JOURNAL OF VIROLOGY, May 1984, p. 563–571
0022-538X/84/050563-09$02.00/0

Vol. 50, No. 2

Transcription of the Hepatitis B Surface Antigen Gene in Cultured Murine Cells Initiates Within the Presurface Region

DAVID N. STANDRING,[1] WILLIAM J. RUTTER,[1] HAROLD E. VARMUS,[2] AND DON GANEM[2,3]*

Departments of Biochemistry and Biophysics,[1] Microbiology and Immunology,[2] and Medicine,[3] University of California, San Francisco, California 94143

Received 4 October 1983/Accepted 7 February 1984

Cloned hepatitis B virus (HBV) DNA directs the synthesis of the viral surface antigen (HBsAg) when introduced into mouse L cells by DNA transformation. We have used recombinants between the Rous sarcoma virus long terminal repeat and subgenomic fragments of HBV DNA to localize regions of the HBV genome required for HBsAg expression. Examination of HBV-specific RNA from such transformants indicates that transcription initiates at three distinct sites (153, 163, and 183 nucleotides upstream from the translation initiation codon for mature HBsAg). Thus in these cells, a large segment of the presurface reading frame is not represented in HBsAg mRNA. The termination site of this RNA lies within the coding sequences for the viral core antigen, some 1,094 ± 10 base pairs downstream from the TAA stop codon for HBsAg. Two additional open reading frames are present in the resultant unspliced HBsAg RNA.

Hepatitis B virus (HBV) is a small DNA virus of humans which causes acute and chronic hepatitis and is strongly associated with the development of hepatocellular carcinoma (9). Little is known about the replicative and transcriptional program of HBV, owing to the inability to propagate the virus in cultured cells. Nevertheless, considerable recent progress has been made in elucidating the structure of the HBV virion and its genome (see reference 25 for a review). This work has shown that the virion is a 42-nm particle whose outer coat is comprised principally of a single protein, known as hepatitis B surface antigen (HBsAg), a 24,000-dalton polypeptide which exists in both a glycosylated and a nonglycosylated form.

Sequence analysis of cloned HBV DNA (8, 16, 26) has revealed that the coding region for HBsAg is only part of a much larger open translational reading frame. Immediately 5′ to the ATG which encodes the first amino acid of HBsAg is a 522-base pair (bp) translatable region known as preS (for presurface) (see Fig. 1). This region is preceded by the sequence TATATAAG, which has been suggested to function as part of the promoter governing HBsAg transcription (18). Thus the DNA sequence raises the possibility that HBsAg might be derived by cleavage of a much larger protein generated by transcription from this putative preS promoter and translation of the entire preS region.

Although HBV replication has not yet been achieved in cell culture, introduction of cloned HBV DNA into cultured cells does result in limited viral gene expression (2, 4, 11, 18). Dubois et al. (4) observed that mouse L cells transfected with head-to-tail dimers of HBV DNA synthesized small amounts of HBsAg; monomeric genomes interrupted at the unique *Eco*RI site within preS did not direct HBsAg biosynthesis. In this paper, we report detailed studies of the transcription of the HBsAg gene in the mouse L cell system. Our results show that, in contrast to earlier suggestions (10, 18), a large section of the preS region is not represented in stable HBsAg mRNA and that transcription of HBsAg sequences presumably occurs from a promoter located within the preS region.

* Corresponding author.

MATERIALS AND METHODS

Cells, DNA, and RNA. The growth of LTK$^-$ cells, $CaCl_2$-mediated cotransformation with cloned HBV DNA and herpes simplex virus thymidine kinase (TK) DNA, and selection of TK$^+$ transformants were performed as previously described (4). TK$^+$ foci were tested for receipt of HBV DNA by Southern blotting and 5 to 8 HBV-positive recipient clones from each transfection were assayed for HBsAg production, using the commercial AUSRIA II solid-phase radioimmunoassay kit (Abbott Laboratories, North Chicago, Ill.). In this test, results are expressed as the ratio P/N, where P = ^{125}I-labeled anti-HBsAg counts per minute bound to anti-HBsAg-coated beads after preincubation with a test sample and N = ^{125}I-labeled anti-HBsAg counts per minute bound to similar beads which had been preincubated with HBsAg-negative serum. Values of P/N greater than two are considered positive. As previously reported (4), independent clones bearing the same HBV plasmid vary over a 10- to 20-fold range in P/N ratios when tested in this way. Serial dilution experiments demonstrate that the P/N ratio is not linearly related to HBsAg concentration at the levels produced by most clones, so that assay results are only a semiquantitative reflection of the amounts of antigen produced.

Plasmid DNA was extracted by the method of Birnboim and Doly (1) and purified by ethidium bromide-cesium chloride density gradient centrifugation before use. Restriction enzymes were purchased from New England Biolabs (Beverly, Mass.) and used according to the instructions of the manufacturer. Polyadenylated RNA was extracted from cell monolayers as described by Payne et al. (17); RNA (2 μg per lane) was analyzed by agarose gel electrophoresis in 2.2 M formaldehyde followed by transfer to nitrocelluose paper, by the method of Thomas (24).

Hybridization probes (Fig. 1A) were prepared by appropriate restriction cleavages, followed by recovery of the indicated fragment from agarose gels. The DNA fragments were radiolabeled with [^{32}P]dCTP with random calf thymus DNA primers and avian myeloblastosis virus polymerase, as previously described (17). Long terminal repeat (LTR) probes specific for the U5 and U3 regions were prepared as described by Payne et al. (17).

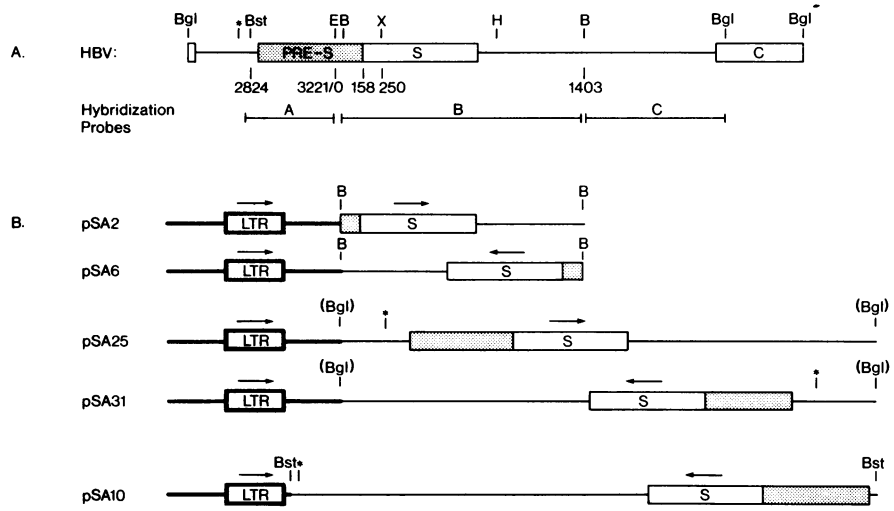

FIG. 1. (A) Schematic map of the HBV genome. PRE-S, Presurface reading frame; S, HBsAg coding region; C, HBV core antigen coding region. *, Position of putative promoter sequences previously suggested to direct initiation of HBsAg transcription (18). Nucleotide positions of key sites (in accordance with the nucleotide numbering scheme discussed in reference 22) are indicated below the map, as are the structures of the fragments used as radiolabeled probes for the hybridization studies in Fig. 2B. Abbreviations: B, *Bam*HI; Bgl, *Bgl*II; Bst, *Bst*EII; E, *Eco*RI; X, *Xba*I; and H, *Hpa*I. (B) Schematic maps of RSV-HBV recombinant plasmids. In all cases, the indicated sequences are flanked by 3.5 kb of pBR322 DNA. Arrows, Transcriptional orientation of the underlying segment. Heavy line, RSV sequences; remaining sequences are from HBV. Stippled area, preS sequences. Parentheses surround *Bgl*II sites of HBV which are ligated to plasmid *Bam*HI sites.

Plasmid construction. The HBV genome used in all experiments was that cloned at its unique *Eco*RI site by Valenzuela et al. (26). This plasmid is termed pHBV3.2. The construction of the Rous sarcoma virus (RSV) LTR-containing plasmid pAV-2 has been previously described (13). This plasmid bears an 850-bp insert of RSV DNA which includes the entire LTR, flanked on its 5′ side by 260 bp of RSV DNA from the *src* region and on its 3′ side by 260 bp of RSV sequences which terminate 8 nucleotides before the *gag* ATG initiation codon. The *Sau*IIIA site normally found at this position was converted to a unique *Bam*HI site by linker insertion (P. Luciw, unpublished data). To prepare pSA2 and pSA6, the 1.4-kilobase (kb) HBV *Bam*HI fragment was cleaved from pHBV3.2 by *Bam*HI, gel purified, and inserted into the *Bam*HI site of pAV2. To prepare pSA25 and pSA31, the entire HBV genome was excised from pHBV3.2 with *Eco*RI, purified by agarose gel electrophoresis, and self-ligated with T4 DNA ligase at a DNA concentration of 100 μg/ml. A sample of the ligation products was then cleaved with *Bgl*II and cloned into the *Bam*HI site of pAV2. For construction of pSA10, a sample of the same ligation mix was cleaved with *Bst*EII and cloned into the unique *Bst*EII site of pAV2.

For construction of HBV-containing plasmids lacking an LTR, a pBR322 derivative was employed which bears a 140-bp herpes simplex virus DNA insertion at its *Bam*HI site; the inserted DNA harbors a unique *Bgl*II site. Into this site was cloned a *Bgl*II-cleaved sample of the above ligation mix. One resulting recombinant, pSA90, proved to contain the entire (3.2-kb) HBV genome permuted at the HBV *Bgl*II site at position 2,432. The deletion derivative pSA90ΔB10 was generated by opening the plasmid at its unique *Bst*EII site within preS, exposing the DNA to nuclease Bal31, and religating the resulting DNA. Plasmids were re-treated with *Bst*EII and then recloned in *Escherichia coli* HB101; plasmids shorter than pSA90 were detected by gel electrophoresis of DNA from crude lysates of transformants. Plasmid pSA90ΔB10 harbors a 550-bp deletion centered around the *Bst*EII site.

S1 mapping and primer extension. The methodology for the production of 5′-end-labeled probes, the conditions for the RNA-DNA annealing, and the subsequent S1 digestion reactions have been described in detail previously (5, 12, 22). The 3′-end-labeled probes were labeled by a fill-in reaction with avian myeloblastosis virus polymerase to add [^{32}P]dGTP to probe 3 (*Bam*HI site) and [^{32}P]dCTP to probe 4 (*Xba*I site). Hybridizations were performed at 50°C for the 5′-end-labeled probes and 49°C for the 3′ probes. Further details can be found in the legends to Fig. 3, 5, and 6. The single-stranded synthetic 20-nucleotide primer was generously provided by P. J. Barr, Chiron Corp., Emeryville, Calif. The 31-nucleotide *Bam*HI-*Eco*RI primer was prepared from pHBV3.2. The plasmid was cleaved with *Bam*HI, end-labeled with [α-^{32}P]ATP, and then cleaved with *Eco*RI. After extraction with phenol and ethanol precipitation, purification and strand separation were accomplished simultaneously on a 15% polyacrylamide–8.3 M urea gel. The conditions used for the primer extension reactions are given in the legend to Fig. 4. The products from S1 or primer extension reactions were denatured in 90% formamide containing tracking dyes and resolved on 5% polyacrylamide–8.3 M urea gels which were dried and autoradiographed for 1 to 3 days at −70°C with an intensifying screen and X-Omat AR film. DNA sequencing reactions were carried out as described by Maxam and Gilbert (15).

RESULTS

Expression of HBsAg in L cells. To survey the HBV genome for regions important for HBsAg expression, we constructed a series of recombinant plasmids in which HBV DNA fragments bearing the HBsAg coding region are cloned (in both orientations) downstream from a transcriptionally active eucaryotic promoter in the LTR of RSV. If the HBV fragment carries all the signals required for surface antigen biosynthesis, cells transfected with the LTR-HBV recombinant should express HBsAg irrespective of the orientation of the HBV fragment to the LTR. However, if the HBV fragment lacks the HBsAg promoter, the LTR-HBV chimera

will direct HBsAg expression only when the HBV segment and the LTR are in the same transcriptional orientation.

Three HBV fragments were subcloned 3′ to the LTR in plasmid pAV2, which contains a single copy of the RSV LTR (Fig. 1). The first was a 1.4-kb *Bam*HI fragment bearing the HBsAg structural gene, 127 bp of preS sequences, and 568 bp of 3′ noncoding sequences (26). In plasmid pSA2, the HBsAg gene is in the correct transcriptional orientation relative to the LTR, whereas in pSA6 it is in the opposing orientation. These plasmids were introduced into LTK^- cells by cotransfection with cloned *Herpesvirus* TK DNA, and TK^+ foci were picked and assayed for HBsAg by solid-phase radioimmunoassay. Plasmid pSA6 does not direct HBsAg synthesis, whereas the presence of pSA2 leads to the synthesis of small amounts of immunoreactive material (Table 1). Control experiments documented that chromosomal DNA of cells transformed with pSA6 DNA contained multiple integrated copies of the recombinant plasmid (data not shown).

Next, a 2.8-kb *Bgl*II fragment of HBV DNA was inserted in both orientations into pAV2. This fragment includes the entire preS region and 424 bp of 5′ noncoding DNA, as well as 1,148 bp 3′ to the termination codon for HBsAg. Plasmid pSA25 bears the insert in the same transcriptional orientation as the LTR, whereas pSA31 bears the same insert in the opposing orientation. Both plasmids direct the synthesis and secretion of HBsAg in readily detectable amounts (Table 1). Analysis of the protein product of lines 25b and $31e_2$ by [^{35}S]methionine radiolabeling and immunoprecipitation shows that both cell lines are producing approximately equal amounts of antigen of normal size (24 kDa); both glycosylated and nonglycosylated forms are detectable (D. Ganem and H. E. Varmus, unpublished data).

TABLE 1. HBsAg synthesis[a]

Plasmid	Cell line	HBsAg (P/N)
pSA2	$2a_2$	5.6
	2b	3.0
pSA6	$6a_1$	<2.0
	$6b_2$	<2.0
pSA25	25a	38.5
	25b	147.0
	25e	58.4
pSA31	$31b_2$	145.4
	$31d_1$	100.8
	$31e_2$	125.0
pSA10	$10a_1$	81.6
	$10a_2$	58.2
	$10a_3$	143.7
pSA90	$90a_1$	10.5
	90b	68.4
pSA90ΔB10	$B10a_2$	45.0
	B10c	15.5

[a] LTK^- cells were cotransfected with 10 ng of HSV TK DNA, 100 ng of the indicated HBV plasmids, 10 μg of carrier salmon sperm DNA, and TK^+ foci selected in hypoxanthine-aminopterin-thymidine medium. Individual colonies were picked and expanded into 60-mm dishes. At 48 h after the cells reached confluence, 0.2 ml of medium was tested for HBsAg by solid-phase radioimmunoassay (AUSRIA II) as described in the text. Results of assay of several representative independent transformants from each experiment are shown. All TK^+ foci were tested for receipt of HBV DNA sequences by Southern blotting of *Hin*dIII-digested chromosomal DNA with [^{32}P]HBV DNA probes. All clones listed above were found to contain multiple copies of integrated HBV DNA.

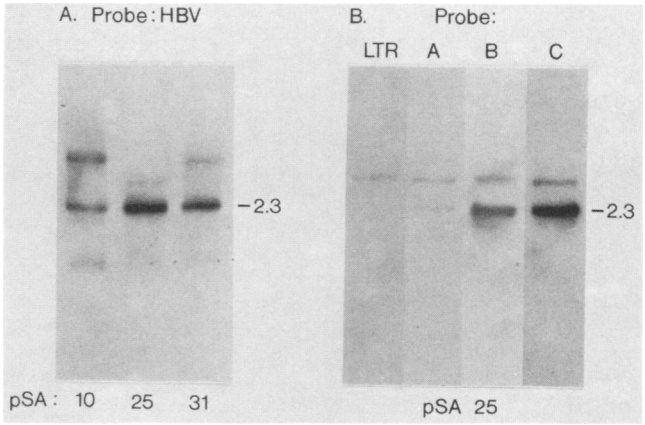

FIG. 2. (A) Northern blot analysis of HBV-specific RNA from transfected L cells. Two micrograms of polyadenylated RNA from cell lines $10a_2$ (left lane), 25b (center lane), and $31e_2$ (right lane) were electrophoresed through 1% agarose–2.2 M formaldehyde gels, transferred to nitrocelluose paper, and hybridized to radiolabeled HBV DNA. After being washed at 50°C in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate for 2 h, the filters were dried and subjected to autoradiography with Kodak X-Omat R film; exposure time was 4 days. Sizes were calibrated by comparison to rRNA bands visualized with ethidium bromide staining and murine leukemia virus RNAs electrophoresed in parallel. (B) Analysis of 25b RNA by hybridization with the radiolabeled subgenomic probes illustrated in Fig 1A. Two micrograms of polyadenylated 25b RNA was electrophoresed in parallel lanes of a 1% agarose–formaldehyde gel and transferred to nitrocellulose as in Fig. 2A. The filter was then cut into strips, and each strip was separately annealed to the radiolabeled probe indicated above the corresponding lane. Probe designated LTR is specific to the U5 region of the RSV LTR. Autoradiography was performed as in Fig. 2A.

Lastly, the 3.2-kb HBV genome was cloned at its unique *Bst*EII site (nucleotide 2,824) 3′ to the LTR in pAV2; this site is 32 bp 5′ to the first ATG of preS, and separates the preS ATG from the TATA-like sequence some 40 bp further upstream which is part of the putative promoter preceding preS. Because not all *Bst*EII sites have identical termini, only the recombinant plasmid bearing the opposing orientation (pSA10) was recovered. In plasmid pSA10, this putative preS promoter is separated from the HBsAg and preS region by a 3.5-kb insertion of plasmid pAV2 DNA, and the RSV LTR is not in the correct orientation to direct HBsAg expression (Fig. 1). Nevertheless this plasmid efficiently directs HBsAg biosynthesis and secretion (Table 1).

RNA transcripts in HBsAg-producting cells. Polyadenylated RNA was prepared from cells transformed with each of the plasmid DNAs listed and analyzed by agarose gel electrophoresis and hybridization to [^{32}P]HBV DNA. No HBV-specific RNA was detectable in cells harboring multiple inserts of pSA6 DNA. By contrast, cells transfected with pSA2 DNA (clone 2b) contain two species of HBV-specific RNA in low abundance; both species also anneal to probes from the U5 (but not the U3) region of the RSV LTR (data not shown). These results suggest that in pSA2, all transcription of HBV sequences is promoted by the LTR; when the LTR is unavailable (as in pSA6), no transcription of HBV DNA occurs.

Figure 2A shows the HBV-specific polyadenylated RNA

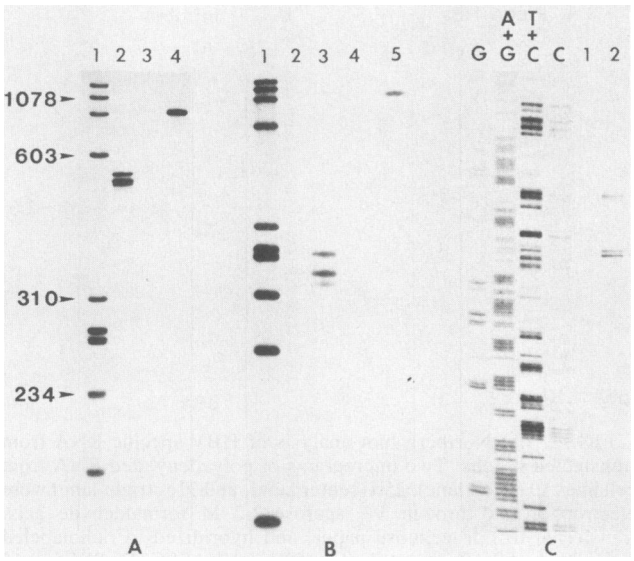

FIG. 3. Autoradiograms showing the S1 analysis of the 5′ end of the HBsAg RNA. (A) *Rsa*I-*Bst*EII probe (probe 2 in Fig. 7). Lane 1, DNA size markers; lanes 2 and 3, S1-resistant products of annealing with 5 μg of polyadenylated RNA from the cell line containing pSA25 (lane 2) and control yeast tRNA (lane 3); and lane 4, undigested probe. (B) *Xba*I-*Bgl*II probe (probe 1 in Fig. 7). Lane 1, DNA markers; lanes 2 through 4, S1-resistant products of annealing to about 2 μg of polyadenylated RNA from cell line 25b (lane 2), to about 6 μg of polyadenylated 25b RNA from a different RNA preparation (lane 3), and control yeast tRNA (lane 4); and lane 5, undigested probe. (C) Alignment between the S1-protected species (lanes 1 and 2) and a sequencing ladder generated from the same probe (*Xba*I-*Bgl*II). The positions of the S1-protected species were read directly against the sequence and are presented in Fig. 7B. The DNA sequencing reactions used are indicated above the respective lanes. The DNA size markers shown in (A) and (B) are ^{32}P-end-labeled, *Hae*III-digested φX174 DNA with sizes of 1,353, 1,078, 872, 603, 310, 281, 271, 234, 194, and 118 bp. The conditions for the formation of RNA-DNA hybrids were as described previously (12, 22). Each experiment contained 5,000 to 10,000 Cerenkov cpm of probe and a total of 20 μg of RNA (experimental RNA plus yeast tRNA carrier). The control experiments contained only yeast tRNA (20 μg). S1 nuclease digestion was performed for 1 h at 37°C with 500 U of enzyme per ml in a volume of 220 μl. The reactions were extracted with phenol, and the protected fragments were precipitated with ethanol and then denatured in 90% formamide containing tracking dyes. Approximately one-third of each sample was analyzed on a 5% polyacrylamide–8.3 M urea gel.

transcripts directed by the remaining three recombinants. A single band of ca. 2.2 to 2.3 kb is common to all three cell lines. This transcript is identical in size to that identified by Pourcel et al. (18). Cells bearing pSA10 (line $10a_2$) and pSA31 (line $31e_2$) show, in addition to the 2.3-kb species, a larger RNA (Fig. 2A); this RNA also hybridizes to pBR322 DNA (data not shown). These transcripts presumably arise via readthrough from or into surrounding plasmid or cellular sequences or both.

Cells bearing pSA25 (cell line 25b) show two HBV-related species, which have been examined in greater detail (Fig. 2). The predominant transcript in this line is the 2.3-kb species, which anneals to HBV DNA (Fig. 2A, center lane) but not to pBR322 (not shown). A much less abundant RNA species of about 3.2 kb can also be detected by hybridization to HBV DNA (Fig. 2A, center lane). When RNA from these cells is annealed to probes from the U5 region of the RSV LTR, only the larger species is detected (Fig. 2B, probe LTR); no annealing of either species occurred with probe from the U3 region of the LTR (not shown). These results, together with its presence irrespective of its orientation relative to the LTR (Fig. 2A, right lane), argue strongly that the 2.3-kb RNA is the product of promotion within HBV DNA.

The fine structure of pSA25-directed RNA was further examined by hybridization with subgenomic fragments of HBV DNA (Fig. 1). As expected from earlier work (10, 18), probes from the coding domain (probe B) and 3′ noncoding region (probe C) annealed strongly to HBsAg mRNA (Fig. 2B). However, probe prepared from the portion of the preS region between *Bst*EII and *Eco*RI annealed only weakly to the HBV-promoted species, though it detected the RSV-promoted species with the same sensitivity as did coding-

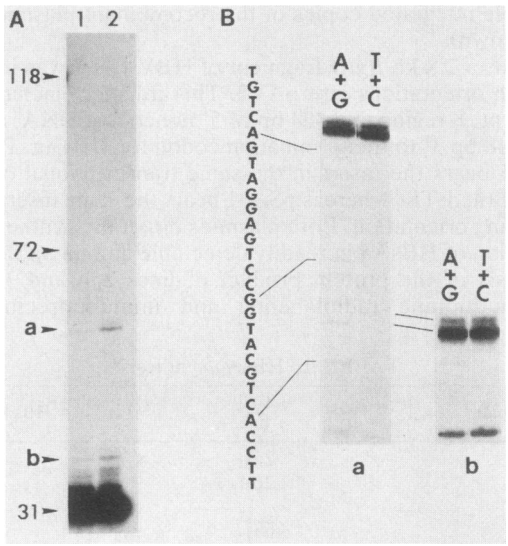

FIG. 4. Mapping of the 5′ end of the HBsAg RNA by primer extension. (A) Autoradiogram showing the extended products observed with the 31-nucleotide *Bam*HI-*Eco*RI primer. The numbers on the left indicate the positions of the primer (31 bp) and DNA size markers of 72 and 118 bp, respectively. No bands were visible in the original autoradiogram above ~56 bp. The extended products subjected to sequence analysis are indicated by arrows (a and b). Lanes 1 and 2 show the extended products obtained from two different RNA preparations from cells transformed with pSA25. (B) Sequence analysis of the extended species a and b shown in A. Only the G+A and T+C reactions are shown here; in the original autoradiogram, the G, A+C, and C reactions could also be read. The sequence of the region (nucleotides 3 through 3,197) is presented alongside B. The sequence is complementary to that shown in Fig. 7B. Each primer extension reaction was performed with ~2 × 10^6 cpm of probe and 25 μg of polyadenylated RNA from line 25b (A, lane 1) or 50 μg of RNA (A, lane 2). RNA plus probe was dissolved in 45 μl of 80% formamide containing 0.4 M NaCl and 10 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)], pH 6.4. After 10 min at 65°C, the samples were slowly cooled to 20°C. The hybrids were precipitated with ethanol, dissolved in 10 μl of water containing 20 mM Tris-chloride (pH 8.4), 50 mM NaCl, 6 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM deoxynucleotide triphosphates, and 26 U of reverse transcriptase, and incubated at 48°C for 1 h. Approximately 1% of each sample was analyzed on an analytical 5% polyacrylamide–8 M urea gel (A). The remainder was electrophoresed on a preparative gel of the same type. The extended products were detected by autoradiography, eluted from the gel, and subjected to sequence analysis (15). The products (~400 cpm per lane) were analyzed on a 15% polyacrylamide–8 M urea gel. Autoradiography was for 2 weeks at −70°C with a Quanta III screen (Dupont).

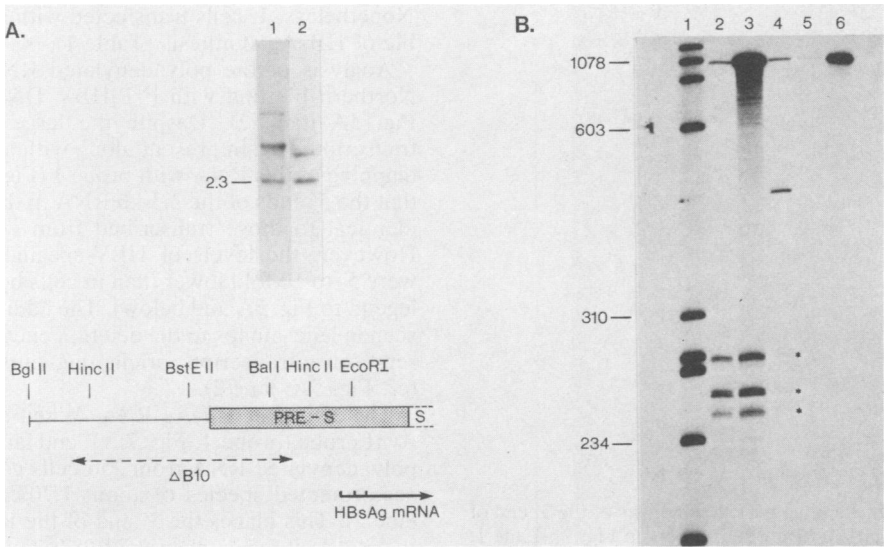

FIG. 5. Effects of RSV LTR and HBV preS region on HBsAg gene transcription. (A) Top panel: Northern gel analysis of polyadenylated RNA from cell lines 90b and B10c (Table 1). Two micrograms of polyadenylated RNA from line 90b (lane 1) and 5 μg of similar RNA from line B10c (lane 2) were electrophoresed in parallel through 1% agarose–2.2 M formaldehyde, transferred to nitrocellulose, and hybridized to [^{32}P]HBV DNA. Autoradiographic exposure was for 16 h (90b RNA) or 4 days (B10c RNA). Bottom panel: Schematic illustration of the preS region of HBV, showing the extent of deletion ΔB10. (B) S1 nuclease mapping of polyadenylated RNA from lines 90b and B10c. Lanes 2 through 5 contain S1-resistant products after annealing ^{32}P-labeled *Xba*I-*Bgl*II probe (probe 1 in Fig. 7) to 6 μg of 25b RNA (lane 2), 10 μg of line 90b RNA (lane 3), 8 μg of line B10c RNA (lane 4), and 20 μg of yeast tRNA (lane 5). Lane 1, *Hae*III-ϕX174 DNA size markers; lane 6, undigested probe. The experiment was performed as described in the legend to Fig. 3. Autoradiography was for 6 days at −70°C.

region probes (Fig. 2B, probe A). This result suggested that sequences from the 5′ portion of the preS region might be underrepresented in HBsAg mRNA.

Fine structure of the 5′ end of the HBsAg mRNA. Because the hybridization results indicated that at least some preS sequences were not present in HBsAg mRNA, we undertook to map the 5′ end of this RNA more precisely. Accordingly, cytoplasmic polyadenylated RNA from cell line 25b was analyzed by the S1 nuclease method (27) with the 5′-end-labeled probes depicted in Fig. 7.

RNA annealed to probe 1 (labeled at the *Xba*I site, position 250) protected three discrete species of DNA from S1 digestion (Fig. 3B, lanes 2 and 3). The size of these species (245, 255, and 275 nucleotides) corresponds to RNAs that extend 153, 163, and 183 bases upstream from the ATG codon at position 158 that marks the start of the mature HBsAg coding region. Three species were similarly protected when this RNA was annealed to probe 2 (end labeled at the *Rsa*I site, position 507); the sizes of these three fragments confirm that their 5′ ends correspond exactly to those detected with probe 1 (Fig. 3A, lane 2).

To precisely map the 5′ ends of the species protected by probe 1, these protected fragments were electrophoresed in parallel with a DNA sequencing ladder generated from the same probe. These results are shown in Fig. 3C, and the sequence alignments are summarized in Fig. 7B.

In each S1 mapping experiment with 25b RNA, a small amount of protected full-length probe was evident (Fig. 3A, lane 2 and 3B, lanes 2 and 3); such species are not seen when 25b RNA is omitted from the reaction (Fig. 3A, lane 3 and 3B, lane 4). This material presumably represents the minor fraction of HBV RNA that originates from the RSV LTR.

S1 mapping experiments of this type do not discriminate between initiation events and RNA splicing. Indeed sequences resembling consensus splice acceptor sites are found in the HBV sequence close to two of the sites mapped by S1 analysis (see Fig. 7). To distinguish between splicing and initiation, the 5′ end of the major RNA species was determined by primer extension analysis with two single-stranded primers (Fig. 4). A 31-nucleotide *Bam*HI-*Eco*RI primer (5′ end labeled at the *Bam*HI site, nucleotide 31) was annealed to HBV RNA from line 25b and extended with avian myeloblastosis virus reverse transcriptase; this reaction gave two major products of about 36 and 56 bases (Fig. 4A). These products were purified and subjected to sequence analysis (Fig. 4B), which revealed that the extended products were colinear with the HBV sequence and contained no sequence information derived from an additional exon. Comparison of the termination sites of the extended products with the sites mapped by S1 analysis revealed an exact alignment (summarized in Fig. 7) for the two groups of sites furthest upstream from the start of HBsAg. Products mapping to the same two sets of locations were observed when a synthetic 20-mer (nucleotides 31 through 50) was used as a primer against 25b RNA (data not shown). No clear band was seen mapping to the third S1 site (see Fig. 7B); however, this band is relatively weakly detected in the S1 analysis and was probably below the threshold of detection in the primer extension experiments.

Taken together, these data indicate that at least two and perhaps all three of the sites mapped in these experiments represent true initiation sites occurring at the positions summarized in Fig. 7.

HBsAg transcription in the absence of an LTR. Recent studies document that the RSV DNA sequences we employed in these experiments contain so-called enhancer sequences, *cis*-acting elements which activate the expression of neighboring genes (13). We considered the possibility that the transcripts defined above resulted from enhancer-mediated activation of a minor or aberrant promoter not

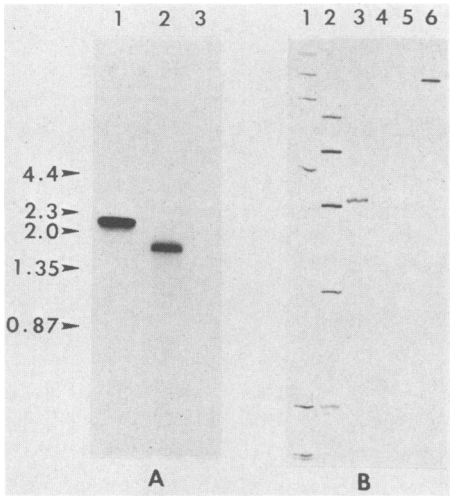

FIG. 6. Autoradiograms showing the S1 analysis of the 3′ end of the HBsAg RNA. (A) *Xba*I-*Ava*I probe (probe 4 in Fig. 7). Lane 1, Undigested probe; lane 2, S1-resistant products of annealing to about 6 μg of polyadenylated 25b RNA; lane 3, S1-resistant products of annealing to control yeast tRNA. The arrows on the left of the figure depict the position of DNA size markers visualized by UV after staining with ethidium bromide. (B) *Bam*HI-*Ava*I probe (probe 3 in Fig. 7). Lane 1, ^{32}P-labeled DNA size markers of 1,353, 1,078, 872, 603, 310, and 281 bp, respectively; lane 2, DNA size markers of 770, 652, 527, 404, and 309 bp; lane 3, S1-resistant products of annealing to 6 μg of polyadenylated 25b RNA; lane 4, identical analysis with control yeast RNA; lane 5, a sample of the *Bam*HI-*Ava*I probe digested with *Bgl*II; and lane 6, undigested *Bam*HI-*Ava*I probe. S1 mapping was performed as described in the legend to Fig. 3 and in the text.

normally functional in vivo. Accordingly, the entire HBV genome, permuted at the *Bgl*II site at position 2,432, was subcloned into a pBR322 derivative which contained no RSV sequences. As expected, when this plasmid (pSA90) was introduced into LTK$^-$ cells by cotransformation with HSV TK DNA, the resulting cell clones synthesized and secreted HBsAg (Table 1). Polyadenylated RNA prepared from one such transformant (90b) was examined by agarose gel electrophoresis and blot hybridization to [^{32}P]HBV DNA (Fig. 5A, lane 1). In this line, the 2.3-kb HBsAg transcript is again seen, as well as two larger species of over 4 kb in length. When this RNA was examined by S1 nuclease mapping (Fig., 5B, lane 3) with the *Xba*I-*Bgl*II probe (probe 1) described above, the same three fragments previously seen in 25b RNA were protected (cf. Fig. 5B, lane 2). No other major start sites or splice-acceptor sites within HBV DNA were identified. Significant protection of full-length probe 1 was observed, indicating that the larger transcripts seen in this line originated outside the HBV genome. Thus, the HBV RNA species mapped in Fig. 3 and 4 are not the result of activation of extraneous promoters by RSV.

We next examined the effect of a large deletion of preS sequences on HBsAg transcription. Plasmid pSA90 was cleaved at its unique *Bst*EII site in preS, and deletions were generated by exposure of the linearized genome to nuclease Bal31. One resulting mutant derivative (ΔB10) was selected for analysis. The fine structure of this mutant as determined by restriction mapping is summarized in Fig. 5A. The lesion in ΔB10 consists of a ~550-bp deletion, which removes at least 154 bp of the preS coding region, along with the region containing the putative promoter preceding preS. Nonetheless, L cells transfected with this plasmid are capable of HBsAg synthesis (Table 1).

Analysis of the polyadenylated RNA of clone B10c by Northern blotting with [^{32}P]HBV DNA probe is shown in Fig. 5A (lane 2). Despite the large deletion, the 2.3-kb transcript is again present along with a longer transcript. S1 mapping of this RNA with probe 1 (Fig. 5B, lane 4) revealed that the 5′ ends of the 2.3-kb RNA in the deletion mutant are identical to those transcribed from wild-type HBV DNA. However, the levels of HBV-specific RNA in these cells were 5- to 10-fold lower than in cells bearing pSA90 (see the legend to Fig. 5A and below). The additional band (~450 bp) seen in lane 4 maps to the deletion end point and presumably represents transcripts originating outside HBV sequences (cf. Fig. 5A, lane 2).

The 3′ end of HBsAg RNA. When annealed to an *Xba*I-*Ava*I probe (probe 4, Fig. 7), 3′ end labeled at the *Xba*I site, polyadenylated RNA from 25b cells gave a single S1 nuclease-protected species of about 1,700 nucleotides (Fig. 6A, lane 2). This places the 3′ end of the RNA in the vicinity of the *Bgl*II site at nucleotide 1,987. To map the 3′ end more precisely, probe 3 (labeled at the *Bam*HI site, nucleotide 1,403) was used. A fragment of 530 ± 10 bases was protected from S1 nuclease when this probe was annealed to 25b RNA (Fig. 6B, lane 3), placing the 3′ end of the HBsAg RNA at nucleotide 1,933 ± 10 on the HBV map, some 50 nucleotides upstream from the *Bgl*II site (cf. Fig. 6B, lane 5). These data, together with the results from the 5′ probes, indicate that the predominant HBsAg mRNA from this cell line is an unspliced molecule containing up to about 1,950 bases of HBV sequence. This is in reasonable agreement with the estimated size of 2.2 to 2.3 kb for the polyadenylated HBsAg message.

DISCUSSION

In agreement with other studies, our experiments show that fibroblast lines transfected with appropriate regions of cloned HBV DNA synthesize and secrete small amounts of HBsAg (4, 11). This synthesis is regularly associated with the presence of a stable polyadenylated transcript of ca. 2.3 kb (18; Fig. 2).

Analysis of HBsAg synthesis by the LTR-HBV recombinants pSA2 and pSA6 demonstrated that sequences 5′ to the *Bam*HI site at position 32 are required for HBsAg gene expression (Table 1). This is consistent with the earlier suggestion of Pourcel et al. (18) that transcription of HBsAg might initiate at a putative promoter element just upstream from the preS region. However, experiments with plasmid pSA10 (Table 1; Fig. 2A) indicate that this element can be separated from the remaining HBV sequences by a 3.5-kb insertion of plasmid DNA without disrupting HBsAg transcription. In addition, hybridization studies with preS region probes (Fig. 2B) indicate that in this system large sections of the preS region are not represented in the final HBsAg mRNA. This is confirmed by S1 mapping and primer extension studies which localize the 5′ ends of the RNA to three sites within a 35-bp region around the *Eco*RI site in preS. The positions of these sites (determined with an accuracy of 1 to 2 bp) are diagrammed in Fig. 7. Our data indicate that these ends represent discrete transcription initiation events rather than RNA splicing. Interestingly, inspection of the DNA sequence upstream from this region reveals no candidate TATAAA-like element. This unusual feature is also observed in the simian virus 40 late-transcription unit, which is likewise characterized by the production of RNAs with heterogeneous 5′ ends.

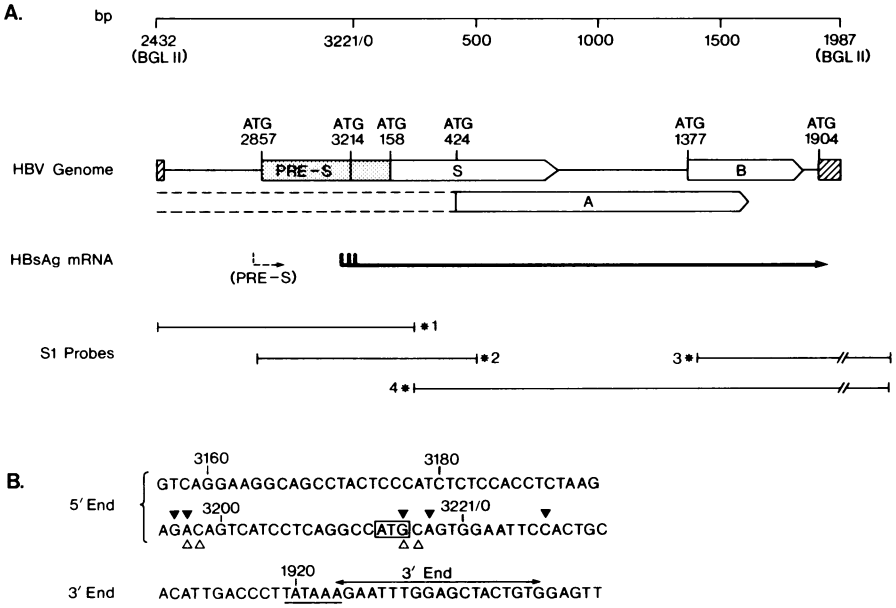

FIG. 7. (A) Structural features of the HBsAg 2.3-kb RNA. Top line shows the scale for the 2.8-kb *Bgl*II fragment of HBV DNA which contains the signals required for HBsAg synthesis in L cells. The unique HBV *Eco*RI site is at nucleotide 1 in this numbering system. Below it is depicted the arrangements of open reading frames in this region of the viral genome. The A frame is open throughout most of this region. Only the section which follows the first available ATG codon within this frame in HBsAg mRNA is depicted with a solid line; the earlier portion is indicated by a dashed line. Heavy line, Position of the 2.3-kb HBV RNA; arrowhead, direction of transcription. Dashed arrow (PRE-S), Location of putative preS promoter seen to be active in in vitro transcription systems. Below this is diagrammed the structures of the end-labeled probes used in the S1 mapping experiments of Fig. 3, 5B, and 6. Probe 1, *Xba*I-*Bgl*II; probe 2, *Rsa*I-*Bst*EII; probe 3, *Bam*HI-*Ava*I; and probe 4, *Xba*I-*Ava*I. *, Radiolabeled end. (B) Nucleotide sequence at the 5' and 3' ends of the 2.3-kb HBV RNA. The closed and open triangles denote the location of 5' ends of HBsAg RNA determined by the S1 and primer extension methods, respectively. These positions should be accurate to within 1 to 2 bp. The boxed ATG codon represents the initiating methionine for the HBsAg precursor containing 55 additional amino acids (see text).

Our experiments involve the introduction into mouse fibroblasts of viral DNA which is normally expressed only in human hepatocytes. Therefore the relationship of these findings to HBV transcription in virally infected liver cells in vivo requires examination. While this manuscript was in preparation, Cattaneo et al. (3) reported similar studies of RNA extracted from the liver of an HBV-infected chimpanzee, as well as from fibroblasts transformed with HBV DNA. These workers detected an RNA in the infected liver whose 5' end appears to be identical to the largest of the three transcripts we have identified in L cells. (The two smaller RNA species were not observed.) These results indicate that HBsAg gene transcription in L cells closely approximates that occurring in normal viral infection.

We do not detect in L cells stable RNA species initiating in the vicinity of the putative preS promoter. Nonetheless, such transcripts can be detected (along with the transcripts identified here) in monkey cells infected with certain simian virus 40-HBV recombinants (12) and in in vitro transcription experiments (20). Why this promoter is active in certain systems and not others is not understood. Our results do not exclude the possibility that sequences upstream of preS may be involved in HBsAg expression in mouse cells. Pourcel et al. (18), for example, observed that L cells transformed with HBV DNA deleted for sequences 5' to preS did not synthesize HBsAg. Although our results indicate that this is not the result of eliminating the start site for mRNA synthesis, it is conceivable that this region contains other sequences which might modulate the activity of the HBsAg promoter. In this context it is of note that clone B10C bearing the large preS deletion (ΔB10) produces five times less HBsAg and 2.3-kb HBV RNA than do some clones bearing the wild-type plasmid pSA90. However, there is considerable variation in levels of expression from clone to clone (see Table 1 and reference 4), and other clones bearing the ΔB10 deletion produce substantially more HBsAg (e.g., clone $B10a_2$, Table 1). Further experimentation is required to examine the possible role of upstream HBV sequences in transcription.

The disposition of ATG codons relative to the transcripts identified here is of interest. DNA sequencing studies show that the preS region harbors three in-phase ATG codons before the initial ATG of the mature HBsAg (26). The first two ATGs are absent from all three transcripts we have identified; the third ATG, at position 3,213, is present only in the largest transcript (see Fig. 7B).

The largest RNA species could therefore encode an HBsAg-related protein containing an additional 55 N-terminal amino acids derived from the preS region. Such a protein has recently been detected in the serum of HBV-infected humans (14, 23). Mature (24 kDa) HBsAg could arise by cleavage of this protein or by internal translation initiation at the second ATG of this RNA. Alternatively, mature HBsAg could be translated directly from the two smaller transcripts.

Our results and those of Cattaneo et al. (3) raise new questions concerning the possible functional role of the preS sequences. Comparison of the DNA sequence of HBV with those of the related hepadnaviruses of woodchucks (7) and ground squirrels (C. Seeger, D. Ganem, and H. E. Varmus, submitted for publication) reveals that a comparable preS open reading frame is present in all three genomes. This

would argue for an important role for this open reading frame in the viral life cycle. As noted above, the 55 amino acids encoded by the 3′ portion of preS may be involved in HBsAg biogenesis or function or both (14). What role the remaining preS sequences might play remains a mystery.

Our S1 mapping experiments suggest that HBsAg mRNA terminates at or near position 1,933 (Fig. 6); recent work has shown that this termination site represents the authentic site used in vivo (3). Inspection of the surrounding sequence reveals no AATAAA sequence, the hexanucleotide which normally precedes eucaryotic mRNA termini (6, 19). Interestingly, a variant of this sequence, TATAAA, is found at position 1,919 and may serve as a polyadenylation signal. However, the absence of a classic termination signal at this site leaves open the formal possibility that the RNA is spliced to downstream sequences in this region. We consider this possibility unlikely in view of the fact that HBsAg mRNA molecules of identical size are directed by plasmids pSA10 and pSA25 (Fig. 2). Plasmid pSA25 has only 55 bp of HBV sequence downstream from position 1,933, whereas pSA10 includes over 800 bp of HBV DNA distal to this site. If the sequence at 1,932 was a splice donor site, then in pSA25 the splice acceptor would have to be located within the cellular or plasmid DNA, whereas in pSA10 it would presumbly be within the HBV genome. It is unlikely that such disparate events would generate RNAs of identical size.

Inspection of the HBV DNA sequence revealed that position 1,933 lies within the coding region for the viral core antigen (HBcAg). The presence of a putative mRNA polyadenylation signal within this region raises new questions about how HBcAg mRNA might be generated in vivo. Removal of this signal by splicing would result in the loss of HBcAg coding information, although a small splice which does not disrupt the core reading frame is possible in principle. Another mechanism would be transcriptional readthrough past site 1,933. Resolution of this question will require examination of viral transcription in systems in which significant levels of core antigen gene expression can be observed.

The coding structure of the 2.3-kb HBsAg transcript (summarized in Fig. 7) also appears to be unusual. Approximately 500 nucleotides beyond the HBsAg termination codon is a second complete 154-amino acid open reading frame (termed B in Fig. 7). Moreover, a substantial segment of another open reading frame (A), which covers 80% of the genome and which has been speculated to encode the viral polymerase, is present in this RNA. Starting with the first available initiation codon at position 424 to the termination codon at position 1,624, the RNA could encode a 400-amino acid protein by using the A frame. Although eucaryotic mRNAs encoding more than one protein have recently been described (21), it remains to be determined whether the 2.3-kb RNA we have characterized can promote the translation of these potential gene products.

ACKNOWLEDGMENTS

We thank P. Luciw for performing the L cell transformations and for many helpful discussions, R. Cattaneo and H. Schaller for communication of unpublished data, and Leslie Spector for world-class manuscript preparation.

This work was supported by Public Health Service grants from the National Institutes of Health to H.E.V. and D.G. (AI/CA 18782) and W.J.R. (AI 19744).

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **Cattaneo, R., H. Will, G. Darai, E. Pfaff, and H. Schaller.** 1983. Detection of an element of the SV40 late promoter in vectors used for expression studies in COS cells. EMBO J. **2**:511–514.
3. **Cattaneo, R., H. Will, N. Hernandez, and H. Schaller.** 1983. Signals regulating hepatitis B surface antigen transcription. Nature (London) **305**:336–338.
4. **Dubois, M. F., C. Pourcel, S. Rousset, C. Chany, and P. Tiollais.** 1980. Excretion of hepatitis B surface antigen particles from mouse cells transformed with cloned viral DNA. Proc. Natl. Acad. Sci. U.S.A. **77**:4549–4553.
5. **Favoloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis of two-dimensional nuclease S1 mapping. Methods Enzymol. **65**:718–749.
6. **Fitzgerald, M., and T. Shenk.** 1981. The sequence of 5′AAUAAA3′ forms part of the recognition site of polyadenylation of late SV40 mRNA. Cell **24**:251–260.
7. **Galibert, F., T. Chen, and E. Mandart.** 1982. Nucleotide sequence of a cloned woodchuck hepatitis virus genome: comparison with the hepatitis B virus sequence. J. Virol. **41**:51–65.
8. **Galibert, F., E. Mandart, F. Fitoussi, P. Tiollais, and P. Charnay.** 1979. Nucleotide sequence of the hepatitis B virus genome. Nature (London) **281**:646–650.
9. **Ganem, D.** 1982. Persistent infection of humans with hepatitis B virus: mechanisms and consequences. Rev. Infect. Dis. **4**:1026–1047.
10. **Gough, N.** 1983. Core and e antigen synthesis in rodent cells transformed with cloned hepatitis B virus DNA is associated with greater than genome length viral mRNAs. J. Mol. Biol. **165**:683–699.
11. **Gough, N., and K. Murray.** 1982. Expression of hepatitis B surface, core and e antigen genes by stable rat and mouse cell lines. J. Mol. Biol. **162**:43–68.
12. **Laub, O., L. B. Rall, M. Truett, Y. Shaul, D. N. Standring, P. Valenzuela, and W. J. Rutter.** 1983. Synthesis of hepatitis B surface antigen in mammalian cells: expression of the entire gene and the coding region. J. Virol. **48**:271–280.
13. **Luciw, P., J. M. Bishop, H. E. Varmus, and M. R. Capecchi.** 1983. Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell **33**:705–716.
14. **Machida, A., S. Kishimoto, H. Ohumura, H. Miyamoto, K. Baba, K. Oda, T. Nakamura, Y. Miyakawa, and M. Mayumi.** 1983. A hepatitis B surface antigen polypeptide (P31) with the receptor for polymerized human as well as chimpanzee albumins. Gastroenterology **85**:268–274.
15. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **Pasek, M., T. Goto, W. Gilbert, B. Zink, H. Schaller, P. Mackay, G. Ledbetter, and K. Murray.** 1979. Hepatitis B virus genes and their expression in E. coli. Nature (London) **282**:575–579.
17. **Payne, G., S. Courtneige, L. Crittendon, A Fadly, J. M. Bishop, and H. E. Varmus.** 1981. Analysis of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for the maintenance of the transformed state. Cell **23**:311–322.
18. **Pourcel, C., A. Louise, M. Gervais, N. Chenciner, M.-F. Dubois, and P. Tiollais.** 1982. Transcription of the hepatitis B surface antigen gene in mouse cells transformed with cloned viral DNA. J. Virol. **42**:100–105.
19. **Proudfoot, N., and G. Brownlee.** 1976. 3′ noncoding sequences in eucaryotic mRNA. Nature (London) **263**:211–214.
20. **Rall, L. B., D. N. Standring, O. Laub, and W. J. Rutter.** 1983. Transcription of hepatitis B virus by RNA polymerase II. Mol. Cell Biol. **3**:1766–1773.
21. **Shaw, M. W., P. W. Choppin, and R. A. Lamb.** 1983. A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral neuraminidase. Proc. Natl. Acad. Sci. U.S.A. **80**:4879–4883.
22. **Standring, D. N., L. B. Rall, O. Laub, and W. J. Rutter.** 1983. Hepatitis B virus encodes an RNA polymerase III transcript. Mol. Cell Biol. **3**:1774–1782.

23. **Stibbe, W., and W. Gerlich.** 1983. Structural relationships between minor and major proteins of hepatitis B surface antigen. J. Virol. **46**:626–628.
24. **Thomas, P.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
25. **Tiollais, P., P. Charnay, and G. Vyas.** 1981. Biology of hepatitis B virus. Science **213**:406–411.
26. **Valenzuela, P., M. Quiroga, J. Zaldivar, P. Gray, and W. J. Rutter.** 1980. Nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. ICN-UCLA Symp. Mol. Cell. Biol. **18**:57–70.
27. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.