Volume 185, number 1 FEBS 2617 June 1985

Subtype ayw variant of hepatitis B virus

DNA primary structure analysis

V. Bichko, P. Pushko, D. Dreilina, P. Pumpen and E. Gren*

Institute of Organic Synthesis, Latvian SSR Academy of Sciences, Aizkraukles 21, Riga, USSR

Received 28 March 1985

The entire genome of human hepatitis B virus (HBV) occurring in Latvia was sequenced. This sequence, which is 3182 nucleotides long, was compared with the other previously published HBV genomes and was shown to share maximum homology with HBV subtype ayw DNA. The coordinates of 4 main open reading frames as well as hairpin structures are very well conserved in the two genomes. The distribution of nucleotide substitutions among different HBV genomes suggests that the open reading frames P and X can fulfil a coding function. On the basis of primary stucture comparison for hepadnaviral DNAs several evolutionary conclusions can be drawn.

Hepatitis B virus *DNA sequence* *Genetic variation* *Evolution*

1. INTRODUCTION

The human hepatitis B virus has the smallest genome among the known eukaryotic viruses and is composed of circular double-stranded DNA with a nick in the noncoding chain and a long gap with the fixed 5′-terminal end in the coding one [1]. Three HBV-associated antigens are known including the surface antigen (HBsAg), core antigen (HBcAg) and antigen e (HBeAg). The HBsAg has a group-specific antigenic determinant 'a' and two pairs of mutually exclusive subtype-specific determinants: 'd' or 'y' and 'r' or 'w', therefore, it can be classified into 4 major subtypes: adr, adw, ayw, ayr. The cloning in bacterial cells has been performed and the primary structure of HBV DNA determined for the following HBsAg subtypes: ayw, 3182 bp [2]; adw_2, 3221 bp [3]; adw, 3200 bp [4] and 4 variants of adr genomes 3214 and 3188 bp long [4,5] as well as for the nearly complete genome of the adyw subtype [6]. Since individual HBV subtypes populate various geographical areas their evolution apparently proceeds, to a large extent, independently. Hence, a comparison of DNA primary structures among various HBV subtypes permits investigation of the evolution of this virus. However, variation is observed within a single HBV subtype in the restriction maps of DNA [7], while the comparison of 4 variants of the HBV DNAs subtype adr shows that differences are observed within the subtype even with respect to the genome length [5]. Naturally, a comparative study of various HBV subtypes is expedient only when the normal range of variation within each subtype is known. To gain this knowledge, one has to compare at least several primary structure variants of the genome within the same subtype. Four genomic variants are known at present for the HBV subtype adr [5] and two for subtype adw [3,4]. This study offers another primary structure variant for the HBV subtype ayw genome DNA.

Abbreviation: bp, base pairs

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

2. EXPERIMENTAL

2.1. *Isolation, restriction and electrophoresis of plasmid DNA*

The plasmid pHB320 DNA containing the complete HBV genome [8] and the replicative forms of phage vectors M13 mp were isolated from bacterial lysates obtained according to Guerry et al. [9] using chromatography on hydroxyapatite [10]. DNA restriction endonuclease digestion was performed under conditions recommended by the suppliers. Recovery of DNA fragments from agarose gels was performed by electroelution into DE-81 paper [11].

2.2. *Sequence determination procedures*

The HBV DNA fragments were subcloned in phage vectors M13 mp 7, mp 8 and mp 9, and sequenced by the dideoxy chain termination method [12] with some modification.

3. RESULTS AND DISCUSSION

3.1. *The primary structure of HBV DNA*

Fig.1 shows the strategy of HBV DNA primary structure analysis, the determined nucleotide sequence is given in fig.2. The viral DNA in question comprises 3182 bp and shows maximal homology (97%) with the HBV subtype ayw DNA [2]. Neither deletions nor insertions are detected in these two HBV DNAs. Thus the HBV genome dealt with in the present study (clone pHB320) is a new variant of the HBV genome subtype ayw.

3.2. *Open reading frames*

Four major open reading frames were observed in the short chain of the HBV DNA: S (with pre-S), C (with pre-C), P and X whose coordinates coincide completely with those of the subtype ayw DNA. As compared to the same HBV DNA, two amino acid substitutions in gene S (and 4 in pre-s) were detected in positions uncorrelated with subtype change [5]. Three and 7 amino acid substitutions were found in genes C and X, respectively. The coordinates of nonessential open reading frames in the short chain are virtually coincident in the two DNAs. In contrast, considerable variation in open reading frame coordinates in the long chain is observed. For instance, in the pHB320-derived HBV DNA 5 new ATG codons appear (one codon disappears) along with 9 termination codons (7 codons disappear). Interestingly, the new termination codons occur predominantly downstream the new ATG codons or in places where termination codons disappear, despite the small length (20–40 bp) of the frames arising. The coordinates of the open reading frame corresponding to the in vitro transcribed DNA region yielding

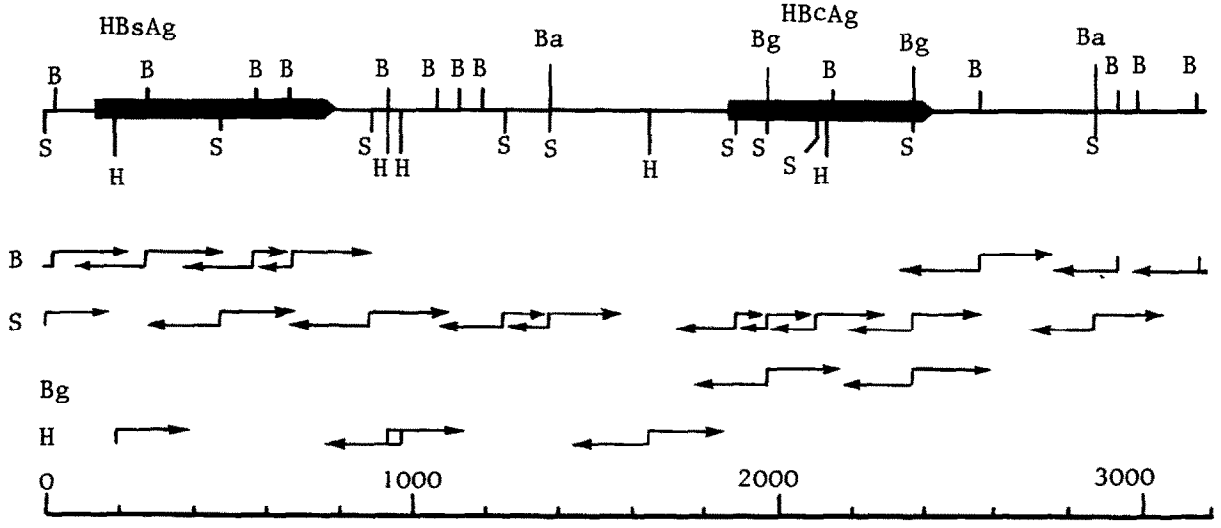

Fig.1. Restriction map of the HBV ayw DNA (clone pHB320). HBsAg and HBcAg genes are designated. Horizontal arrows show the strategy of DNA sequencing. B, *Bsp*I; Ba, *Bam*HI; Bg, *Bg*III; S, *Sau*3A; H, *Hin*dII.

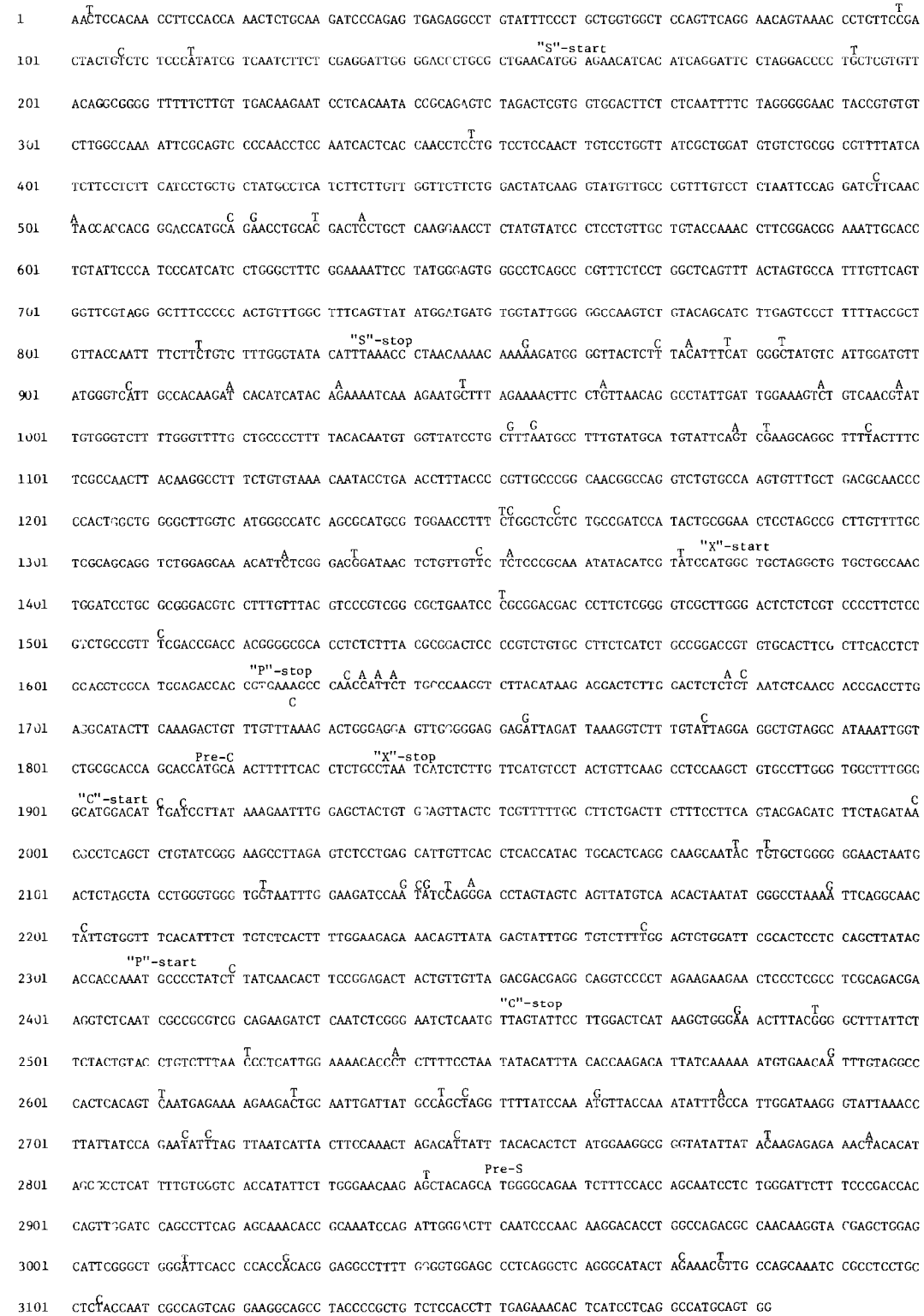

Fig.2. Nucleotide sequence of HBV ayw DNA (clone pHB320). The sequence of the L-strand is shown. Nucleotides are numbered as in [2]. Substitutions of the nucleotide that are observed in HBV ayw DNA [2] are shown above the sequence. The initiation and termination codons of the genes S, C, P and X are designated.

the 700 nucleotide mRNA [13] vary among the two variants of the HBV subtype ayw DNA.

3.3. *Nucleotide substitutions in comparison with published HBV ayw DNA [2]*

Differences in the DNA primary structure affect neither the coordinates of the open reading in the coding (short) chain nor the formation of local hairpin structures. Interestingly, one of the nucleotide substitutions (T–C in position 2774) is located within the putative Hogness box of gene S (TATACAA instead of TATATAA) [14]. The degree of structural conservatism of any region in the HBV ayw genome is strictly correlated with the number of genes located therein (with the frame shift). For example, the number of point mutations in the genome regions carrying two genes is 3-times smaller than the appropriate number in the one gene-carrying regions. The amount of point mutations in the genome segment where frame X overlapped by frame P corresponds to the mutation rate in the HBV ayw DNA regions carrying two genes (0.8 and 1.3%, respectively). In the DNA region where frame X is the only open reading frame, the mutation rate coincides with that for HBV ayw DNA regions carrying a single gene (4.7 and 4.1%, respectively). This suggests that frame X could code for a polypeptide. However, the difference in length between any two HBV genomes under study is a multiple of 3, which allows one to assume that the whole of HBV genome has no regions failing to code for a polypeptide in any of the 3 reading frames (the only exception could be the cloned HBV adr DNA, 3214 bp long [4]). If this holds true, frame X codes for a polypeptide, since there is a region in the genome where it is the only open reading frame, and frame P codes for a polypeptide starting its first ATG codon for the same reasons.

The validity of this assumption is also supported by the distribution of synonymous nucleotide substitutions (not leading to amino acid changes) in genes C, P and X (DNA molecules from the ayw, adw and adr subtypes were compared). The C-terminal region (48 codons) of gene C overlaps with the beginning of gene P and the number of synonymous substitutions in this region is considerably smaller than in the remaining part of gene C (2 and 26%, respectively, fig.3).

The C-terminal part of gene P overlaps with the

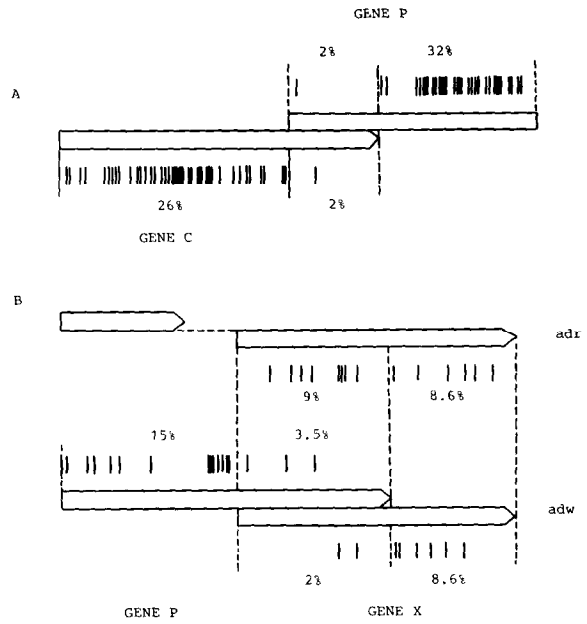

Fig.3. The distribution of synonymous nucleotide substitutions in genes C, P and X from HBV DNAs of different subtypes. (A) HBV ayw DNA (clone pHB320) and HBV adw DNA [3] were compared. (B) HBV ayw DNA (clone pHB320), HBV adw DNA [3] and HBV adr DNA [5] were compared.

beginning of gene X (82 codons), this also decreases the number of synonymous substitutions in this part of gene X from 8.6 to 2%. It is interesting to note that gene P in the adr subtype DNA [5] shows no overlapping with gene X, and the synonymous substitutions in the latter gene are evenly distributed (9 and 8.6%, fig.3). At the same time, gene X imposes restrictions on the variation in the synonymous sites of gene P. The synonymous substitutions in the overlapping regions of the two genes drop from 15 to 3.5%.

3.4. *HBV subtype determinants and evolutionary implications*

Comparative analysis of amino acid sequences of HBsAg derived from 6 HBV DNA structures belonging to 4 different HBV subtypes allows determination of the coordinates of those amino acids whose substitutions are correlated with subtype changes in HBsAg [5]. A similar analysis of the amino acid sequences of pre-S, HBcAg and X, conducted by us leads to a conclusion that the

polypeptides in question also contain amino acids whose substitutions are correlated with HBsAg subtype changes. Interestingly, the polypeptides pre-S and X equally have amino acid substitutions correlated not only with the exchange of alternative determinants (d–y and w–r), as in the case of HBsAg, but also with replacement of their combinations (yw–dw–dr). Thus it is possible to assume the existence of hepatitis B virus HBsAg and HBcAg subtypes correlated with respect to their HBsAg subtypes. This also suggests the absence of genetic recombination among the known HBV subtypes. The study of geographical distribution of HBV subtypes leads to the same conclusion.

This makes it possible to study phylogenetic relationship of different HBV subtypes and other hepadnaviruses by analysing nucleotide substitutions [15]. We compared only those regions of genes C and P that do not overlap with other genes (840–1373, 1903–2307, 2456–2848). Deletions and insertions were taken into account. The rate of accumumalation of synonymous substitutions was taken as $5.1 \pm 0.3 \cdot 10^{-9}$ per site annually [15].

The main conclusions of this part of our study are as follows:

(i) Division of all HBV sequences into 3 phylogenetic groups completely coincides with their division into HBsAg subtypes (HBV adyw DNA and pHB320-derived HBV DNA belong to the ayw group).

(ii) The HBV DNA of the ayw and adw subtypes are closer related to one another than to the adr subtypes and diverged from a common ancestor ~52–55 million years ago.

(iii) HBV divergence began ~60–65 million years ago. This date is approximately coincident with the time of origin of the primate order.

(iv) The common ancestor of HBV and WHV (woodchuck hepatitis virus) existed ~180 million years ago when mammals appeared. This enables one to infer that hepadnaviruses have evolved parallel to the species under consideration. In a recent study of Mandart et al. [16] a similar hypothesis has been proposed on the basis of genome structure comparison conducted for 3 hepatitis viruses: HBV ayw subtype, woodchuck hepatitis virus and duck hepatitis B virus.

REFERENCES

[1] Summers, J., O'Connel, A. and Millman, J. (1975) Proc. Natl. Acad. Sci. USA 72, 4597–4601.
[2] Galibert, F., Mandart, E., Fittoussi, F., Tiollais, P. and Charnay, P. (1979) Nature 281, 646–650.
[3] Valenzuela, P., Quiroga, M., Zaldivar, J., Gray, P. and Rutter, W.J. (1981) in: Animal Virus Genetics (Fields, B. et al. eds.) pp. 57–70, Academic Press, New York.
[4] Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugino, Y. and Nishioka, K. (1983) Nucleic Acids Res, 11, 1747–1757.
[5] Fujiyama, A., Miyanohara, A., Nozaki, C., Yoneyama, T., Ohtomo, N. and Matsubara, K. (1983) Nucleic Acids Res. 11, 4601–4610.
[6] Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., MacKay, P., Leadbetter, G. and Murray, K. (1979) Nature 282, 575–579.
[7] Siddiqui, A., Sattler, F. and Robunson, W.S. (1979) Proc. Natl. Acad. Sci. USA 76, 4664–4668.
[8] Bichko, V., Kozlovskaya, T., Dishler, A., Pumpen, P., Janulaitis, A. and Gren, E. (1982) Gene 20, 481–484.
[9] Guerry, P., LeBlanc, D.J. and Falkow, S. (1973) J. Bacteriol. 116, 1064–1066.
[10] Colman, A., Byers, M.J., Primrose, S. and Lyons, A. (1978) Eur. J. Biochem. 91, 303–310.
[11] Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112, 295–298.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Standring, D., Rall, L., Laub, O. and Rutter, W.J. (1983) Mol. Cell. Biol. 3, 1774–1782.
[14] Rall, L.B., Standring, D.N., Laub, O. and Rutter, W.J. (1983) Mol. Cell. Biol. 3, 1766–1773.
[15] Miyata, T., Yasunaga, T. and Nishida, T. (1980) Proc. Natl. Acad. Sci. USA 77, 7328–7332.
[16] Mandart, E., Kay, A. and Galibert, F. (1984) J. Virol. 49, 782–792.