VIROLOGY 238, 363–371 (1997)
ARTICLE NO. VY978863

Heterogeneity and Common Features of Defective Hepatitis B Virus Genomes

Stephan Günther,[1] Gunhild Sommer, Alicja Iwanska, and Hans Will

Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie, Universität Hamburg, 20251 Hamburg, Federal Republic of Germany

Received July 9, 1997; returned to author for revision August 12, 1997; accepted September 24, 1997

Defective hepatitis B virus (HBV) genomes derived from packaging and reverse transcription of spliced RNA pregenomes were reported to be associated with progression to chronic infection. Since only two types with similarly spliced regions were characterized so far we reasoned that additional ''spliced'' genome variants may exist. Therefore, we isolated a large number of defective HBV genomes from sera of seven chronic carriers by full-length PCR. Forty-eight were found to contain deletions caused by splicing as identified by cloning, subgenomic PCR, and sequencing. In total, 11 types of spliced genomes derived from excision of 10 different introns were present in various combinations in each serum. This diversity resulted from alternative usage of five splice donor and four acceptor sites present in most but not all HBV genotypes. All spliced genomes shared sequence elements essential for replication as well as for transcription of the pre-C and pregenome/C mRNAs and the X mRNA. Moreover, all contained the coding regions for the X protein and for precore/core or precore/core fusion proteins but lacked the pre-S/S gene promoters. These data demonstrate substantial and HBV genotype-dependent diversity of spliced genomes from which a variety of aberrant precore/core fusion proteins and normal X protein but no functional envelope and P proteins could be expressed. These genomes and the encoded proteins may play a role in the viral life cycle, persistence, and pathogenesis.

INTRODUCTION

Hepatitis B virus (HBV) infection causes a variety of liver diseases in humans. The clinical manifestations range from acute, fulminant, to chronic hepatitis; liver cirrhosis; and hepatocellular carcinoma. Most HBV genomes in virions consist of a small partially double-stranded circular DNA which is converted into covalently closed circular (ccc) molecules in nuclei upon infection of cells. The ccc HBV DNA serves as template for synthesis of a set of viral transcripts, namely a 3.5-kb pre-C mRNA for translation of precore/HBe protein, a 3.5-kb pregenome/C mRNA coding for nucleocapsid and P protein, 2.4- and 2.1-kb pre-S/S mRNAs coding for surface proteins (HBs, pre-S1, and pre-S2 proteins), and a 0.9-kb mRNA presumably coding for the X protein. All these transcripts are unspliced, have different 5′-ends, and share the same 3′-end located at the unique polyadenylation/processing site of the HBV genome. Only the pregenomic RNA which contains a functional encapsidation signal at its 5′-end is encapsidated into core particles and reverse transcribed into progeny HBV DNA (for a recent review see Ganem, 1996).

In addition to these unspliced transcripts, a variable fraction of spliced 3.5-kb pre-C and pregenome/C mRNAs (up to more than 50% of the corresponding unspliced RNAs) was identified in HBV DNA-transfected cell lines, in HBV-transgenic mice, and in HBV-infected liver (Chen *et al.,* 1989; Su *et al.,* 1989a,b; Suzuki *et al.,* 1989, 1990; Choo *et al.,* 1991; Kajino *et al.,* 1991; Wu *et al.,* 1991). The simultaneous synthesis of spliced and unspliced HBV transcripts is reminiscent of the situation in HIV. Furthermore, HBV RNAs have a posttranscriptional regulatory element (PRE) (Fig. 1) (Huang and Liang, 1993; Huang and Yen, 1994, 1995), which seems to be functionally similar to a HIV RNA sequence element, designated rev response element (Emerman *et al.,* 1989; Hadzopoulou-Cladaras *et al.,* 1989; Malim *et al.,* 1989). In HIV, and possibly also in HBV, the elements posttranscriptionally regulate the ratio of spliced to nonspliced mRNAs. It is therefore well conceivable that spliced RNAs of HBV play a role in the life cycle of the virus.

From the two major types of spliced HBV RNAs known so far, a single intron or two introns are removed (exon boundaries at positions 2447/489 and 2067/2350 + 2447/282, respectively; see Fig. 1) (Chen *et al.,* 1989; Su *et al.,* 1989a; Suzuki *et al.,* 1989; Choo *et al.,* 1991; Wu *et al.,* 1991). In addition, in two minor spliced transcripts the regions between positions 2985 and 489 and positions 2471 and 489 are removed (Fig. 1) (Suzuki *et al.,* 1990). From these four spliced pregenomic RNAs only two (2447/489 and 2471/489) were detected in the form of reverse-transcribed DNA in virions (Terre *et al.,* 1991). Thus, the number of functional splice sites could be lim-

[1] To whom reprint requests should be addressed. Fax: (0049) 40 48051 221.

0042-6822/97 $25.00

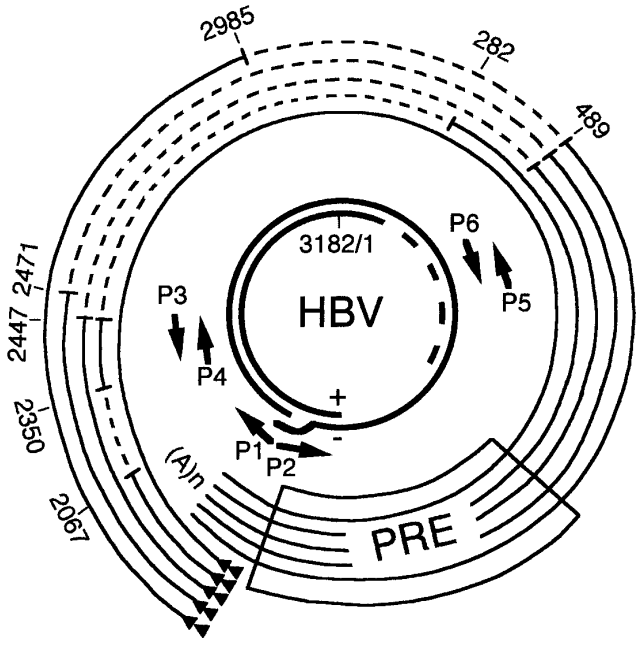


FIG. 1. HBV genome structure and spliced as well as unspliced pregenomic transcripts. Spliced regions (dashed lines), exon boundaries, and primers used for full-length and subgenomic PCRs are indicated. PRE, posttranscriptional regulatory element; (A)n, poly(A) tail.

ited and only a subfraction of spliced pregenomic RNAs may be encapsidated and reverse transcribed. Alternatively, there may be many more functional splice sites which have not been detected so far.

Splicing of the pregenomic RNA is not unique for HBV but was also observed for the related woodchuck and duck (DHBV) hepatitis viruses (Hantz *et al.,* 1992; Ogston and Razman, 1992; Obert *et al.,* 1996). Moreover, a spliced version of the DHBV pregenomic RNA was shown to be essential for productive infection of ducks but dispensable for virus replication when DHBV DNA was introduced into hepatoma cell lines by transfection (Obert *et al.,* 1996). Similarly, the major types of spliced pregenomic HBV RNAs are dispensable for virus replication when tested by transfection of cells with HBV DNA (Su *et al.,* 1989a; Wu *et al.,* 1991), but a function *in vivo* is not excluded. In fact, it was suggested that defective genomes derived from the major singly spliced RNA (2447/489) play a role in virus persistence because they were found predominantly in sera of patients with chronic infection and much less frequently in those with an acute self-limited course (Rosmorduc *et al.,* 1995). Moreover, these genomes were shown to express much higher levels of core and HBe proteins and of X gene transcripts than a wild-type genome in transfection experiments, which was speculated to be eventually important for immune recognition of the virus, viral persistence, and pathogenicity (Rosmorduc *et al.,* 1995). In this report we addressed the question whether additional "spliced" HBV genomes exist and whether or not they share common features with those variants proposed to play a role in viral persistence and pathogenicity. Therefore, we isolated and characterized a large number of such genomes from serum of seven chronically infected patients. Here, we demonstrate that the types of defective genomes derived from spliced pregenomic RNA are much greater than known so far and that most of them could only express X protein and aberrant core fusion proteins.

MATERIALS AND METHODS

Isolation of DNA from serum

Sera were from seven highly viremic patients with hepatitis B surface antigen-positive chronic hepatitis. From each serum sample, 300 μl was incubated at 65°C for 4 h in 20 mM Tris–HCl (pH 8.0), 10 mM EDTA, 0.1% SDS, and 0.8 mg/ml proteinase K. The DNA was then extracted with phenol and subsequently with chloroform and precipitated with ethanol using 20 μg tRNA as carrier. The pellet was dissolved in 20 μl H_2O.

Amplification and cloning of complete virion-encapsidated HBV genomes

Serum HBV DNA was amplified as described previously (Günther *et al.,* 1995) by PCR in 50 μl buffer containing 50 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTP, 0.01% gelatin, 5 U *Taq* DNA polymerase (Boehringer Mannheim, Germany), and 1 μM primer P1 and P2 (P1, CCGGAAAGCTTGAGCTCTTCTTTTTCACCTCTGCCTAATCA; P2, CCGGAAAGCTTGAGCTCTTCAAAAAGTTGCATGGTGCTGG; HBV homologous sequences span nucleotides 1821–1841 and 1823–1806, respectively; see Fig. 1 for position on the HBV genome; *Sst*I sites used for cloning are underlined). A 45-μl reaction premix was heated to 80°C and 5 μl enzyme was then added. The amplification was performed for 20 to 30 cycles, depending on the amount of template DNA, at 94°C for 40 s, 60°C for 1 min, and 72°C for 4 min with an increment of 5 s per cycle in a Perkin–Elmer thermocycler (48 wells). The PCR products were purified by phenol/chloroform (1/1) extraction and digested with *Sst*I for at least 6 h. The cleaved DNA was separated on an agarose gel; fragments between 1.0 and 3.0 kb were recovered from the gel and cloned into pUC19.

Analysis of cloned HBV DNA by PCR and sequencing

The cloned HBV genomes were PCR amplified as described above using 1 μl of bacterial glycerol stock culture and primers P1/P2. The resulting PCR products were diluted 1:50 and 1 μl thereof was used as template for amplification of subgenomic fragments in a 25-μl assay with 0.1 μM primer combinations P1/P3 (P3: 2400–2381,

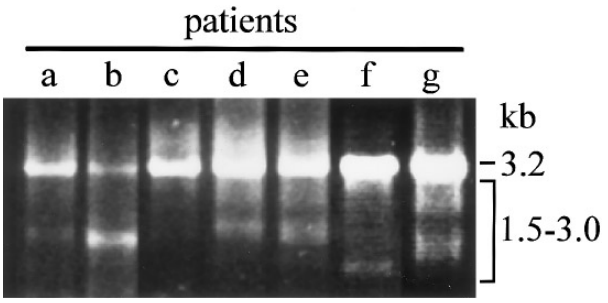


FIG. 2. Size analysis of HBV genomes present in serum from seven patients with chronic hepatitis B. HBV genomes were amplified in their entire length with primers P1/P2, separated in an agarose gel, and stained by ethidium bromide. The same pattern and relative amounts of short genomes were found when the PCR was repeated with the same samples (not shown).

CTTCGTCTGCGAGGCGAGGG), P4/P5 (P4: 2357–2380, GGCAGGTCCCCTAGAAGAAGAACT; P5: 738–716, ATAACTGAAAGCCAAACAGTGGG), and P6/P2 (P6: 634–656, ATTCCTATGGGAGTGGGCCTCAG). The positions of the primers are indicated in Fig. 1. The plasmid DNA of 34 representative clones was prepared and the HBV DNA insert was sequenced with vector-specific reverse and forward primers and primer P4 using the SequiTherm Long-Read Cycle Sequencing Kit (Epicentre Technologies, Madison, WI) and an automated sequencer (LI-COR, Lincoln, NE).

RESULTS

Amplification of full-length HBV genomes from sera

HBV DNA was isolated from sera of seven chronically HBV-infected patients and amplified by PCR. In order to obtain the entire length of the HBV genomes, primers annealing at the very 3′- and 5′-ends of the HBV DNA minus strand were used as described previously (Fig. 1) (Günther *et al.,* 1995). After separation of the PCR products in an ethidium bromide-stained agarose gel, a major band corresponding to HBV genomes of 3.2-kb wild-type length was visible in all samples (Fig. 2). In addition, DNA with faster electrophoretic mobility corresponding to molecules 1.5 to 3.2 kb in length was seen in all but one sample, suggesting the presence of HBV genomes containing large deletions (Fig. 2). Such genomes shorter than full-length HBV were not observed after amplification of cloned wild-type HBV DNA (data not shown), indicating that they occur *in vivo* and are not artificially created by PCR. In samples a, b, and f the HBV "deletion" variants showed up as rather distinct bands, indicating dominant size classes of defective genomes (Fig. 2). In contrast, a diffuse banding pattern was observed in samples d, e, and g, suggesting the presence of a much more heterogeneous population of such genomes. Furthermore, the length of the short PCR products as well as the ratio between the 3.2-kb and the shorter PCR products differed between the samples. Thus, each serum sample seems to contain a characteristic set and amount of short genomes. The short genomes always constituted only a small fraction of the virus population as evident from the lower intensity of the corresponding bands and their more efficient amplification compared to wild-type genomes. The latter was experimentally demonstrated by amplification of a mixture of HBV DNAs 1.2 and 3.2 kb in length. The 1.2-kb HBV genomes were amplified 10 times more efficiently than the 3.2-kb genomes in a 30-cycle PCR with *Taq* polymerase (data not shown). Therefore, even in sample b, which showed a 1:1 ratio of full-length to short PCR fragments (Fig. 2), the latter represent only a minor HBV genome subpopulation.

Cloning and PCR analysis of defective HBV genomes

For the analysis of the structure of the short HBV genomes, the amplified DNAs in the 1.0- to 3.0-kb length range were recovered from the gel and cloned into vector pUC19. Recombinants were obtained from all serum samples and 55 plasmids were randomly chosen for a further structural analysis. The approximate sizes and locations of deletions were first roughly mapped by three overlapping subgenomic PCRs covering the whole genome and agarose gel analysis of the PCR fragments. The primers chosen for this screening hybridized outside of the intron sequences of the previously described singly and doubly spliced pregenomic RNAs (Fig. 1) because we expected that a large fraction of molecules is derived from spliced pregenomic RNA. This subgenomic PCR screen revealed various large deletions between nucleotide positions 2357 and 738 in HBV genomes of most of the plasmids (for representative examples see Fig. 3), in some cases with additional adjacent deletions in the C gene. The PCR analysis of individual genomes from sample e, for example, even revealed five different types of deletions located between positions 2357 and 738 (Fig. 3, patient e, lanes 1 to 5). Deletions identical in length in some genomes from one as well as from different patients were also found (Fig. 3, patient d, lanes 1–5, 7, and 9; patient e, lanes 1, 6, and 7). In a few cases, no DNA fragments were amplified with the primers used, indicating deletions or mutations of the primer binding sites in these genomes as experimentally confirmed below. In conclusion, the PCR analysis revealed the presence of one or more large internal deletions between the 5′-end of the C gene and the 3′-end of the S gene and that the population of defective genomes can be highly heterogeneous in the serum of an individual.

Detailed structural analysis of defective HBV genomes

The precise structures of the defective genomes were determined by sequencing. To avoid redundant sequenc-

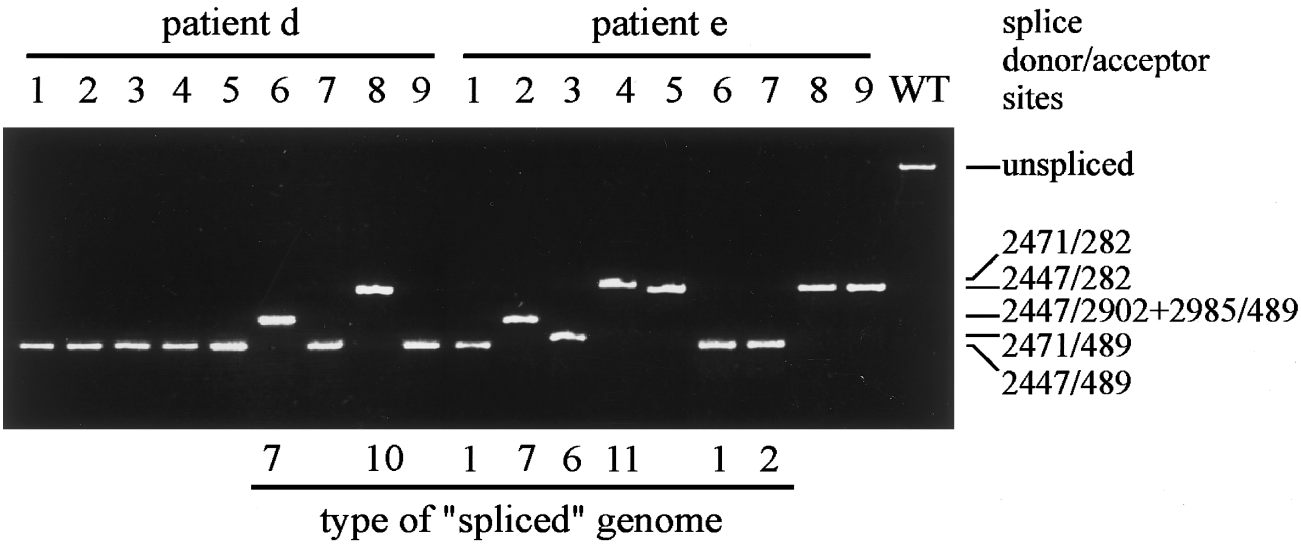


FIG. 3. Size analysis of deletions in cloned defective HBV genomes derived from patients d and e. Subgenomic fragments of nine cloned defective genomes from each patient were amplified using primer combination P4/P5 spanning nucleotides 2357 to 738 (Fig. 1) and were separated in an agarose gel and stained by ethidium bromide. WT, cloned wild-type HBV DNA (Galibert *et al.,* 1979) was used as a template for PCR. The splice donor and acceptor sites which delimit the deletions are indicated. The "splice" types of those cloned genomes whose sequences have been determined (see Fig. 5) are indicated below the lanes.

ing of identical types of genomes, the defective genomes from each serum sample were first classified according to position and length of the deletions by subgenomic PCR analysis as described above. One or two genomes representative of each genome class from each serum were sequenced completely (in total 34 of the 55 cloned genomes). In 27 of the 34 sequenced genomes, representative of 48 of the 55 cloned defective genomes, splice donor/acceptor site-specific sequences were identified at the deletion boundaries, indicating that the deleted sequences correspond to introns removed from pregenomic RNA (Fig. 4). In total, 11 different types of defective genomes derived from spliced pregenomic RNA were found (Fig. 5). Although these 11 types differ drastically in structure, they share some common features. In particular, all defective genomes share sequences from nucleotide position 489 to position 2067 (Fig. 5, bottom). This region contains all known signals required for conventional replication (DR1 and DR2), synthesis and processing of pregenomic/C-mRNA, pre-C mRNA, and X mRNA (C and X gene promoter, enhancer I and II, polyadenylation/processing site), and pregenomic RNA encapsidation (Fig. 5, bottom). The presence of the RNA encapsidation signal and of the direct repeat sequences is consistent with the isolation of the defective HBV genomes from serum-derived viral particles. Furthermore, common to all genomes is the presence of the complete X protein coding region and the complete C ORF or truncated versions thereof fused to other ORFs (Fig. 5). All genomes lack sequences in two regions (positions 2471 to 2902 and 2985 to 282), both coding for parts of the P and pre-S/S proteins and containing the complete pre-S1 (region 1) and pre-S2/S (region 2) promoters (Fig. 5, bottom).

The structures of the defective genomes indicate that 10 different intronic sequences are removed from the corresponding RNA pregenomes. The diversity in structure is caused by alternative usage of five different splice donor and four splice acceptor sites (Fig. 4). Another reason is the removal of more than one of these introns as evident in genomes 2, 4, 7, and 10, with two introns removed, and in genome 8, lacking three intronic sequences (Fig. 5). The number and types of genomes identified in our study indicate that the various splicing sites differ also quantitatively in their function. The most frequently used splice donor site was at position 2447 (in 83% of the spliced genomes), approximately one-third as frequent was that at position 2067, and the remaining three at positions 2087, 2471, and 2985 were rarely functional (Fig. 4). A similar analysis for the four acceptor sites revealed most frequent splicing at position 489 (in 79% of the spliced genomes). Acceptor sites at positions 282 and 2350 were involved in every fifth spliced genome and splicing at position 2902 was 10 times less frequent than at the most prevalent one (Fig. 4).

In view of the quantitative differences in splice site usage we wondered whether this depends on the degree of homology of the HBV splice sites with the splice consensus sequences or on genotype-specific sequence variation within these sites. This analysis revealed a similar degree of homology of the different HBV splice sites with the consensus sequences (Fig. 4). From the minor

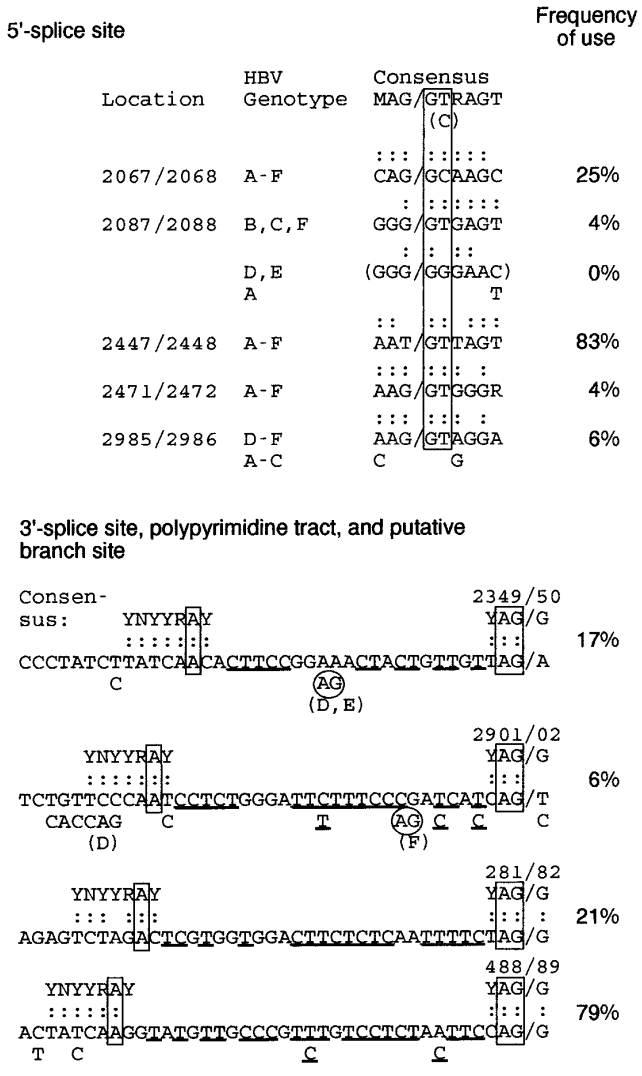


FIG. 4. Putative 5′- and 3′-splice signals in HBV used to generate the spliced genomes shown in Fig. 5 and their conservation in different HBV genotypes. The sequences at the deletion junctions were aligned to the consensus sequences for the 5′-splice site (MAG/GTRAGT; a C nucleotide shown in parentheses at position +2 is compatible with function but occurs rarely), the 3′-splice site (YAG/G), and the branch site (YNYYRAY) (Mount, 1982; Zeitlin and Efstratiadis, 1984; Jackson, 1991). Invariable nucleotide positions are boxed. Only one possible branch site which most perfectly matches the consensus is indicated. Pyrimidine residues forming the polypyrimidine tract located between the putative branch site and the 3′-splice site are underlined. HBV genotype-specific nucleotide changes in splice signals are indicated below the sequences and the corresponding genotype is given in parentheses. Genotype-specific AGs are also indicated (encircled) because they may serve as alternative 3′-splice sites or inhibit splicing as reported in other systems (Smith *et al.,* 1989). The frequency of use of a specific splice site was calculated by dividing the number of genomes spliced at this site by the total number of genomes analyzed ($n = 48$).

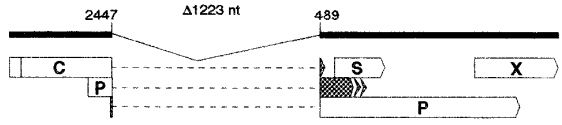
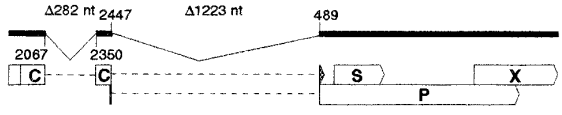
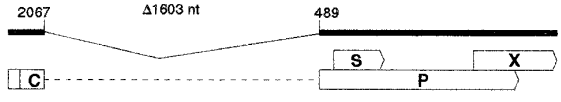
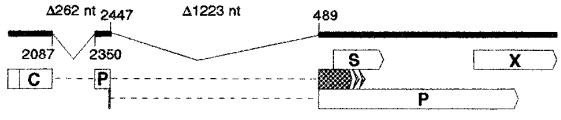
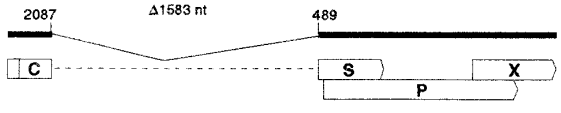
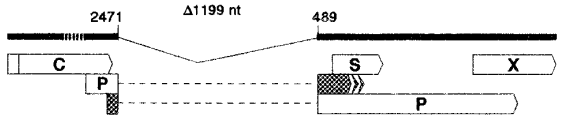
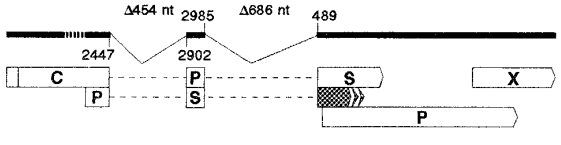
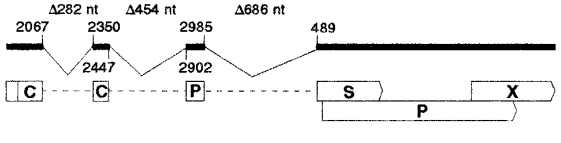
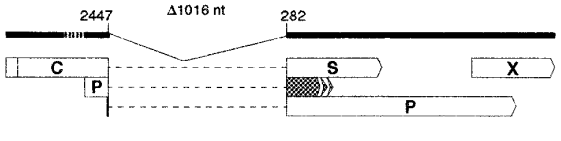
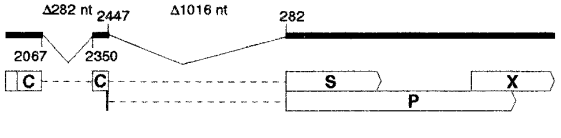
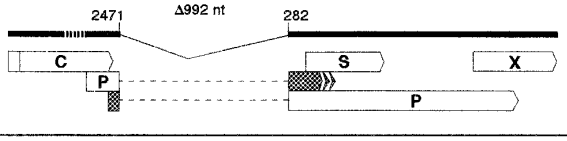
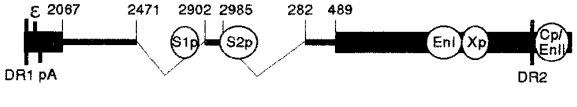
differences in homology no association with the frequency of splicing at a specific site can be inferred. Interestingly, a sequence data bank search uncovered genotype-specific sequence variation within the splice signals (Fig. 4). With one exception, all of these changes are consistent with the splice consensus sequences, but an influence of some of them on recognition by the splicing apparatus is conceivable. The exception concerns the splice donor site at position 2087. In HBV genotypes A, D, and E there is a G residue instead of a T or C in the invariable +2 position of the splice consensus sequence which should inactivate this splice donor site in these genomes. This is consistent with our observation that spliced genomes generated via this splice site did not belong to any of these genotypes but were exclusively of genotype B.

DISCUSSION

In this study we provide evidence for the existence of a large variety of defective HBV genomes derived from spliced pregenomic RNA (HBV splice variants). Sequence analysis revealed four donor and five acceptor splice sites used in different combinations and frequencies depending in part on the HBV genotype. The splice variants represented a minor population compared to wild-type genomes, and their relative amounts and types seem to differ from patient to patient. Common to all the splice variants was the presence of the X gene and all known regulatory sequences required for pre-C and pregenome/C mRNA synthesis and genome replication. In contrast, all lacked the pre-S/S gene promoter regions and contiguous coding regions for the envelope and P proteins. The C gene was fused in most genomes with other reading frames. These findings indicate a more complex splicing of pregenomic HBV RNA than known so far and may imply a role of the corresponding splice variant genomes and expressed proteins in the viral life cycle, viral persistence, and pathogenesis.

One interesting finding of our study is the high degree of inter- as well as intrapatient structural variability of defective spliced genomes which only became obvious because of amplification of entire HBV genomes. The data obtained extend previous investigations based on subgenomic PCRs which led to the identification of "singly spliced" genomes of types 1 and 6 in serum but failed to detect other spliced genomes (Terre *et al.,* 1991; Rosmorduc *et al.,* 1995). In agreement with these reports, we also found splice variants in different relative amounts in individual samples and only as minor populations consisting predominantly of defective genomes of type 1. The latter accounted for 50% of all splice variant genomes analyzed and were present in sera from five of seven patients. All other splice variant genomes were much less prevalent and this may be one reason why they have not been detected before.

Previously described splice sites and the HBV genotype-specific splice donor site 2087 identified in our study sum up to five donor and four acceptor sites used for

Type		Patients	Clones analyzed by: Sequencing	PCR	HBV genotypes
1		a,c,d,e,f	7	24	A,D
2		b,e,g	3	3	A,B
3		f,g	5	5	A,D
4		b	1	1	B
5		b	1	1	B
6		e	1	1	A
7		d,e	2	2	A
8		g	1	1	A
9		e	2	6	A
10		b,d	3	3	A,B
11		e	1	1	A
1-11		a-g	27	48	A,B,D

splicing of pregenomic RNA. The donor sites 2067 and 2447 and the acceptor sites 2350, 282, and 489 are frequently involved in pregenomic RNA splicing (Chen *et al.,* 1989; Su *et al.,* 1989a; Suzuki *et al.,* 1989; Choo *et al.,* 1991; Wu *et al.,* 1991), whereas splicing was rarely detected at the donor site 2985 and occurred only after inactivation of the 2447 donor site at a cryptic site at position 2471 (Suzuki *et al.,* 1990). Based on the deletion of nucleotides 2448 to 2901 in chromosomally integrated HBV DNA from the human hepatoma cell line PLC/PRF/5 (Ziemer *et al.,* 1985), the existence of a functional donor site at position 2902 was so far only suggested (Terre *et al.,* 1991). Our study demonstrates the function of the splice site at position 2902 and suggests moreover that the HBV DNA integrated in these cells arose from infection with a corresponding splice variant. Thus, splice variant genomes may also play a role in hepatocarcinogenesis. Strikingly, most of the splice sites which gave rise to the defective genomes described here were previously identified on the RNA level, but in only 3 of the 11 splicing patterns, namely types 1, 6, and 10. The splicing patterns 1, 10, and 6 correspond to the most prevalent singly and doubly spliced pregenomic RNAs (Chen *et al.,* 1989; Su *et al.,* 1989a; Suzuki *et al.,* 1989; Choo *et al.,* 1991; Wu *et al.,* 1991) and to the rare variant in which splicing occurs at the cryptic site at position 2471, respectively (see Fig. 1) (Suzuki *et al.,* 1990).

It is not clear why the efficiency and type of splicing differs from patient to patient. One possible reason could be sequence variations which create, destroy, or modulate the function of splice signal sequences. Four sequence elements on the RNA are important in the splicing process: the 5′-splice site (consensus: MAG/GTRAGT), the 3′-splice site (consensus: YAG/G), and the branch site (consensus: YNYYRAY) 20–50 nucleotides upstream thereof, as well as a polypyrimidine tract between the 3′-splice and the branch site (Mount, 1982; Zeitlin and Efstratiadis, 1984). All HBV splice donor and acceptor sites identified in our study are in close agreement with these consensus sequences. However, we noted that they differ in some HBV genotypes which may have quantitative and qualitative effects on the splicing of the genotype-specific RNA pregenomes. This speculation needs to be experimentally examined because the sequences involved in splice site recognition are not highly conserved. A notable exception is a genotype-specific T to G mutation in position +2 of the splice donor site 2087 which is not compatible with splice site recognition (Mount, 1982; Jackson, 1991). Mutations outside of splice site recognition sequences, such as the deletions in the C gene of some of the genomes studied or the HBV genotype-specific insertions close to the 2350 and 2902 acceptor sites, may also influence splice site selection by altering the RNA structure. Provided splicing in HBV is functionally important, the presumed HBV genotype-specific variation in splicing and the complex pregenomic RNA splicing patterns observed may imply that some splice sites are functionally redundant, e.g., they could have the same effects. In this case, knockout of specific sites would be tolerable. This conclusion does not exclude an essential function of a subset of splice sites for some aspects of infection. Consistent with the latter speculation is the use of the splice acceptor sites 282 and 489 in all genomes. These two sites may be functionally important for the HBV life cycle.

An intriguing observation is the apparent absence of splice sites downstream of position 489 and upstream of position 2067. This may be due to the lack of such splicing sites, to too low sensitivity of our assays, or to the fact that we have analyzed only genomes greater than 1.0 kb. Since very small HBV genomes could be generated if splicing started in the C gene and extended up to the X gene region, we analyzed in an independent study amplified HBV DNA fragments smaller than 1.0 kb by cloning and sequencing of PCR fragments. In these experiments, defective genomes with extremely large deletions were found but none was generated by splicing (G.S., unpublished observation). Therefore, we currently believe that there are no splice sites downstream of position 486 and upstream of 2067.

Taking our and previously published data (Chen *et al.,* 1989; Su *et al.,* 1989a; Suzuki *et al.,* 1989, 1990; Choo *et al.,* 1991; Terre *et al.,* 1991; Wu *et al.,* 1991) together it is evident that both the pre-C and the C mRNAs can be extensively spliced. As a consequence a large number of aberrant precore and core fusion proteins are presumably expressed from these RNAs. Some of these proteins could negatively interfere with the viral assembly process as shown for artificial core fusion proteins (Scaglioni *et al.,* 1994, 1996; von Weizsäcker *et al.,* 1996). Such defec-

FIG. 5. Types of defective HBV genomes classified according to the splice pattern. Deletions generated by splicing are indicated by thin lines, whereas in-frame deletions in the C gene which do not result from splicing are shown as a dashed line. Nucleotide positions at the exon boundary and the lengths of deletions calculated on the basis of the genotype D wild-type genome (Galibert *et al.,* 1979) with numbering starting in the middle of the *Eco*RI site are given. *Note.* Positions or lengths are different in other genotypes due to genotype-specific insertions. Conventional and novel open reading frames (ORFs) present in each type of genome are shown and capital letters indicate the origin of each part of these ORFs with respect to the conventional HBV ORFs (C, P, S, and X as shown on top), whereas nonconventional ORF parts are filled. The nonconventional 3′-ORF parts in types 1, 4, 6, 7, 9, and 11 terminate in a genotype-specific manner at heterogeneous positions as indicated by multiple arrows at the end of these frames. HBV sequences and regulatory DNA and RNA elements shared by (black boxes) and lacking in all (interruptions) spliced genomes are schematically shown at the bottom line. DR1 and DR2, direct repeat sequences; S1p and S2p, S gene promoters; Xp, X gene promoter; EnI, enhancer I; Cp/EnII, C gene promoter/enhancer II; pA, processing/polyadenylation signal; ϵ, RNA encapsidation signal.

tive proteins are predictably abundant in the cell relative to the nondefective proteins since a large fraction and sometimes even most of the pre-C and C mRNAs are spliced in infected liver and in HBV DNA-transfected cells (Chen *et al.,* 1989; Su *et al.,* 1989a,b; Suzuki *et al.,* 1989, 1990; Kajino *et al.,* 1991; Wu *et al.,* 1991). Their predicted abundance makes them ideal candidates for acting as inhibitors of virus multiplication similarly to defective interfering particles (Holland, 1990). A carboxy-terminal P (fusion) protein with RNase H activity could eventually be translated from all spliced pregenomic/C mRNAs provided that translational scanning occurs as shown for the P ORF in wild-type C mRNA (Fouillot *et al.,* 1993).

It is also conceivable that the splice sites in the pre-C and pregenomic/C mRNAs facilitate the export of both their spliced and their unspliced versions. Since splicing seems to be a prerequisite for efficient nuclear export (Izaurralde and Mattaj, 1992), it could be that recognition of the splice sequences, but not necessarily completion of splicing, initiates nuclear transport of the HBV mRNAs into the cytoplasm. However, the observation that inactivation of the HBV splice signals does not reduce the level of the unspliced transcripts (Su *et al.,* 1989a; Wu *et al.,* 1991) argues against this possibility. In addition, the PRE present on both RNAs (see Fig. 1) can alleviate the need for introns as mediators of efficient RNA export from the nucleus to the cytoplasm (Huang and Yen, 1994).

The spliced pregenomic RNAs may not have a direct function but may have an indirect one after they are reverse transcribed into DNA in viral particles. Upon reentry into the nucleus or infection of a new cell, the defective genomes can serve as templates for HBV gene expression provided their genome structure is compatible with nuclear transport and conversion into ccc HBV DNA. Due to the fact that all types of spliced genomes share the complete X gene, including its promoter region, the enhancers I and II, and the polyadenylation signal, splice variants should be able to synthesize X mRNA. This was already demonstrated for the major splice variant genome which expresses much more abundant X mRNA than do wild-type genomes (Rosmorduc *et al.,* 1995). The same may be true for most of the additional spliced genomes identified in our study. One argument for this speculation is based on the observation that the S gene promoter sequences are deleted in all splice variant genomes. This could result in a stronger activation function of enhancers I and II on the C and X gene promoters. The reported activation of the X gene promoter by a factor of 3 upon deletion of all sequences upstream of enhancer I (Zhang *et al.,* 1992) is consistent with this speculation. If one further takes into account that spliced genomes can be the dominant genome population within infected liver cells (Terre *et al.,* 1991), it is well conceivable that splice variants rather than the wild-type virus are the major source for X protein expression. If splice variants can synthesize ccc DNA and a corresponding pregenome-like RNA, they could presumably replicate indefinitely when supplemented *in trans* with the essential virus proteins from wild-type virus. This may contribute to their accumulation and, consequently, to enhanced expression of X protein and of aberrant core proteins. The data presented and the clones obtained in our study provide the basis to prove or disprove these predictions as well as the possible role of the splice variants in the viral life cycle, persistence, and pathogenesis.

ACKNOWLEDGMENTS

We greatly appreciate the critical reading of the manuscript by P. Forster. This work was supported by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (Verbundvorhaben, Project 01KI9560) and the Wilhelm-Sander-Stiftung. The Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie is supported by the Freie und Hansestadt Hamburg.

REFERENCES

Chen, P. J., Chen, C. R., Sung, J. L., and Chen, D. S. (1989). Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. *J. Virol.* 63, 4165–4171.

Choo, K. B., Liew, L. N., Chong, K. Y., Lu, R. H., and Cheng, W. T. (1991). Transgenome transcription and replication in the liver and extrahepatic tissues of a human hepatitis B virus transgenic mouse. *Virology* 182, 785–792.

Emerman, M., Vazeux, R., and Peden, K. (1989). The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization. *Cell* 57, 1155–1165.

Fouillot, N., Tlouzeau, S., Rossignol, J.-M., and Jean-Jean, O. (1993). Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation. *J. Virol.* 67, 4886–4895.

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

Ganem, D. (1996). Hepadnaviridae and their replication. *In* "Fields Virology," 3rd ed. (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), pp. 2703–2737. Lippincott-Raven, Philadelphia.

Günther, S., Li, B.-C., Miska, S., Krüger, D. H., Meisel, H., and Will, H. (1995). A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J. Virol.* 69, 5437–5444.

Hadzopoulou-Cladaras, M., Felber, B. K., Cladaras, C., Athanassopoulos, A., Tse, A., and Pavlakis, G. N. (1989). The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J. Virol.* 63, 1265–1274.

Hantz, O., Baginski, I., Fourel, I., Chemin, I., and Trepo, C. (1992). Viral spliced RNA are produced, encapsidated, and reverse transcribed during *in vivo* woodchuck hepatitis virus infection. *Virology* 190, 193–200.

Holland, J. J. (1990). Defective viral genomes. *In* "Virology" (B. N. Fields and D. M. Knipe, Eds.), pp. 151–165. Raven Press, New York.

Huang, J., and Liang, T. J. (1993). A novel hepatitis B virus (HBV) genetic element with Rev response element-like properties that is essential for expression of HBV gene products. *Mol. Cell. Biol.* 13, 7476–7486.

Huang, Z. M., and Yen, T. S. (1994). Hepatitis B virus RNA element that facilitates accumulation of surface gene transcripts in the cytoplasm. *J. Virol.* 68, 3193–3199.

Huang, Z. M., and Yen, T. S. (1995). Role of the hepatitis B virus posttran-

scriptional regulatory element in export of intronless transcripts. *Mol. Cell. Biol.* 15, 3864–3869.

Izaurralde, E., and Mattaj, I. W. (1992). Transport of RNA between nucleus and cytoplasm. *Semin. Cell. Biol.* 3, 279–288.

Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* 19, 3795–3798.

Kajino, K., Ohta, Y., Miyamura, T., and Saito, I. (1991). Efficient cloning of hepatitis B virus DNA from single-stranded replicative intermediates and its application to S1 mapping of viral RNA in human liver tissue. *J. Med. Virol.* 33, 33–38.

Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V., and Cullen, B. R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 338, 254–257.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* 10, 459–472.

Obert, S., Zachmann-Brand, B., Deindl, E., Tucker, W., Bartenschlager, R., and Schaller, H. (1996). A spliced hepadnavirus RNA that is essential for virus replication. *EMBO J.* 15, 2565–2574.

Ogston, C. W., and Razman, D. G. (1992). Spliced RNA of woodchuck hepatitis virus. *Virology* 189, 245–252.

Rosmorduc, O., Petit, M. A., Pol, S., Capel, F., Bortolotti, F., Berthelot, P., Brechot, C., and Kremsdorf, D. (1995). In vivo and in vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. *Hepatology* 22, 10–19.

Scaglioni, P., Melegari, M., Takahashi, M., Chowdhury, J. R., and Wands, J. (1996). Use of dominant negative mutants of the hepadnaviral core protein as antiviral agents. *Hepatology* 24, 1010–1017.

Scaglioni, P. P., Melegari, M., and Wands, J. R. (1994). Characterization of hepatitis B virus core mutants that inhibit viral replication. *Virology* 205, 112–120.

Smith, C. W., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. (1989). Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. *Nature* 342, 243–247.

Su, T. S., Lai, C. J., Huang, J. L., Lin, L. H., Yauk, Y. K., Chang, C. M., Lo, S. J., and Han, S. H. (1989a). Hepatitis B virus transcript produced by RNA splicing. *J. Virol.* 63, 4011–4018.

Su, T. S., Lui, W. Y., Lin, L. H., Han, S. H., and P'eng, F. K. (1989b). Analysis of hepatitis B virus transcripts in infected human livers. *Hepatology* 9, 180–185.

Suzuki, T., Kajino, K., Masui, N., Saito, I., and Miyamura, T. (1990). Alternative splicing of hepatitis B virus RNAs in HepG2 cells transfected with the viral DNA. *Virology* 179, 881–885.

Suzuki, T., Masui, N., Kajino, K., Saito, I., and Miyamura, T. (1989). Detection and mapping of spliced RNA from a human hepatoma cell line transfected with the hepatitis B virus genome. *Proc. Natl. Acad. Sci. USA* 86, 8422–8426.

Terre, S., Petit, M. A., and Brechot, C. (1991). Defective hepatitis B virus particles are generated by packaging and reverse transcription of spliced viral RNAs in vivo. *J. Virol.* 65, 5539–5543.

von Weizsäcker, F., Wieland, S., and Blum, H. E. (1996). Inhibition of viral replication by genetically engineered mutants of the duck hepatitis B virus core protein. *Hepatology* 24, 294–299.

Wu, H. L., Chen, P. J., Tu, S. J., Lin, M. H., Lai, M. Y., and Chen, D. S. (1991). Characterization and genetic analysis of alternatively spliced transcripts of hepatitis B virus in infected human liver tissues and transfected HepG2 cells. *J. Virol.* 65, 1680–1686.

Zeitlin, S., and Efstratiadis, A. (1984). In vivo splicing products of the rabbit beta-globin pre-mRNA. *Cell* 39, 589–602.

Zhang, P., Raney, A. K., and McLachlan, A. (1992). Characterization of the hepatitis B virus X- and nucleocapsid gene transcriptional regulatory elements. *Virology* 191, 31–41.

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