VIROLOGY 220, 256–262 (1996)
ARTICLE NO. 0311

SHORT COMMUNICATION

Natural Populations of Woodchuck Hepatitis Virus Contain Variant Precore and Core Sequences Including a Premature Stop Codon in the Epsilon Motif

DANIEL H. LI,* JOHN E. NEWBOLD,† and JOHN M. CULLEN*,[1]

*Department of Microbiology, Pathology and Parasitology, College of Veterinary Medicine, North Carolina State University, 4700 Hillsborough Street, Raleigh, North Carolina 27606; and †Department of Microbiology and Immunology, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599

Received February 26, 1996; accepted March 27, 1996

We have determined a consensus sequence and the type and the frequency of spontaneous sequence variations in the woodchuck hepatitis virus (WHV) precore gene and the 5′ region of the core gene in 101 serum samples from 53 naturally WHV-infected woodchucks by polymerase chain reaction sequencing. Twenty of the 53 woodchucks were found to have variant sequences. Ten patterns of variant sequences were identified in these 20 animals. WHV sequences from 4 woodchucks had 1 nucleotide change, 3 had 2 nucleotide changes and 3 had 3 nucleotide changes. The nucleotide changes were not randomly distributed, but were limited to only 8 sites. Four sites were in the epsilon motif of the precore gene and four were in the 5′ region of the core gene. Sixteen of the 53 (30%) woodchucks had precore sequence variants. All altered sites were analogous to previously described mutations in hepatitis B virus. There was a nucleotide change at nucleotide 2016 in codon 29 of the precore region that produced a stop codon in 4 animals. This site is analogous to a common hepatitis B virus e antigen mutation. The sequence from the initial blood samples from 3 of 4 animals with this stop codon producing variant appeared to be the consensus sequence; however, in later samples the variant occurred as a mixed infection with the consensus sequence. The mixed infections were chronic and the proportion of the variant sequence was maintained or increased in the course of infection. In the fourth animal only the variant was found and it persisted for over 14 months of infection. WHV appears to be a valuable model for the study of the structure and function of the hepadnavirus precore region.

The precore gene of hepadnaviruses is an in phase contiguous open reading frame that precedes the core gene. The sequence of this gene is well conserved in hepatitis B virus (HBV) and in the other mammalian hepadnaviruses, woodchuck hepatitis virus (WHV) and ground squirrel hepatitis virus. Since there are several important sequences within the gene, such as the encapsidation signal (ϵ) which is involved in the packaging of the RNA pregenome into the virus core, as well as DR1, a short direct repeat involved in the replication of minus-strand DNA (*1–5*), mutations are not well tolerated. However, spontaneous mutations in the precore region of HBV have been found throughout the world (*6–9*).

Normally, the precore protein is targeted to the endoplasmic reticulum and then to the Golgi apparatus by an encoded signal sequence where it undergoes proteolytic cleavage at the amino and carboxy terminals and is glycosylated to produce a soluble extracellular protein known as e antigen (HBeAg) (*10–13*). The presence of HBeAg in serum is associated with viral replication in the liver. Once antibodies to HBeAg are detected, liver injury diminishes and viral DNA is cleared from the serum. A relatively common mutation at nucleotide (nt) 1896 of HBV produces a stop codon which prematurely terminates the translation of the precore gene and leads to HBeAg-negative mutants. HBeAg-negative mutants have been associated with acute fulminant hepatitis or severe chronic hepatitis, in some studies, but an association with increased liver disease has not always been found (*6–9*).

Woodchuck hepatitis virus (WHV) has been a valuable model for studying HBV pathogenesis. Although there is an extra codon in the WHV precore region, there is over 90% sequence identity in the stem–loop region of the ϵ signal of HBV and WHV (*2, 14*). Numerous isolates of HBV have been sequenced, but only 5 WHV sequences are available (*14–17*). In order to determine the type and frequency of naturally occurring sequence variations in the WHV precore and 5′ core region, we examined 101 serum samples from 53 naturally WHV-infected woodchucks from Pennsylvania and North Carolina. From these data we have developed a consensus sequence

[1] To whom correspondence and reprint requests should be addressed. Fax: (919) 829-4455.

0042-6822/96 $18.00

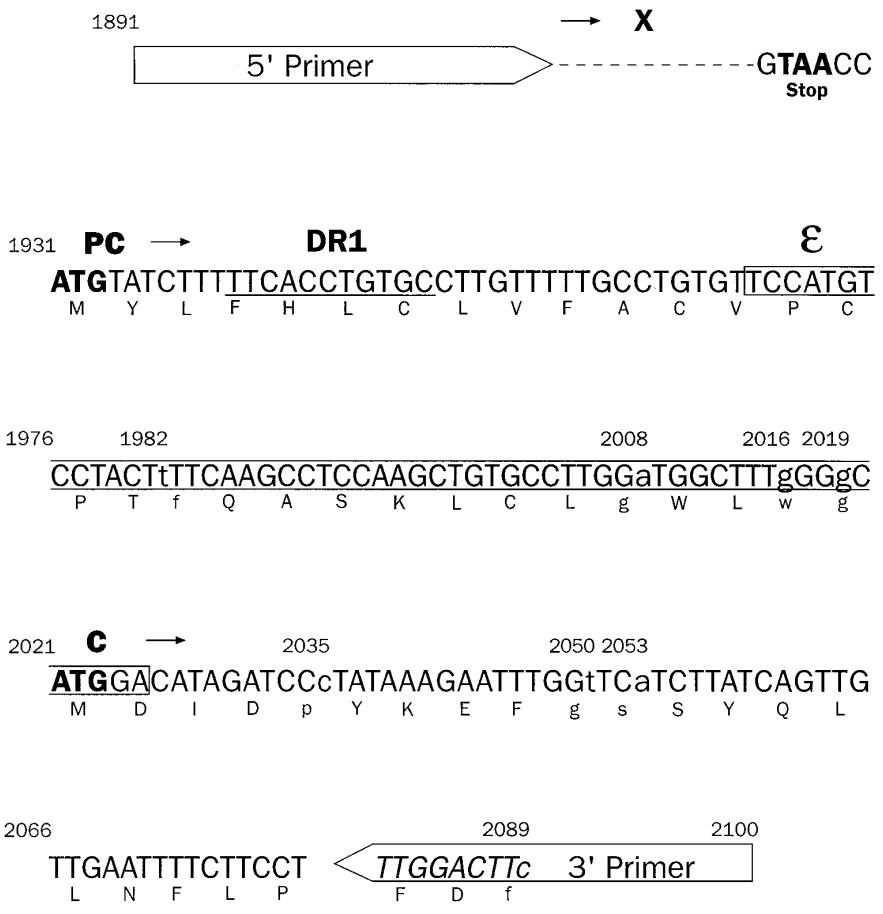

FIG. 1. Consensus sequence of woodchuck hepatitis virus precore and 5′ core region from 101 blood samples from 53 chronically infected woodchucks. Sequencing primers are indicated by the open arrows. The sequences of the primers are shown in the text. Sites between the 5′ primer and consensus sequence are left blank. The DR1 sequence is indicated by an underline. The ϵ stem–loop sequence is indicated by an open box. Positions where variants occurred are designated by lower case letters. The amino acid sequence is shown under the nucleotide sequence. Abbreviations: DR1, direct repeat sequence 1; ϵ, epsilon stem–loop sequence of the pregenome encapsidation signal; X, x open reading frame; PC, precore open reading frame; C, core open reading frame.

for the WHV precore region and used it to assess sequence variants.

The strategy for sequencing and the consensus sequence that was developed are shown in Fig. 1. A commercial DNA purification kit (QIAGEN. Chatsworth, CA 91311) was used to purify DNA from serum. WHV precore and 5′ core region DNA was amplified by polymerase chain reaction. The 5′ primer and 3′ primer were CGGAATTCGTAAGGACCTTTGGACTCC (WHV2 nt 1802–1819) and CGGGATCCACAAGGCAGTACGACTGTC (WHV2 nt 2132–2114), respectively (*15*). PCR amplification was performed according to manufacturer's instructions. To prevent contamination, DNA extraction was performed in a hood, and pre-PCR and post-PCR steps were carried out in separate rooms. PCR reactions were carried out in a room in which WHV was not otherwise present. Sequencing was performed with a commercial kit (PCR-Product Sequenase, United States Biochemical, Cleveland, OH 44128). The primers used were 5′ GCGAATTCTAGGAGGCTGTAGGCATA and 3′ GCGAATTCAAGGTCAGGAAAGAAGTC. Independent PCR-reamplification and resequencing (including forward and reverse DNA strands) were done with 12 precore variant sequences to verify the sequencing results.

The precore region of 33 of the woodchucks was identical and defined of the consensus sequence. Our consensus sequence was identical to 4 of the 5 sequenced WHV strains (WHV 2, 7, 8, 59). There was a single nucleotide difference between the consensus sequence and WHV1. Twenty of the 53 woodchucks were found to have variant sequences in the precore or 5′ core region. Sixteen had variant sequences in precore with or without changes in 5′ core and 4 had changes in the 5′ core region only. The nucleotide changes were not randomly distributed, but were limited to only 4 sites in the precore region and 4 sites in the 5′-core region (Table 1). Ten types of variant sequences were identified in these 20 animals. Four

TABLE 1

Sequence Variations in Precore and 5′ Core Region of Woodchuck Hepatitis Virus Genome Found in 20 of 53 Naturally Infected Woodchucks

Woodchuck (Sex)	Bleed interval (month)	Percore region (nt)				Core region (nt)			
		1982	2008	2016	2019	2035	2050	2053	2089
		T	A	G	G	C	T	A	C
HC134 (F)	0						C		
	6						C		
	9						C		
	17						C		
HC135 (F)	0			G					
	6			G + A					
	9			G + A					
	15			G + A					
	17			G + A					
	17.5			A					
HC150 (M)	0		G						T
	6		G						T
	9		G						T
	17		G						T
HC140 (M)	0			G					
	6			G					
	9			G + A					
	17			G + A					
	17.5			G + A					
HC116 (F)	0								T
	13								T
	22								T
HC073 (F)	0			A					T
	2			A					T
	14			A					T
HC080 (F)	0					T			
	11					T			
HC066 (M)	0								T
	6								T
HC072 (F)	0		G					G	T
	1		G					G	T
	3		G					G	T
CW801 (M)	0			G					
	4			G + A					
CW912 (F)	0	G			G + A	T			
	3	G			G + A	T			
	5	G			G + A	T			
	9	G			G + A	T			
CW1063 (F)	0		G						T
CW1068 (M)	0		G						T
CW1064 (M)	0		G					C	T
	2		G					C	T
CW1061 (M)	0		G					C	T
	1		G					C	T
HC042 (F)	0	G				T			
	12	G				T			
HC161 (F)	0	G				T			
	5	G				T			
HC186 (M)	0	G				T			
	6	G				T			
	11	G				T			
HC181 (F)	0	G				T			
	6	G				T			
HC157 (F)	0	G				T			

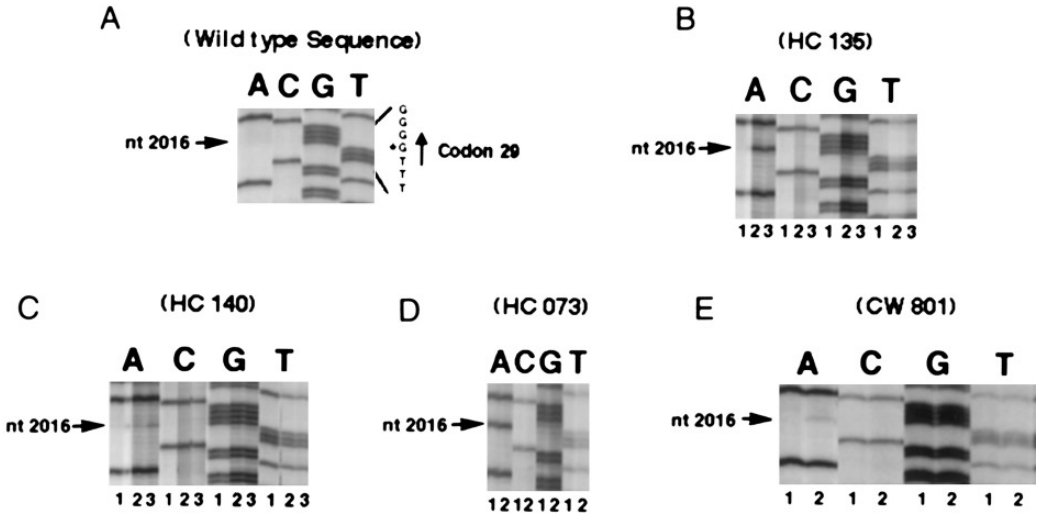

FIG. 2. Naturally occurring sequence variations in codon 29 of the WHV precore region. The sequence of the precore and 5′ core region of WHV from nt 2006 to nt 2023 is shown. (A) wild-type precore sequence, the sequence of nt position 2013 to 2019 is indicated to the right of the panel, the consensus sequence base in codon 29 (nt 2016) is indicated by an asterisk. (B) HC135, note that the relative proportion of A to G in position 2016 changes during the course of infection. The band corresponding to A is light and the band corresponding to G is relatively dark at the first bleed. In subsequent bleeds A becomes the predominant nucleotide in this position. (C) HC140, the abundance of A in position 2016 increases over time as seen in animal HC135 although to a lesser extent. (D) HC073, the mutation at nt 2016 was present in the initial sample and was maintained over 14 months of infection. (E) CW801, a band at nt 2016 corresponding to A becomes apparent within 4 months of the initial bleed. In B through E, 1 indicates initial bleed. The months following the initial bleed are represented as follows: in B, 2 = 15, 3 = 17; in C, 2 = 9, 3 = 17; in D, 1 = 14; in E, 2 = 14.

of them had 1 nucleotide change, 3 had 2 nucleotide changes, and 3 had 3 nucleotide changes.

The frequency of sequence variation in the precore region of HBV and WHV appears to be similar. In two studies of patients from Hong Kong, precore mutations occurred in 32–38% of HBV-infected patients (*6, 18*). In our series we found precore sequence variants in 30% of WHV-infected woodchucks.

One of the more common precore mutations in HBV results in a base change at nt 1896 from G to A in codon 28 that produces a stop codon and leads to HBeAg negative infections. The mutation at nt 1896 occurred in 10–17% of all HBV-infected patients from Hong Kong (*6, 18*). Precore mutations at this same site have also been described in patients from Japan and the Mediterranean region (*8, 9*). We found an analogous stop codon-producing variant in WHV in 4 of the 53 (7.5%) WHV-infected woodchucks (Fig. 2). Because there is an extra codon in the WHV precore region the sequence variation occurred in the 29th codon (*14*). The mutation was also a G to A transition at the corresponding nucleotide site in HBV. In 3 of the 4 animals the sequence from the initial blood sample appeared to be the consensus sequence; however, in later samples the variant occurred as a mixed infection with the consensus sequence. The ratio of the two sequences changed during the course of infection. In animal HC135, the relative abundance of viral DNA with the variant sequence gradually increased relative to the consensus sequence over 12 months. In animal HC140, the levels of the variant sequence were maintained at a low level during 9 months of infection. In CW 801 the variant sequence was apparent 4 months following the initial bleed in which only the consensus sequence was detected. In HC073, the variant sequence predominated and was maintained over 14 months of infection.

Only acute infections resulted when neonatal woodchucks were infected with a mutant WHV-8 containing an experimentally produced mutation in nt 2016 producing a stop codon in codon 29 (*19*). This nucleotide change is identical to the naturally occurring sequence variant we detected in four animals. Our results indicate that this sequence alteration alone is not sufficient to inhibit chronic infection since naturally occurring mutants persisted for 6 to 14 months without being eliminated. It is possible that concurrent infection with wild-type virus may provide some benefit to these presumably WHeAg-negative variants because the wild-type virus produces WHeAg that may afford some protection from the host immune response, as has been postulated in mixed HBV infections (*20*). However, in mixed precore mutant and wild-type precore HBV infections, the virus with the analogous mutation often predominates and becomes the only virus detectable (*6, 20–22*). We observed this pattern in one woodchuck (HC135) in which the variant strain became the only detectable sequence and this evolution may have occurred in HC073 in which only the variant

TABLE 2A

Nucleotide and Cognate Amino Acid Changes in Variant WHVs and Corresponding HBV Precore Mutants

Precore condon		Position (nt)	Nucleotide change	Amino acid change	Prevalence in study animals (%)
WHV	(HBV)				
18[a]	(17)	HBV 1862[b]	GTT to TTT	V to F	
		WHV 1982	TTT to GTT	F to V	6/53 (11.2)
26	(25)	HBV 1888[b]	GGG to GGA	G to G	
		WHV 2008	GGA to GGG	G to G	6/53 (11.2)
29	(28)	HBV 1896[c]	TGG to TAG	W to stop	
		WHV 2016	TGG to TAG	W to stop	4/53 (7.5)
30	(29)	HBV 1899[c]	GGC to GAC	G to D	
		WHV 2019	GGC to GAC	G to D	1/53 (1.9)

[a] WHV has an extra codon in precore region, compared to that in HBV.

[b] Laskus *et al.,* 1993.

[c] Lok *et al.,* 1994.

strain was evident at the time of the first blood sample. In contrast, precore mutant duck hepatitis B virus (DHBV) was outgrown by wild-type DHBV in mixed infections in day-old ducklings which probably do not develop an effective immune response to hepadnavirus infection (*23*).

We found three other WHV precore sequence variations in twelve animals (Table 1). The changes occurred at sites analogous to nucleotide positions of HBV mutations. The first was at WHV nt 1982 in the bulge region, nt 2008 in the upper stem, and nt 2019 in the lower stem of the ϵ signal (Table 2A) (Fig. 3) (*6, 7*). One of the mutations, WHV nt 2019, involved an identical G to A base substitution in WHV and HBV. In the other two sites the base change in WHV was the opposite of the change in HBV. In WHV the change is T to G at nt 1982 and in HBV a G to T occurs. G replaces A at nt 2008 in WHV and A replaces G in HBV. In both WHV and HBV the mutations in the upper and lower stem regions retain the ability to form base pairs within the proposed secondary structure of the ϵ signal (Fig. 3). The sequence variation in the bulge region at position 1982 of WHV, which is the same sequence of the precore region of WHV1, is not likely to alter viral replication, since this area does not form base pairs in the proposed model of ϵ. Several experimentally induced mutations of the bulge region of the stem–loop of HBV, including altering nt 1862 (analogous to 1982 of WHV) from a G to a U did not alter encapsidation significantly (*2, 24*).

A novel sequence (1 nucleotide change in precore and 1 in the 5′ core region) was found in all five woodchucks from North Carolina (HC042, HC157, HC161, HC181, HC186) compared to the consensus sequence developed from the woodchucks from Pennsylvania (Table 1). This suggests that there may be geographic variation in WHV sequences.

The WHV 5′ core sequence variations occurred at four positions (Table 1). All these sequence variations occurred in the third base of the codon and did not change the amino acids encoded (Table 2B).

Certain areas of the stem–loop structure of the ϵ signal in HBV are regarded as critical for viral replication (*2, 24, 25*). Spontaneous mutations have not been described in these sequence specific areas (*6, 7, 21*). These areas are the 3′ loop (5′-GUGC-3′), the unpaired U region, the lower portion of the right upper stem, and the 3′ bulge (5′-UUC-3′) (Fig. 3). Our data support this view in that we did not detect sequence variation in any of these sites. Base pairing and possibly primary sequence are important in the upper portion of the upper stem and upper portion of the lower stem of HBV (*2, 25*). In our study we found no sequence variations in the upper part of the upper stem. Two sequence variants at positions nt 2016 and 2019 were detected in the upper portion of the lower stem that would be consistent with increased strength of base pairing in the lower stem region (6). Tavis and Ganem recently reported that the upstream region (5′-AAU-3′) of DR1 and the 5′ region of DR1 (5′-UACA-3′) are a critical acceptor element for the minus strand DNA transfer of DHBV (*26*). In this study, we found no sequence variations in DR1 and the upstream region of DR1 up to the termination codon of the X gene. It is conceivable that part of this region may play a role in accepting minus-strand DNA primer in the initiation of minus-strand DNA synthesis or play a role as a cis-acting element for transcription factors in precore protein expression.

Lok and coworkers indicated that HBV precore mutations occur mainly at the 4-nt segments in the upper part of the lower stem region of the ϵ signal and postulated that the primary function of the mutations in the precore region is to enhance the stability of the stem–loop structure by increasing the number or strength of the base pairs that occur in the ϵ signal to improve viral replication (*6*). Mutations at nt 1896 and 1899 of HBV frequently occurred together and were proposed to increase stability by the introduction of new or stronger base pairs. Our study also found the same sequence variations at the corresponding positions (nt 2016 and 2019) in WHV. The substitution of G with A in each of these sequence variations could increase the strength of base pairing in the upper part of the lower stem. However, in our study these two sequence variations did not associate with each other. We did not find strong evidence for a greater growth capability of the variant isolates. There were five animals that were infected with viruses containing these sequence variations. In one animal (HC073) the variant sequence was the only isolate detected. In three of them

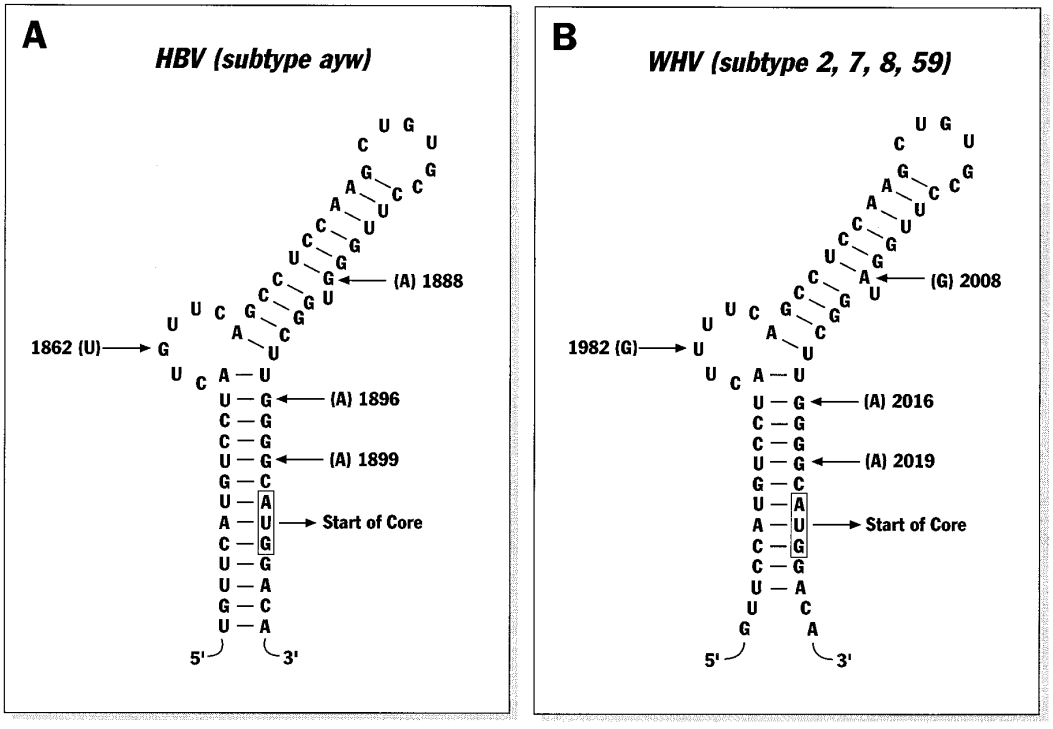

FIG. 3. (A) Diagram of proposed structure for the ϵ stem–loop of HBV. Sites of previously described mutations are indicated by arrows. The base changes and nucleotide positions are also indicated. (B) Consensus sequence of WHV ϵ stem–loop with sites of sequence variants indicated by arrows. The corresponding base changes and nucleotide positions are indicated.

(HC140, CW801, CW912) the variant was present or appeared, but did not increase significantly. In one of them (HC135) the variant sequence became predominant.

We have established a consensus sequence for the precore region of WHV based on samples from 53 woodchucks. It is identical to 4 of the 5 previously sequenced WHV strains. All four of the sequence variants found in the naturally infected woodchucks were analogous to previously described mutations in HBV, suggesting that sequence changes in some sites may be better tolerated or even convey some advantage to viral replication. Using mutant WHV to infect woodchucks can be a valuable approach to study the effects of mutations in the precore region on the pathogenesis of hepadnavirus-induced liver disease and the host immune response. Areas of particular interest include the effect on the viability of WHV by mutations in ϵ signal region and inhibition of WHeAg synthesis.

TABLE 2B

Silent Nucleotide Changes and Cognate Amino Acid in Variant WHV 5′ Core Region

Core codon WHV	Position (nt)	Nucleotide change	Amino acid	Prevalence in study animals (%)
5	WHV 2035	CCC to CCT	P	7/53 (13.2)
10	WHV 2050	GGT to GGC	G	1/53 (1.9)
11	WHV 2053	TCA to TCC/TCG	S	3/53 (5.6)
23	WHV 2089	TTC to TTT	F	10/53 (18.8)

ACKNOWLEDGMENTS

We thank Cindy Brown and Laura Gaylord for assistance with sample collection and Sandie Harris for technical assistance with sequencing. This work was supported by a gift from Glaxo-Wellcome Co.

REFERENCES

1. Junker-Niepmann, M., Bartenschlager, R., and Schaller, H., *EMBO J.* 9, 3389–3396 (1990).
2. Pollack, J. R., and Ganem, D., *J. Virol.* 67, 3254–3263 (1993).
3. Wang, G. H., and Seeger, C., *J. Virol.* 67, 6507–6512 (1993).
4. Tavis, J. E., Perri, S., and Ganem, D., *J. Virol.* 68, 3536–3543 (1994).
5. Rieger, A., and Nassal, M., *J. Virol.* 70, 585–589 (1996).
6. Lok, A. S. F., Akarca, U., and Greene, S., *Proc. Natl. Acad. Sci. USA* 91, 4077–4081 (1994).
7. Laskus, T., Persing, D. H., Nowicki, M. J., Mosley, J. W., and Rakela, J., *Gastroenterol.* 105, 1173–1178 (1993).
8. Omata, M., Ehata, T., Yokosuka, O., Hosoda, K., and Ohto, M., *N. Engl. J. Med.* 324, 1699–1704 (1991).
9. Tur-Kaspa, R., Klein, A., and Aharonson, S., *Hepatology* 16, 1338–1342 (1992).
10. Ou, J., Laub, O., and Rutter, W. J., *Proc. Natl. Acad. Sci. USA* 83, 1578–1582 (1986).
11. Schlicht, H. J., Salfeld, J., and Schaller, H., *J. Virol.* 61, 3701–3709 (1987).

12. Standring, D. N., Ou, J., Masiarz, F. R., and Rutter, W. J., *Proc. Natl. Acad. Sci. USA* 85, 8405–8409 (1988).
13. Carlier, D., Jean-Jean, O., and Rossignol, J. M., *J. Gen. Virol.* 75, 171–175 (1994).
14. Galibert, F., Chen, T. N., and Mandart, E., *J. Virol.* 41, 51–65 (1982).
15. Kodama, K., Ogasawara, N., Yoshikawa, H., and Muralami, S., *J. Virol.* 56, 978–986 (1985).
16. Cohen, J. I., Miller, R. H., Rosenblum, B., Denniston, K., and Gerin, J. L., *Virology* 162, 12–20 (1988).
17. Girones, R., Cote, P. J., Hornbuckle, W. E., Tennant, B. C., Gerin, J. L., Purcell, R. H., and Miller, R. H., *Proc. Natl. Acad. Sci. USA* 86, 1846–1849 (1989).
18. Akarca, U. S., Greene, S., and Lok, A. S. F., *Hepatology* 19, 1366–1370 (1994).
19. Chen, H. S., Kew, M. C., Hornbuckle, W. E., Tennant, B. C., Cote, P. J., Gerin, J. L., Purcell, R. H., and Miller, R. H., *J. Virol.* 66, 5682–5684 (1992).
20. Brunetto, M. R., Giarin, M. M., Oliveri, F., Chiaberge, E., Baldi, M., Alfarano, A., Serra, A., Saracco, G., Verme, G., Will, H., and Bonino, F., *Proc. Natl. Acad. Sci. USA* 88, 4186–4190 (1991).
21. Lok, A. S. F., Akarca, U. S., and Greene, S., *Hepatology* 21, 19–24 (1995).
22. Tran, A., Kremsdorf, D., Capel, F., Housset, C., Dauguet, C., Petit, M., and Brechot, C., *J. Virol.* 65, 3566–3574 (1991).
23. Chuang, W. L., Omatoa, M., Ehata, T., Yokosuka, O., Hosoda, K., Imazeki, F., and Ohto, M., *Hepatology* 19, 569–576 (1994).
24. Knaus, T., and Nassal, M., *Nucleic Acids Res.* 21, 3967–3975 (1993).
25. Fallows, D. A., and Goff, S. P., *J. Virol.* 69, 3067–3073 (1995).
26. Tavis, J. E., and Ganem, D., *J. Virol.* 69, 4283–4291 (1995).