–223.
17. **Mitchell, A. R., R. S. Beauchamp, and C. J. Bostock.** 1979. A study of sequence homologies in four satellite DNAs of man. J. Mol. Biol. **135**:127–149.
18. **Mizusawa, H., M. Taira, K. Yaginuma, M. Kobayashi, E. Yoshida, and K. Koike.** 1985. Inversely repeating integrated hepatitis B virus DNA and cellular flanking sequences in the human hepatoma-derived cell line huSP. Proc. Natl. Acad. Sci. USA **82**:208–212.
19. **Ogston, C. W., G. J. Jonak, C. E. Rogler, S. M. Astrin, and J. Summers.** 1982. Cloning and structural analysis of integrated woodchuck hepatitis virus sequences from hepatocellular carcinomas of woodchuck. Cell **29**:385–394.
20. **Sambrook, J., R. Green, J. Stringer, T. Mitchison, S. L. Hu, and M. Botchan.** 1980. Analysis of the site of integration of viral DNA sequences in rat cells transformed by adenovirus of SV40. Cold Spring Harbor Symp. Quant. Biol. **44**:569–584.
21. **Segal, S., M. Garner, M. E. Singer, and M. Rosenberg.** 1976. In situ hybridization of repetitive monkey genome sequences isolated from defective simian virus 40 DNA. Cell **9**:247–257.
22. **Shafritz, D. A., and M. C. Kew.** 1981. Identification of integrated hepatitis B virus DNA sequences in human hepatocellular carcinomas. Hepatology **1**:1–8.
23. **Shaul, Y., W. J. Rutter, and O. Laub.** 1985. A human hepatitis B viral enhancer element. EMBO J. **4**:427–430.
24. **Shaul, Y., M. Ziemer, P. D. Garcia, R. Crawford, H. Hsu, P. Valenzuela, and W. J. Rutter.** 1984. Cloning and analysis of integrated hepatitis virus sequences from a human hepatoma cell line. J. Virol. **51**:776–787.
25. **Smith, G. P.** 1976. Evolution of repeated DNA sequences by unequal crossover. Science **191**:528–535.
26. **Southern, E. M.** 1975. Long range perodicities in mouse satellite DNA. J. Mol. Biol. **94**:51–69.
27. **Standring, D. N., W. J. Rutter, H. E. Varmus, and D. Ganem.** 1984. Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. J. Virol. **50**:563–571.
28. **Tiollais, P., C. Pourcel, and A. Dejean.** 1985. The hepatitis B virus. Nature (London) **317**:489–495.
29. **Valenzuela, P., M. Quiroga, S. Zaldivat, P. Gray, and W. J. Rutter.** 1980. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes, p. 57–70. *In* B. Fields, R. Jaenisch, and C. Fox (ed.), Animal virus genetics. Academic Press Inc., New York.
30. **Yaginuma, K., M. Kobayashi, E. Yoshida, and K. Koike.** 1985. Hepatitis B virus integration in hepatocellular carcinoma DNA: duplication of cellular flanking sequences at the integration site. Proc. Natl. Acad. Sci. USA **82**:4458–4462.
31. **Ziemer, M., P. Garcia, Y. Shaul, and W. J. Rutter.** 1985. Sequence of hepatitis B virus DNA incorporated into the genome of a human hepatoma cell line. J. Virol. **53**:885–892.