Hepatitis B Virus Particles Contain a Polypeptide Encoded by the Largest Open Reading Frame: A Putative Reverse Transcriptase

DAVID H. MACK,1 WILL BLOCH,2 NRAPENDRA NATH,3 AND JOHN J. SNINSKY1*

Department of Infectious Diseases1 and Department of Protein Chemistry,2 Cetus Corporation, 1400 Fifty-Third Street, Emeryville, California 94608, and Blood Services Laboratories, American Red Cross, Bethesda, Maryland 200143

Received 28 March 1988/Accepted 29 August 1988

A segment of the largest open reading frame of hepatitis B virus (HBV) was inserted into an open reading frame vector directing the expression in Escherichia coli of a fusion molecule containing 143 HBV-encoded amino acids. The fusion protein was used to generate antiserum which served in immunoblots to identify a polypeptide with a molecular mass of 65 kilodaltons in HBV particles. Because of the small number of molecules in virus particles, unambiguous detection required the development of a highly sensitive immunoblot procedure.

The largest open reading frame (ORF P) of human hepatitis B virus (HBV) could encode a polypeptide containing approximately 845 amino acids; its size alone led Calibret et al. (4) to suggest that it is the gene for the DNA polymerase activity typically identified in virus particles (5). The other members of the hepatadnaviruses isolated from woodchucks (3), ground squirrels (10), and ducks (7) contain a similarly sized and positioned ORF. Summers and Mason (14) initially demonstrated that duck hepatitis B virus replication proceeds through a reverse transcription step in immature cytoplasmic nucleocapsids. Other laboratories have extended this observation to the mammalian viruses (8, 11). Accordingly, the enzyme responsible for DNA polymerase activity in mature extracellular particles may also be able to use RNA as a template. Consistent with this proposal was a report by Toh et al. (15) that a region of the polypeptide predicted by ORF P shows limited but striking similarity in amino acid sequence to the reverse transcriptases of Rous sarcoma virus and murine leukemia virus. This report describes the cloning and bacterial expression of the internal segment of the putative hepatadnaviral polymerase, previously shown to be conserved among elements which replicate through an RNA intermediate, as a portion of a tripartite fusion protein. This protein was used to generate rabbit antiserum which permitted the identification of an ORF P gene product within virus particles. Unambiguous confirmation that such proteins exist within virus particles, presumably because of the low copy number of the proteins, required the development of a highly sensitive detection procedure for immunoblotted proteins.

Efforts to increase the sensitivity of solid-phase nonisotopic detection for Southern blot analysis by using horseradish peroxidase have benefited dramatically by the use of tetramethylbenzidine (TMB) rather than diaminobenzidine (DAB) as a substrate (12, 13; W. Bloch and D. Birch, unpublished results). We chose to determine initially the quantitative advantage that TMB provides in Western blots (immunoblots) since the number of copies of the viral polymerase polypeptide(s) in a virus particle could be extremely low. Two identical immunoblots (16) were generated which contained a dilution series of purified aminoglycoside phosphotransferase (3')I [APH(3')I] (Fig. 1).

Development of immunoblots with DAB (Sigma Chemical Co.) was carried out as previously described (2). For TMB-based detection, washed filters were agitated in fumarate buffer (100 mM sodium fumarate, 10 mM EDTA [pH 5.4]) (fumarate buffer) for 5 min. Blots were then incubated with fumarate buffer containing 1% dextran sulfate for 15 min with agitation, followed by a 5-min wash with fumarate buffer alone to remove bulk dextran sulfate. Next, the buffer was removed and replaced with freshly made substrate solution (2.5 ml of a 0.1-mg/ml TMB [ICN Biologicals] stock in 100% ethanol per 50 ml of fumarate buffer) for 15 min with agitation. Protein bands were visualized by the addition of hydrogen peroxide to a final concentration of 0.0015%; after development for 5 min, blots were gently agitated with five changes of distilled H2O. DAB and TMB provide for the detection of 5 ng and 125 pg of electrophoresed protein, respectively. The 40-fold increased sensitivity provided by TMB over DAB allows greater dilution of primary and secondary antibodies, thereby decreasing the background caused by weak cross-reactivity to unrelated polypeptides.

FIG. 1. Comparative immunoblot sensitivity with alternate peroxidase substrates. Amounts of purified APH(3')I lanes 1 through 8, 30 ng, 15 ng, 5 ng, 625 pg, 125 pg, 25 pg, 5 pg, and 0 pg, respectively) were loaded on two sodium dodecyl sulfate–12.5% polyacrylamide gels. Following fractionation, the gels were electrobotted onto nitrocellulose membranes (16). The blots were treated identically with a 1:3,000 dilution of anti-APH(3')I serum and a 1:10,000 dilution of goat anti-rabbit immunoglobulin horseradish peroxidase conjugate (Bio-Rad Laboratories). Immunoblots were developed with either DAB (panel A) or TMB (panel B) (see text).
Immunoblots used to monitor patient exposure to various pathogens and to confirm the enzyme-linked immunosorbent assays which measure antibodies to the recently identified exogenous human retroviruses may benefit from peroxidase-linked detection in which the chromogenic substrate is TMB.

The open reading frame (ORF) expression vector chosen for this study (pTACNEO 2.1) is one of a series of nine vectors specifically designed to allow expression of DNA fragments of any translational reading frame containing single-stranded termini with the sequence GATC (S. Kwok, unpublished results). Briefly, transcription is regulated by the hybrid TAC promoter and efficient translation is provided by the ribosome-binding site of the β-galactosidase (β-Gal) gene and eight N-terminal codons. The β-Gal codons are followed by a polylinker region containing a unique BamHI endonuclease site and the entire gene for a modified 262-amino-acid APH(3′)I (F. C. Lawyer, F. C. Stoffel, and D. H. Gelfand, Abstr. Mol. Biol. Yeast 1983, p. 190). In summary, these coding segments allow the production of in-frame tripartite (tribrid) proteins containing 8 amino acids of β-Gal at the N terminus, the amino acids encoded by an inserted DNA fragment, and the 262 amino acids encoded by the antibiotic resistance gene at the C terminus (Fig. 2).

The HBV (adw2) Sau3A DNA fragment chosen for insertion into an expression vector has the potential to encode 143 amino acids corresponding to codons 502 to 644 of ORF P (17). This fragment contains the region shown to be similar in amino acid sequence to the largest predicted polypeptides of the other hepadnaviruses and reverse transcriptases of retroviruses (15; D. H. Mack and J. J. Sninsky, Proc. Natl. Acad. Sci. USA, in press). An additional codon is contributed because the linker region of the vector encodes the same residue as the viral sequence at position 645 (proline). Since the HBV fragment chosen for expression lacks other interrupted reading frames, the production of a fusion protein of the expected size unambiguously defines the amino acids directed by the inserted DNA as encoded by ORF P.

Protein synthesized by the parent and recombinant plasmids with the HBV insert in the correct orientation for expression, before and after induction of transcription with isoprropyl β-D-galactopyranoside, was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 3, panel A) and immunoblot analysis (16) by using rabbit antiserum to the portion of the fusion protein encoded by the bacterial APH(3′)I gene (Fig. 3, panel B). The rabbit antiserum recognized a major polypeptide of the predicted molecular mass of ca. 45 kilodaltons (kDa) before induction (Fig. 3, panel B, lane 3) and with increased intensity after induction (Fig. 3, panel B, lane 4). The detected protein of lower molecular mass after induction was attributed to proteolytic degradation. A band migrating at ca. 29 kDa was

![Diagram of vector construction](image-url)
also identified by this analysis from induced bacteria harboring the parent expression vector (Fig. 3, panel B, lane 2) and bacteria carrying the recombinant plasmid (Fig. 3, panel B, lanes 3 and 4). This polypeptide is thought to be the inefficiently expressed, truncated APH(3')I translated from the internal ATG. Neither the ca. 45-kDa or 29-kDa bands were identified by preimmune serum (for example, see Fig. 4, panel A). After sonication and centrifugation of bacteria carrying the recombinant plasmid (pHBpoldDNEO-1), the same two proteins were present in the pellet (Fig. 3, panel B, lane 6) but not in the supernatant (Fig. 3, panel B, lane 5). After preparative sodium dodecyl sulfate-polyacrylamide gel electrophoresis purification (19), the tribrid fusion protein represented the only readily detectable band by Coomassie blue staining (Fig. 3, panel A, lane 7).

The purified fusion protein was used to generate rabbit antiserum. With the TMB-based detection system, the rabbit anti-tribrid serum clearly identified a protein from virus particle preparations migrating at ca. 65 kDa (Fig. 4, panel C, lane 1). The polypeptide recognized by the generated antiserum is similar in molecular mass to one of two DNA polymerase-reverse transcriptase activities (~70 kDa and ~90 kDa) in virus particles from a differentiated human hepatoma cell line transfected with genomic HBV DNA (1). Bacteria harboring a recombinant plasmid that directs the synthesis of a hybrid fusion protein containing eight amino acids of β-Gal at the N terminus and 262 amino acids of APH(3')I at the C terminus (directed by pTACNEO 1.1) served as a source for control hybrid protein (Fig. 4, panels A through E, lanes 3). Neither DAB visualization of horse-radish peroxidase-linked secondary antibody nor autoradiographic visualization of iodinated protein A had the sensitivity needed to detect the ORF P-encoded polypeptide in virus particles with this antiserum. This band was not recognized by preimmune serum from the same rabbit (Fig. 4, panel A, lane 1) or by antiserum to APH(3')I (Fig. 4, panel B, lane 1). Uninfected human sera prepared and probed in a manner similar to that for sera containing HBV particles failed to generate the ca. 65-kDa band (data not shown). Preadsorption of antiserum with various antigens in a manner previously described (9) confirms that the recognition of the ca. 65-kDa protein in virus particle preparations by the generated antiserum is specifically related to the ORF P determinants within the molecule. Anti-tribrid serum preincubated with the tribrid protein no longer recognized the viral polypeptide (Fig. 4, panel D, lane 1), tribrid, or hybrid proteins (Fig. 4, panel D, lanes 2 and 3, respectively). Preincubation of the anti-tribrid serum with an equivalent concentration of the hybrid protein did not block recognition of the viral polypeptide or tribrid fusion protein (Fig. 4, panel E, lanes 1 and 2, respectively) but eliminated recognition of the hybrid protein (Fig. 4, panel E, lane 3).

If the HBV ORF P region was transcribed and translated in its entirety, a 92-kDa polypeptide would result. The identification of a ca. 65-kDa protein suggests that the primary translation product is proteolytically cleaved or that the detected polypeptide results from translation initiation at one of the multiple internal ATGs or is translated from a spliced mRNA. The precedent of proteolytic maturation of the reverse transcriptase of the related retroviruses, and preliminary results by Will et al. (18) suggesting the presence of core-meramerase fusion proteins in extracts of hepatocellular carcinoma-containing HBV sequences, would perhaps favor proteolytic processing as the likely maturation event. The central location within the ORF of the ORF P segment expressed in the tribrid protein prevents a determination of whether the protein identified from virus particle preparations is truncated at either or both predicted termini. Preliminary experiments with polyclonal sera generated to the region encompassed by codons 34 to 314 of ORF P expressed as a tripartite protein suggest that the ca. 65-kDa protein lacks the extreme N-terminal residues predicted by the ORF (unpublished results).

Antiviral strategies designed to intercede in persistent
viral infection will benefit from the characterization of proteins required for viral replication. Since precursor polypeptides often are functionally inactive, our demonstration of the size of the HBV particulate polymerase (and perhaps reverse transcriptase) should assist in the eventual production of enzymatically active polypeptides in heterologous or homologous expression systems. In addition, the sequence similarity of the ORF P region used in the tribrid fusion protein to the other hepadnaviruses and the reverse transcriptades of retroviruses suggest that the anti-tribrid serum may recognize the reverse transcriptase of other viruses.

We kindly thank E. Schaeffer for critical discussions, D. Birch and R. Goodson for helpful discussions concerning the optimization of TMB-based detection of Southern blots prior to publication, S. Kwok for providing pTAACEO ORF vectors, S. Stoffel and D. Gelfand for providing purified APH(3')I protein and rabbit polyclonal anti-APH(3')I sera, S. Nilson and E. Ladner for preparation of the figures, and R. Bengelsdorf for excellent manuscript preparation.

The research described in this report was supported by Cetus Corporation.

LITERATURE CITED

