genome, EIAV appears slightly more related to HTLV-III (p26, protease, segments 2, 3, 4, and 7 of *pol*), or visna (p15, p11, segments 1 and 6 of *pol*), but with the possible exception of *pol* segment 6, these differences are not large. Since visna and HTLV-III appear no closer to each other than either is to EIAV, we conclude that the three viruses are about equidistant from each other.

Chiu *et al.* (*34*) recently compared the translated nucleotide sequence at the NH_2-terminus of RT (corresponding to our *pol* segments 1 and 2) in EIAV, CAEV, and HTLV-III. They concluded that the three viruses are about equidistant from each other. Since CAEV is a close relative of visna (81 percent amino acid identities in the region sequenced by Chiu *et al.*), our results are in good agreement with theirs. They also concluded that divergence among the lentiviral *pol* genes is about the same as that observed between the *pol* genes of BLV and HTLV-I. This conclusion, however, while true for *pol* segments 1 and 2 does not apply to *pol* segments 6 and 7 (see Table 1). There, as throughout *gag* and protease, BLV and HTLV-I are more highly related to each other than are EIAV, visna, and HTLV-III to each other. Similarly, within RT the BLV–HTLV-I and -II group appears more closely related to the avian type C viruses than to the mammalian type C viruses (*2*, *21*, *34*), but in other genomic regions closer relatedness to the mammalian viruses is seen (*21*). Thus a phylogenetic tree that accurately reflects relationships in one segment of the genome may not apply to other segments. This may result from real differences in evolutionary rate, from limitations in our methods of analyzing relatedness, from viral recombination, or from some combination of these factors. Until this issue is clarified further, we view the lentiviruses, the BLV–HTLV-I and -II group, the mammalian type C viruses, and the avian type C viruses as four major retroviral groups that are approximately equidistant from each other.

REFERENCES AND NOTES

1. M. A. Gonda *et al.*, *Science* **227**, 173 (1985).
2. P. Sonigo *et al.*, *Cell* **42**, 369 (1985).
3. M. Gonda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press
4. M. Gonda *et al.*, *Am. J. Vet. Res.* **39**, 731 (1978).
5. O. Narayan *et al.*, *J. Gen. Virol.* **41**, 343 (1978); Y. Kono *et al.*, *Arch. Virol.* **41**, 343 (1973); R. Montelara *et al.*, *J. Biol. Chem.* **259**, 10539 (1984).
6. L. Montagnier *et al.*, *Ann. Virol.* **135E**, 119 (1984).
7. J. M. Casey *et al.*, *J. Virol.* **55**, 417 (1985).
8. L. S. Levy *et al.*, unpublished results.
9. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
10. For sequencing, fragments were subcloned into pUC8 or modified pUC8 vectors. Reactions for G, G + A, C, C + T, and A + C were performed. For about 64 percent of the DNA, the sequence of both strands was determined. There were no compressions within the remaining 36 percent. Storage and analysis of the sequence data were performed on an IBM PC as described (*11*)
11. R. Stephens, *Gene Anal. Tech.* **2**, 67 (1985).
12. L. Ratner *et al.*, *ibid.*, p. 450; S. Wain-Hobson *et al.*, *Cell* **40**, 9 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985).
13. G. Peters and C. Glover, *J. Virol.* **35**, 31 (1980).
14. F. Harada *et al.*, *J. Biol. Chem.* **250**, 3487 (1975); G. Peters *et al.*, *J. Virol.* **21**, 1031 (1977); G. Peters and C. Glover, *ibid.* **33**, 708 (1980).
15. S. M. Mount, *Nucleic Acids Res.* **10**, 459 (1982).
16. L. Henderson *et al.*, in preparation.
17. B. Parekh *et al.*, *Virology* **107**, 520 (1980); R. C. Montelaro *et al.*, *J. Virol.* **42**, 1029 (1982).
18. The computer program ALIGN (*19*) brings two sequences into optimal alignment based on the mutation data matrix, a matrix bias of 6, and a gap penalty of 6, and then assigns a score to the comparison. A score for a random comparison of two sequences of the same length and composition is obtained by jumbling the two and comparing them. This latter operation was performed 300 times, yielding an average score for a random comparison. The align score (AS) expresses in standard deviation (SD) units how far the test score is above the average score. A score of 5 SD (probability of chance occurrence 3×10^{-7}) has been taken to imply evolutionary relatedness of two sequences, and a score between 3 and 5 SD is suggestive of such relatedness.
19. M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1979), vol. 5, suppl. 3, pp. 1–8.
20. S. Oroszlan *et al.*, unpublished results.
21. N. R. Rice *et al.*, *Virology* **142**, 357 (1985).
22. T. Shinnick *et al.*, *Nature (London)* **293**, 543 (1981)
23. Y. Yoshinaka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1618 (1985).
24. D. E. Schwartz *et al.*, *Cell* **32**, 852 (1983).
25. M. Seiki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3618 (1983).
26. K. Shimotohno *et al.*, *ibid.*, **82**, 3101 (1985).
27. T. D. Copeland *et al.*, *J. Virol.* **36**, 115 (1985).
28. T. D. Copeland *et al.*, *Virology* **143**, 676 (1985).
29. Y. Yoshinaka *et al.*, *J. Virol.*, in press.
30. H. Toh *et al.*, *Nature (London)* **305**, 827 (1983).
31. P. Hippenmeyer and D. Grandgenett, *Virology* **137**, 358 (1984).
32. Y. Yoshinaka *et al.*, *J. Virol.* in press.
33. N. R. Rice *et al.*, *ibid.* **29**, 907 (1979).
34. I.-M. Chiu *et al.*, *Nature (London)* **317**, 366 (1985).
35. Conservative groupings are A, G, P, S, and T; I, L, M, and V; D, E, N, and Q; H, K, and R; F, Y, and W; R. M. Schwartz and M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1978), vol. 5, suppl. 3, pp.353–358.

36. We thank J. Elser for help in sequencing and S. Gregory for manuscript preparation. We thank L. Henderson and S. Oroszlan for sharing results with us. Supported by NCI DHHS (contract NOI-C)-23909 with Litton Bionetics).

2 October 1985; accepted 14 November 1985

Putative Reverse Transcriptase Intermediates of Human Hepatitis B Virus in Primary Liver Carcinomas

H. WILL,* J. SALFELD, E. PFAFF, C. MANSO, L. THEILMANN, H. SCHALER

Nucleocapsid-pol fusion proteins have been detected by serological screening hepatocellular carcinoma tissues that contain hepatitis B virus (HBV) DNA. The existence of these fusion proteins suggests that HBV may synthesize its reverse transcriptase in a fashion analogous to the way that retroviruses synthesize and process a precursor. The accumulation of HBV reverse transcriptase intermediates in tumorous tissues and not in other tissues may be related to the absence of viral core particles and possibly contributes to tumor development.

HUMAN HEPATITIS B VIRUS (HBV) contains a DNA genome but resembles retroviruses (i) in its mode of replication involving reverse transcription of an RNA pregenome, (ii) in its gene organization, and (iii) in its association with tumor development (*1*). At the level of transcription, both differences and similarities between HBV and retroviruses exist (*2*). In contrast to retroviruses (*3*), all transcripts identified so far are unspliced and originate from different promoters; however, as in retroviruses, they are coterminal. For HBV and the related duck hepatitis B virus (DHBV) the major messenger RNA's (mRNA's) identified have been related to the hepatitis B surface antigen (HBsAg) and nucleocapsid protein, hepatitis B core antigen (HBcAg), but not to the reverse transcriptase believed to be encoded by the pol frame (*4*). Only the HBcAg mRNA covers the complete pol frame; for expression of the pol protein from this mRNA, internal translation initiation would have to occur. Alternatively, in analogy to retroviruses (*3*), the HBV reverse transcriptase could be synthesized via a HBcAg-Pol (c-pol) fusion protein. This is conceivable since the HBcAg coding region (C gene) of HBV overlaps with that of the pol frame reminiscent of the gag-pol frame arrangement in most retroviruses. As confirmation of this hypothesis, we have identified and charac-

Mikrobiologie and ZMBH, University of Heidelberg, Im Neuenheimer Feld 230 and 282, 6900 Heidelberg, Federal Republic of Germany.

*Present address: Max-Planck-Institut für Biochemie, Am Klopferspitz, 8033 Martinsreid/München, Federal Republic of Germany.

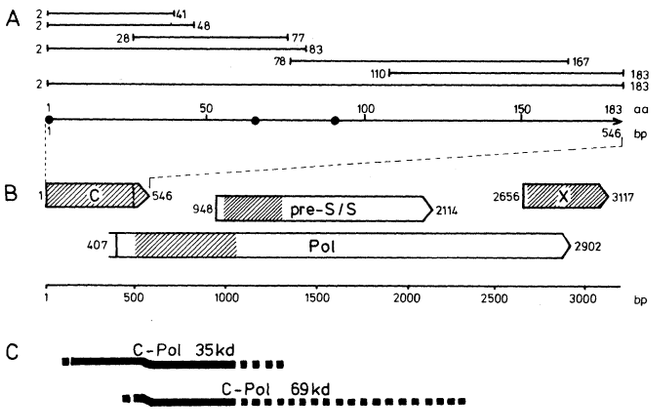

Fig. 1. Sections of HBcAg or HBeAg expressed in *E. coli*, gene organization of HBV, and deduced structure of the 35- and 69-kD c-pol fusion proteins. (A) Polypeptides of the C gene expressed in *E. coli* (*8*) are schematically drawn and the positions of the NH_2- and COOH-terminal amino acids are indicated. (B) The HBV genome is presented in a linear form. Open arrows represent known and predicted genes. Areas covered by antisera to HBV proteins expressed in *E. coli* (*8*) are shaded. The COOH-terminus of the HBeAg protein (*9*) is indicated by the vertical line within the C gene. The nucleotide positions of all open reading frames are according to the nomenclature of Pasek *et al.* (*10*). (C) Structure of the 35- and 69-kD c-pol polyproteins as deduced from the immunoblot experiments. Continuous lines are drawn in areas covered by reacting antisera. Broken lines indicate possible locations of the NH_2- and COOH-terminal amino acids.

terized such fusion proteins in this study.

All liver samples (Table 1) were screened by immunoblotting with an antiserum to HBcAg (c2-183 in Fig. 1A) in order to detect polypeptides antigenically related to the major viral HBcAg protein. In the livers with ongoing viral replication (Table 1, patients 1 and 4) a strong 21,000-dalton (21-kD) signal was observed (Fig. 2, lanes 1 and 4), a position corresponding to the size of the major HBcAg protein (*5*). In addition, with the extract of liver from patient 4, a minor specific signal was observed at 38 kD (Fig. 2, lane 4) (see below). In the livers where there was no acute HBV infection, the 21-kD HBcAg was not detected (Fig. 2, lanes 2 and 3 and 6 to 11), a finding consistent with the absence of HBc-related antigens in serum and liver (Table 1). Most of the livers contained only integrated HBV DNA, a fact known to be associated with the lack of HBcAg and HBeAg expression (*6*). In two of the hepatocellular carcinoma tissues tested the same antiserum reacted specifically with a 35-kD protein (Fig. 2, lane 5). In addition, in one primary liver carcinoma (PLC) tissue a 69- and 92-kD polypeptide reacted specifically (Fig. 2, lane 5). These results suggest that all these polypeptides contain HBcAg sequences.

The high molecular weight HBcAg proteins were likely to contain cellular sequences or HBV-encoded sequences of the pol, pre-S, or X open reading frames (Fig. 1B). To test for pol-related protein sequences, we used a rabbit antiserum to the NH_2-terminal amino acid sequence of the pol frame (p30-201) and reacted it with proteins of a liver from a patient with ongoing viral replication (patient 4) and with proteins of the PLC tissue of patient 5. (Patient 10 was tested in the same way.) A specific immune reaction was observed with the 35-, 38-, and 69-kD proteins but not with the 92-kD protein (Fig. 3A, lanes 1 to 3, and Fig. 2, lane 4). This suggests that HBcAg sequences are fused with pol frame sequences in all these polypeptides except that of 92 kD. Because of the high background in a control experiment (Fig. 3B, lane 3) this could not be proved unequivocally for the 38-kD polypeptide (legend to Fig. 3). Liver extracts of patients 4, 5, and 10 screened by immunoblot analysis with antisera to X frame and pre-S frame proteins revealed no reaction with HBcAg-related proteins. Thus, these proteins do not contain the corresponding pre-S or X amino acid sequences.

To determine which parts of the HBcAg protein sequence were contained in the 35-, 38-, 69-, and 92-kD polypeptides, we again used the immunoblotting technique. Rabbit antisera to short fragments of the HBcAg protein were tested with liver extracts of patients 4, 5, and 10. The antisera were used either directly or after preliminary incubation (indicated by PI + or −) with denatured HBcAg (c2-183) to demonstrate the specificity of the immune reaction (Fig. 3, A and B). All antisera to HBcAg except those to NH_2- and COOH-termini reacted specifically with the 35-kD protein (Fig. 3A, lanes 4 to 14). This suggests that the 35-kD protein lacks the NH_2- and COOH-termini of the 21-kD HBcAg protein. The strong reaction of antiserum c28-77 with the 35-kD protein compared to the weak reaction with the 21-kD HBcAg protein (only seen after longer exposure) may be due to the fact that the NH_2-terminus of the 35-kD polypeptide lies within the amino acids 28 to 77 of the C gene sequence. The HBcAg sequence may thus start at the second AUG of the C gene (amino acid 66), extend at least to amino acid 110 and then be fused to pol frame sequences extending to about amino acid 200 of the pol frame (Fig. 1C). The 69-kD protein reacted specifically only with the antiserum to the COOH-terminus (Fig. 3A, lanes 13 and 14) and the complete HBcAg protein (Fig. 2, lane 5), suggesting that it contains only a short stretch of the COOH-terminus of the 21-kD HBcAg protein. Therefore, the 69-kD protein appears to begin close to the COOH-terminus of the C gene and is fused to pol protein sequences.

Table 1. Origin and viral markers of liver tissues. Diagnoses were made by clinical, serological, and autopsy findings. All liver samples were obtained from autopsies. The HBV antigens and antibodies were determined by commercially available radioimmunoassays or enzyme immunoassays (Abbott Diagnostics). The HBV DNA was determined by Southern blot analyses (*11*); IgM, immunoglobulin M.

Patient	Infection or diagnosis	Serum markers			Liver markers		Liver HBV DNA	
		HBsAg	HBeAg	Anti-HBc (IgM)	HBc/cAg	HBsAg	Free	Integrated
1	Acute	+	+	+	+	+	+	−
2	Acute*	−	−	+	−	−	−	+
3	PLC†	+	−	−	−	+	−	+
4	Acute	+	+	+	+	+	+	−
5	PLC†	+	−	−	−	+	+	+
6	Acute	−	−	+	−	−	+	−
7	Cirrhosis	+	−	−	−	+	−	+
8	Cirrhosis				−	−	−	−
9	PLC†	+			−	−	−	−
10	PLC†	+	−	+	−	−	−	+
11	Acute*	+	−	+	−	−		

*Fulminant hepatitis. †Primary liver carcinoma.

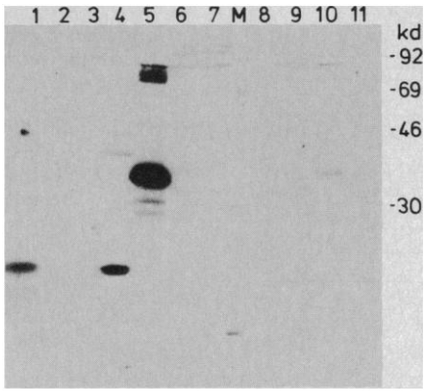

Fig. 2. Immunoblot analysis of cell extracts of HBV-infected liver tissues with an antiserum to HBc. The proteins of liver cell extracts (*11*) (Table 1) were immunoblotted with a rabbit antiserum to HBc (c2-183). ^{14}C-Labeled size markers were cytochrome (14 kD), carbonic anhydrase (30 kD), ovalbumin (46 kD), bovine serum albumin (69 kD), and phosphorylase B (92 kD).

The amino acid sequence could extend up to amino acid 700 of the pol frame (Fig. 1C). In agreement with this, an antiserum to a pol protein covering amino acid sequence 170 to 376 react with the 69-kD but not with the 35-kD polypeptide. The 92- and the 38-kD polypeptides were not characterized in detail.

C-pol fusion polypeptides but no normal 21-kD HBcAg protein were found in two PLC tissues (Figs. 2 and 3A), and more recently also in a third PLC tissue. In the productively infected livers the 21-kD HBcAg protein and, in one case, an additional minor putative c-pol fusion protein were detected (Figs. 2 and 3B). The absence of the 21-kD HBcAg protein in the PLC tissues implies that no viral core particles are present in these cells. If we assume that similar c-pol proteins are also produced in the normal replication cycle but are more rapidly packaged and processed in viral cores destined to be exported, accumulation of the c-pol fusion proteins may be a consequence of the absence of core particles in the PLC tissue. The c-pol transcripts in the PLC tissues are probably produced from integrated DNA because this was common to both PLC tissues studied in detail. Integration into the HBV C gene could account for the changed C gene expression. This could explain the altered start site of the 35-kD protein within the C gene, the lack of the 21-kD HBcAg, and overexpression of c-pol fusion proteins.

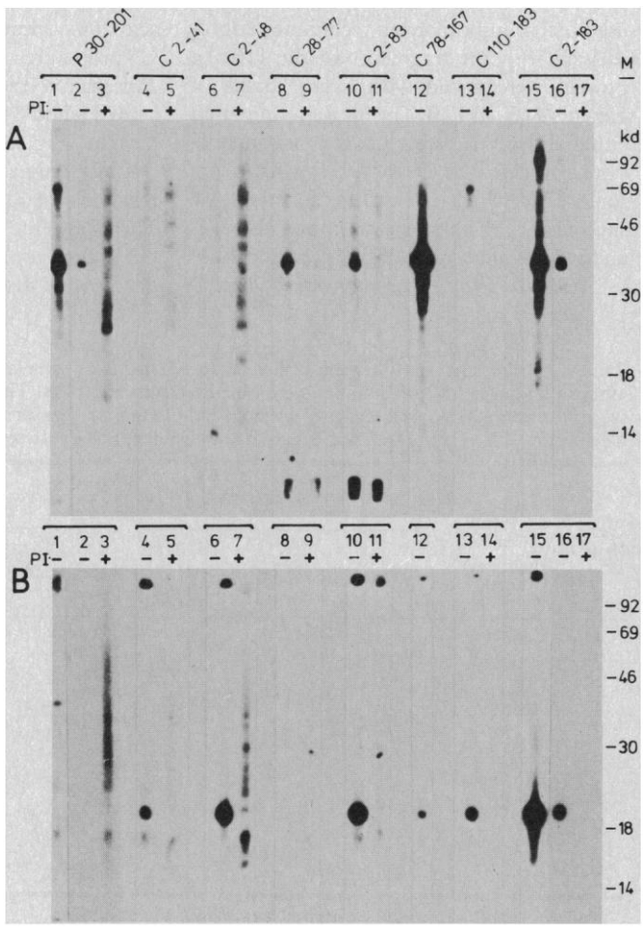

Fig. 3. Immunoblot analyses of liver extracts with antisera to known and predicted HBV proteins. Immunoblot analyses of liver extracts of patient 5 (A) and 4 (B) performed with antisera to HBc and pol protein sequences delimited by the amino acid position indicated [(*8*) and Fig. 1] on top of each lane. PI + or − indicates whether the antiserum had been first incubated with the corresponding antigen. All lanes are the same exposure except lanes 1 and 15 which were longer exposed. Unspecific bands observed only in the PI + lanes are due to the fact that only partially purified antigens were used for the competition experiment. HBcAg-specific signals on top of the gel in (B) could be caused by aggregated HBcAg.

As the C gene and the pole frame are in different reading frames, and as there is no evidence for mRNA splicing in HBV, we favor ribosomal frameshifting as the mechanism for fusion protein synthesis. Alternatively, c-pol fusion protein expression might also be due to a frameshift mutation in the corresponding DNA template, but the detection of 35-kD fusion proteins in three PLC tissues argues against this possibility.

The 35-, 38-, and 69-kD c-pol fusion proteins could be the result of incomplete synthesis or processing of a HBV reverse transcriptase precursor protein. If the fusion proteins detected also reflect the presence of enzymatically active reverse transcriptase synthesized from a common precursor, the enzyme would not be encapsulated in viral particles in the PLC tissues. Consequently, one intriguing possibility would be that reverse transcription of cellular genes may occur, eventually leading to cellular transformation as speculated for retroviral reverse transcriptase (*7*).

REFERENCES AND NOTES

1. R. Cattaneo, R. Sprengel, H. Will, H. Schaller, in *Primers in Developmental Biology*, G. Malacinski, Ed., (Macmillan, New York, in press); W. S. Robinson, R. H. Miller, L. Klote, P. L. Marion, S. C. Lee, in *Viral Hepatitis and Liver Disease*, G. N. Vyas, J. L. Dienstag, J. H. Hoofnagle, Eds. (Grune & Stratton, Orlando, FL, 1984).
2. R. Cattaneo, H. Will, H. Schaller, *EMBO J.* **3**, 2191 (1984); R. Cattaneo, H. Will, N. Hernandez, H. Schaller, *Nature (London)* **305**, 336 (1983); M. Büscher, W. Reiser, H. Will, H. Schaller, *Cell* **40**, 717 (1985).
3. R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds., *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1982); H. E. Varmus, *Nature (London)* **314**, 583 (1985).
4. H. Toh, H. Hayashida, T. Miyata, *Nature (London)* **305**, 827 (1983).
5. H. Will, R. Cattaneo, E. Pfaff, C. Kuhn, M. Roggendorf, H. Schaller, *J. Virology* **50**, 335 (1984).
6. S. J. Hadzyannis *et al.*, *Hepatology* **3**, 656 (1983).
7. D. Baltimore, *Cell* **40**, 481 (1985).
8. Antisera were prepared in rabbits with fusion proteins expressed in *Escherichia coli*. Protein c2-183 contained eight amino acids of the β-galactosidase protein and amino acids 2 to 183 of the C gene of HBV (provided by P. Wingfield, Biogen, SA, Geneva). All other viral proteins (J. Salfeld and E. Pfaff, unpublished) contained the first 99 amino acids of the MS2-polymerase from vector pPlc 24 [Remaut *et al.*, *Gene* **15**, 81 (1981)], or derivatives thereof (E. Beck, unpublished data). For immunization, the MS2 fusion proteins were purified according to Küpper *et al.* [*Proceedings of the Fourth International Symposium on Genetics of Industrial Microorganisms* (Japan, 1982)], p. 222.
9. K. Takahashi *et al.*, *J. Immunol.* **130**, 2903 (1983).
10. M. Pasek *et al.*, *Nature (London)* **282**, 575 (1979).
11. Liver cell extracts were prepared by homogenization in a mixer (Sorvall) in phosphate-buffered saline (0.14*M* NaCl, 10 m*M* sodium phosphate, *p*H 7.4) including 0.1 percent NP-40. The HBV DNA of the liver cells was extracted [R. Sprengel *et al.*, *J. Virol.* **52**, 932 (1984)] and analyzed by Southern blot analyses [T. Maniatis, *Molecular Cloning*, Cold Spring Harbor Laboratory (Cold Spring Harbor, NY, 1982)] with cloned and ^{32}P-labeled nick-translated HBV DNA as a probe.

12. We thank W. Keller, A. Krämer, K. Mölling, L. Phillipson, and G. Schütz for reading the manuscript. Supported by the Deutsche Forschungsgemeinschaft, Forschergruppe Genexpression.

4 June 1985; accepted 30 September 1985