Volume 262, number 2, 345–348 FEBS 08282 March 1990

Amino acid sequence similarity between the terminal protein of hepatitis B virus and predicted hepatitis delta virus gene product

Yu.E. Khudyakov and A.M. Makhov

The D.I. Ivanovsky Institute of Virology, Academy of Medical Sciences of the USSR, Gamaleya st. 16, Moscow 123098, USSR

Received 17 January 1990

The comparative analysis of primary and secondary structures, and hydropathy plots of hepatitis B virus (HBV) and hepatitis delta virus (HDV) proteins was carried out. Two short regions belonging to the HBV terminal protein were shown to be homologous to two regions; one encoded by HDV ORF5, and the other encoded by small ORF of the HDV antigenomic RNA strand. We propose a new protein containing both these regions may be synthesized in HDV infected cells. Stiking structural homology between the terminal protein of HBV and this predicted protein called HDAg′ of HDV may indicate a possible functional similarity. We hypothesize the HDAg′ may interact with and inhibit the polymerase activity of HBV.

Protein homology; Hepatitis B virus; Hepatitis delta virus; Terminal protein

1. INTRODUCTION

Hepatitis delta virus (HDV) is a recently identified human pathogenic agent [1]. This virus is a defective naturally occurring satellite of human hepatitis B virus (HBV). The hepatitis delta virions are 36-nm particles with an envelope composed of hepatitis B surface antigen [1]. The internal part of these particles consists of a HDV-specific protein, hepatitis delta antigen (HDAg), and a small genomic RNA [1,2]. The HDAg observed in the sera and liver of infected humans, or experimentally infected animals, has been shown to be composed of two protein species of between 24 and 30 kDa [3,4], or a major 26 kDa protein and multiple minor proteins [5]. All of these proteins are related in sequence [4] and are encoded by the large open reading frame (ORF5) of the HDV antigenomic RNA [6,7]. No other HDV specific proteins have been detected.

The relationship between HDV and HBV suggests that HDV may use some unrecognized HBV-specific function(s) for its own replication despite the clear distinction in genetic organization of these viruses. We now report protein structure similarities of these viruses which suggests a possible functional interaction between HDV and HBV.

2. MATERIALS AND METHODS

The primary analysis of sequences was carried out with the use of matrix homology program DNASEQ (Institute of Protein, Academy of Sciences of the USSR). The statistical significance of sequence alignment was evaluated according to Feng et al. [8]. The correlation coefficient (K) of hydropathy plots [9] was calculated by the method described in Sweet et al. [10]. The method of Chow-Fasman was used for calculation of protein secondary structure [11]. Sources of sequences of HBV-specific proteins were taken from Miller [12]. HDV sequence information was obtained from published data [6,7,13,14].

Correspondence address: Yu.E. Khudyakov, D.I. Ivanovsky Institute of Virology, Academy of Medical Sciences of the USSR, Gamaleya St. 16, Moscow 123098, USSR

3. RESULTS AND DISCUSSION

Initially, we conducted a computer-assisted translation of published HBV and HDV genomes. Terminator codons were considered as coding for amino acid X. The only strand of the HBV genome comprising the known ORFs for HBV proteins was translated in all 3 phases. For HDV, both antigenomic and genomic strands were translated. Thus, the sequences of 3 HBV large polyproteins and 6 HDV polyproteins were obtained. Each of the HBV polyproteins were compared with all of the HDV polyproteins. Short regions of 10 residues containing no less than 6 identical amino acids were chosen in comparing sequences. For the primary analysis of sequence similarity the subtype ayw HBV genome [14] and the HDV genome described by Wang et al. [7,13] were used. To define homologous regions more precisely other known sequences of HDV and HBV were analysed.

Two protein regions encoded by two segments of the HBV genome (2372–2461 and 2477–2542) have been identified as having amino acid sequence homology with two regions encoded by two segments of the HDV antigenomic strand (459–543 and 650–724). These two HBV protein regions were identified as two functional determinants in the N-terminal part of gene P product [15] which has been previously shown to function as a

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

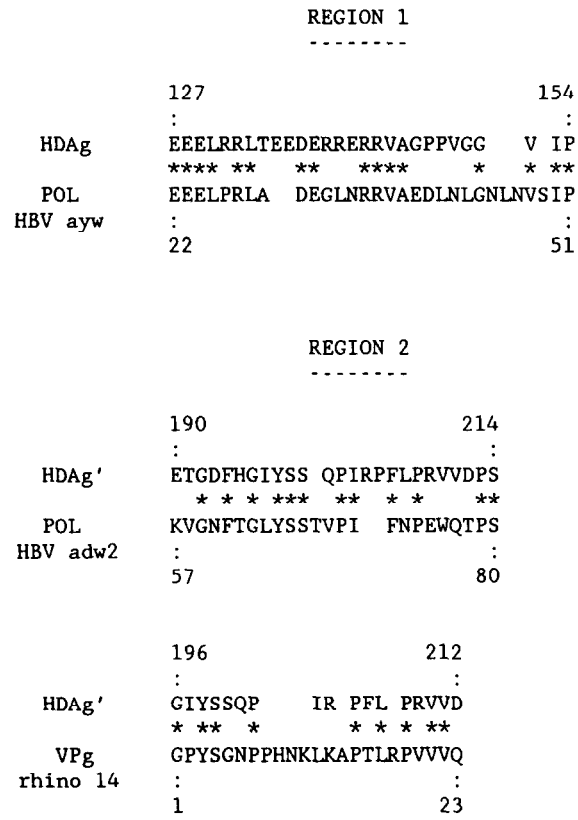

Fig. 1. Comparison of region 1 and region 2 of HDAg' with the N-terminal sequence of the gene P product of HBV and VPg of rhinovirus 14. Asterisks indicate identical residues in the comparing sequences. HDAg', sequences of region 1 and 2 of the predicted frameshift polypeptide of HDV; POL HBV adw2, appropriate regions of the terminal protein of HBV subtype adw2; VPg rhino 14, the genome linked protein of rhinovirus 14.

terminal protein [15–17]. The first region encodes a negatively charged amphipathic α-helix. Such helices have been identified in the terminal proteins of adenoviruses and phi-29 phage just before a nucleic acid linkage site [18]. Helices of this type have also been found to play a significant role in the interaction of some transcriptional factors with other proteins of the transcription apparatus. These protein helices contain determinants of protein interaction [19]. The second region has been demonstrated to be homologous to the VPg of picornaviruses [15]. Thus, two functionally essential regions of HBV gene P product are similar to two protein regions encoded by the HDV antigenomic RNA (Fig. 1). The percentage of coinciding amino acids in region 1 is 50% ($P < 10^{-5}$). The correlation coefficient (K) of hydropathy plots for the region of HBV and HDV polyproteins is 0.561 ($P < 0.01$). Region 2 has 46.2% identical residues ($P < 10^{-5}$). The K value is 0.463 ($P < 0.02$). Region 2 of HDV has significant similarity (39.1%; $P < 10^{-3}$) to the VPg of rhinovirus 14 (fig. 1).

It is important to note that while the HBV regions belong to the same protein, the homologous regions of HDV are encoded by two different reading frames. Region 1 corresponds to the fragment 127–154 of HDAg and is encoded by ORF5. Region 2 is coded by a short ORF. The stop codons of the ORF5 and the short ORF are separated by only one nucleotide (fig. 2). A frame shift mechanism or splicing event must occur to form a HDV protein encoded by both ORF5 and the short ORF. We did not, however, find sequences resembling canonical splice sites; consequently, frame shifting seems to be a more plausible mechanism. Such a mechanism for protein synthesis is known to be widely used in viral genome expression [20]. It has been shown that protein coding sequences can carry an additional message in the form of a universal 3-base periodical pattern (G-nonG-N)m. At recognized slippage sites the G-periodical pattern has been found to be interrupted. It reappears downstream from the slippage sites in a new frame [21]. We used this rule to find G-periodicity interrupted in the ORF5 just at the border of the short ORF. The periodicity is restored in the short ORF (fig. 3). Thus, according to the G-periodicity rule the possible frame shifting site may be

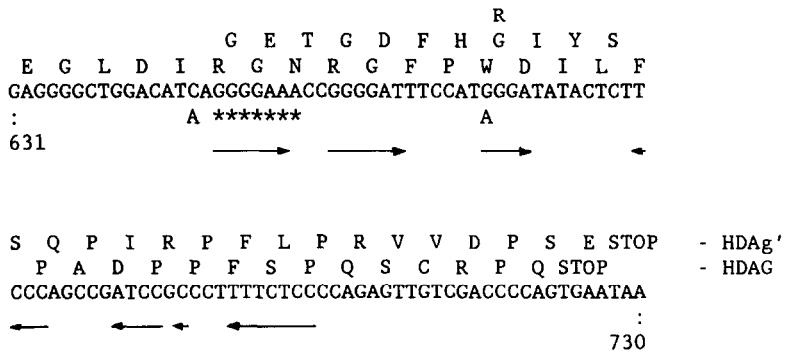

Fig. 2. Genetic organization of the region 631–730 of HDV antigenomic RNA strand. Sequences of the translation products of ORF5 (HDAg) and small ORF (HDAg') are shown over the line in single letter code. Symbol A under the line indicates two substitutions in HDV genome primary structure leading to the appearance of termination codon in these ORFs. Symbol R shown over the sequence of HDAg' demonstrates substitution Gly for Arg in some protein sequences of the protein. Asterisks mark probable slippage site. Arrows indicate reverse repeats in RNA sequence.

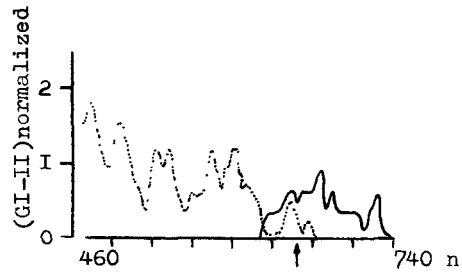

Fig. 3. Distribution of guanines in the codons of ORF5 (dotted line) and small ORF (solid line). Arrow indicates position of probable ribosome frameshifting.

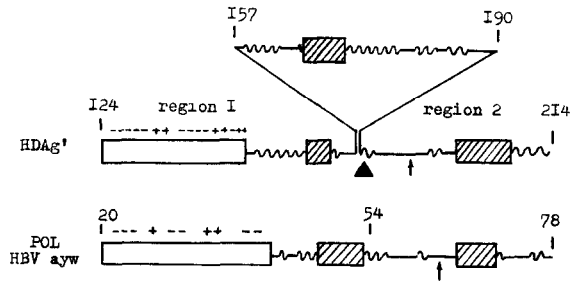

Fig. 4. Comparison of secondary structures predicted for the homologous regions of the terminal protein of HBV subtype ayw, POL HBV ayw, and HDAg'. Hollow box, α-helix. Shaded box, β-strand. Wavy line, random coil. Straight line, β-turn. Distribution of positively and negatively charged residues in α-helices is depicted by + and −, correspondingly. Triangle indicates the possible site of ribosome slippage. Arrow shows localization of conserved Tyr in the predicted nucleic acid linkage site of the protein.

located in a region coding for the segment 185–195 of HDAg. A stretch of four G residues is exactly placed within this region. Frameshifts are known to occur often in sites containing stretches of identical nucleotides [20]. Furthermore, the whole 3'-terminal part of ORF5 is saturated with reverse repeats (fig. 2), which suggests the formation of a stable secondary structure within this mRNA region. Strong hairpin structures are also known to occur near frame shifting sites [20].

These data demonstrate the possibility for the synthesis of a new protein encoded by HDAg mRNA. This new protein designated HDAg', may be synthesized by a translation slippage mechanism. The C-terminal part of the HDAg' contains a new domain homologous to proposed DNA linkage site of the terminal protein of hepadnaviruses [15]. In addition to primary structure homology we found significant similarities of the protein secondary structure predicted by the method of Chow-Fasman (fig. 4).

The striking structural similarities of the predicted new product of HDAg gene and the terminal protein of HBV suggest that they may share a common function(s). The N-terminal part of the gene P product has been proposed to participate in priming of HBV DNA replication [15–17]. However, fundamental differences in genome organization and in the mechanism of genome replication of these two viruses suggests that HDAg' priming for HDV RNA synthesis is unlikely. Nonetheless, the possibility of HBV polymerase interaction with HDAg', like the terminal protein of HBV, cannot be excluded. Region 2 of HDAg' contains a conserved Tyr residue (fig. 1). The terminal protein of HBV is known to be linked with DNA through Tyr [16,17]. In a previous publication [15] we proposed that the Tyr located in region 2 may be responsible for this linkage. Observation of essential structural homology between HDAg' and HBV-specific terminal protein suggests that HDAg' containing a single Tyr residue may also be connected with nucleic acid. Furthermore, HDAg' may compete with the terminal protein of HBV for interaction with the polymerase and/or for interaction with pregenome RNA to inhibit protein expression encoded by this mRNA and replication of HBV virus. Thus, the HDAg' may interfere with the mechanism of HBV replication. Multiplication of HBV is known to be inhibited by HDV [1].

In conclusion, it is important to note that some published HDV genome sequences were found to contain a nucleotide substitution (fig. 2) leading to the appearance of an additional stop codon in ORF5 [6,7,13,14]. As a result of this change, Gly is substituted for Arg before Tyr in region 2 (fig. 1). This mutation may influence the activity of HDAg' because Gly is part of a consensus for a nucleic acid linkage site [15]. If that mutation is formed during HDV infection and the corresponding region is the hot spot of HDV genome, then the modulation of the primary structure of HDAg' might regulate activity of the predicted protein.

Acknowledgements: The authors wish to express their thanks to Dr Howard A. Fields for critical review of the manuscript and valuable discussion.

REFERENCES

[1] Rizzetto, M., Hoyer, B., Canese, M.G., Shih, J.W.-K., Purcell, R.H. and Gerin, J.L. (1980) Proc. Natl. Acad. Sci. USA 77, 6124–6128.
[2] Bonino, F., Hoyer, B., Shih, J.W.-K., Rizzetto, M., Purcell, R.H. and Gerin, J.L. (1984) Infect. Immunol. 43, 1000–1005.
[3] Bonino, F., Heermann, K.H., Rizzetto, M. and Gerlich, W.H. (1986) J. Virol. 58, 945–950.
[4] Pohl, C., Baroudy, B.M., Bergmann, K.F., Cote, P.J., Purcell, R.H., Hoofnagle, J. and Gerin, J.L. (1987) J. Infect. Dis. 156, 622–629.
[5] Puig, J., Anderson, L.J. and Fields, H.A. (1987) in: Hepadna Viruses (Robinson, W. et al. eds) pp.579–590, Alan R. Liss, New York.
[6] Makino, S., Chang, M.-F., Shih, C.-K., Kamahora, T., Vannier, D.M., Govindarajan, S. and Lai, M.M.C. (1987) Nature 329, 343–346.
[7] Wang, K.-S., Choo, O.-L., Weiner, A.L., Ou, J.-H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L. and Houghton, M. (1986) Nature 323, 508–513.

[8] Feng, D.F., Johnson, M.S. and Doolittle, R.F. (1985) J. Mol. Evol. 21, 112–125.
[9] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[10] Sweet, R.M. and Eisenberg, D. (1983) J. Mol. Biol. 171, 479–488.
[11] Chow, P. and Fasman, G. (1978) Adv. Enzymol. 47, 45–148.
[12] Miller, R.H. (1987) Science 236, 722–725.
[13] Wang, K.-S., Choo, O.-L., Weiner, A.J., Ou, J.-H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L. and Houghton, M. (1987) Nature 328, 456.
[14] Kuo, M.Y.P., Goldberg, J., Coates, L., Mason, W., Gerin, J.L. and Taylor, J. (1988) J. Virol. 62, 1855–1861.
[15] Khudyakov, Yu.E. and Makhov, A.M. (1989) FEBS Lett. 243, 115–118.
[16] Bosch, V., Bartenschlager, R., Radziwill, G. and Schaller, H. (1988) Virology 166, 475–485.
[17] Bartenschlager, R. and Schaller, H. (1988) EMBO J. 7, 4185–4192.
[18] Vartapetian, A.B. and Bogdanov, A.A. (1987) Prog. Nucleic Acids Res. Mol. Biol. 34, 209–251.
[19] Giniger, E. and Ptashne, M. (1987) Nature 330, 670–672.
[20] Jacks, T., Madhani, H.D., Masiarz, F.R. and Varmus, H.E. (1988) Cell 55, 447–458.
[21] Trifonov, E.N. (1987) J. Mol. Biol. 194, 643–652.