

## Analysis of Processing and Polyadenylation Signals of the Hepatitis B Virus Surface Antigen Gene by Using Simian Virus 40-Hepatitis B Virus Chimeric Plasmids

CHRISTIAN C. SIMONSEN AND ARTHUR D. LEVINSON\*

*Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080*

Received 25 July 1983/Accepted 8 September 1983

We examined the transcription of the hepatitis B virus surface antigen (HBsAg) gene in COS cells transfected with simian virus 40-based recombinant plasmids. When positioned behind the simian virus 40 late promoter, three transcripts were identified which hybridized to the HBsAg gene: a 2,000-nucleotide transcript colinear with a gene, a 1,100-nucleotide transcript representing a spliced molecule in which a major portion of the sequences encoding HBsAg were deleted, and an 800-nucleotide transcript derived primarily from sequences 3' to the HBsAg gene. The splice acceptor site utilized by the 1,100-nucleotide transcript is located immediately upstream of an open reading frame of unknown function contained within the 3' nontranslated region of the HBsAg gene. The HBsAg-specific mRNA species terminate 12 to 19 base pairs 3' of the sequence UAUAAA, similar to the consensus hexanucleotide which is thought to promote polyadenylation (AAUAAA). We constructed a series of plasmids with progressive deletions from the region surrounding where these transcripts terminate. Analysis of mRNA produced by cells transfected with these plasmids indicated that the signal hexanucleotide is in itself unable to promote the efficient processing of mRNA in the absence of downstream hepatitis B virus sequences. Processing proceeds properly, however, from plasmids containing an additional 30 nucleotides 3' of this signal.

Although it is estimated that hepatitis B virus (HBV) chronically afflicts 200 million persons worldwide (41, 53), the molecular biology of the virus is not well understood due to the inability of the virus to propagate in any tissue culture (53). The recent cloning of the 3,200-base-pair (bp) viral genome (4, 7, 47) and the determination of the complete nucleotide sequence of the cloned DNA (19, 36, 54) have allowed the definition of regions which encode viral polypeptides. These include the major nucleocapsid core antigen and the surface antigen (HBsAg). The surface antigen is found in the sera of infected individuals as a 22-nm particle devoid of DNA (41, 53) and in lesser amounts as a 42-nm particle (Dane particle) which represents the infectious virion (41). HBsAg represents the major neutralizing antigen of the virus and as such has proven effective as a vaccine against the disease (50). In addition, two other open reading frames have been identified (19, 36, 54), although no identification of the gene products has yet been made. It has been proposed that one such open reading frame may encode a viral polymerase (19, 36); the other has been designated the X region (54).

In view of the inability of cultured cells to

propagate the virus, cell lines containing integrated HBV sequences have been utilized to study the expression of viral proteins. The human hepatoma line PLC/PRF/5 has proven useful in studies designed to study transcription from, and organization of, the integrated HBV DNA (14, 29, 32, 38). Several groups have reported the introduction and expression of cloned HBV DNA in mammalian tissue culture cells by utilizing dominant selectable markers (8, 9, 16, 22, 38) or simian virus 40 (SV40) viral vectors (27, 33). Recently, the construction of SV40/HBV/pBR322 vectors which express high levels of HBsAg when introduced into COS cells (21) has been described (6, 12, 46). These chimeric vectors, which are nonlytic and thus not constrained by packaging considerations (20, 35), replicate to approximately  $10^5$  copies per cell and produce HBsAg particles similar, if not identical, to those found in the serum of infected patients (12, 27).

To define viral regulatory sequences which affect HBsAg expression, we constructed a series of plasmids containing HBV sequences with various amounts of HBsAg 3' untranslated sequences positioned behind the SV40 late pro-

moter. We identified two major transcripts which arise when the HBsAg gene is transcribed. Examination of such transcripts demonstrated that the hexanucleotide 5'-TATAAA at position 104 of the viral DNA appears to represent the polyadenylation signal; however, additional sequences downstream of this hexanucleotide are required for efficient production of cytoplasmic mRNA.

## MATERIALS AND METHODS

**DNA constructions.** All restriction enzymes, T4 polynucleotide kinase, and T4 DNA ligase were products of either Bethesda Research Laboratories, Inc. (Gaithersburg, Md.) or New England BioLabs, Inc. (Beverly, Mass.) and were used according to the prescribed reaction conditions. Synthetic oligonucleotides were synthesized according to previously described methods (11). The procedures for isolating plasmid DNA (12), electrophoresis, and electroelution (23) were as described previously. Plasmid pDLRI used in this study was derived from p342L (12) by removal of one of the two *EcoRI* sites, accomplished by filling in the cleaved restriction site with Klenow DNA polymerase (see Fig. 1). Bal 31 digestions were performed after digestion of DNA with the appropriate restriction nuclease. The DNA was phenol extracted and ethanol precipitated, and the ends were made blunt by the addition of Klenow DNA polymerase in the presence of the four nucleoside triphosphates (23). The samples were again phenol extracted and ethanol precipitated and then digested with a second restriction enzyme. Linear fragments of the appropriate length were isolated from polyacrylamide gels by electroelution and ligated into the indicated vector. The resulting plasmids were sequenced to verify the extent of the deletions introduced by the exonuclease. *Escherichia coli* 294 (1) was used as recipient in all bacterial transformations.

**DNA transfections.** To study the transcription of the recombinant SV40/HBV plasmids in monkey cells, monolayers of COS cells were grown to 50 to 60% confluency in 15-cm plastic dishes and transfected with DNA as described (12, 48). Briefly, the cells were washed with Dulbecco modified medium, and 10 ml of medium containing 2  $\mu$ g of plasmid DNA and DEAE-dextran at 200  $\mu$ g/ml was applied for 12 h at 37°C. The DNA solution was removed, the cells were washed twice with medium, 25 ml of medium containing 10% fetal calf serum was added, and the cells were incubated at 37°C before RNA extraction.

**RNA isolation and manipulations.** Cytoplasmic RNA was isolated from clarified extracts of  $10^8$  transfected COS cells which had been lysed by treatment with 20 ml of 0.15 M NaCl-1.5 mM MgCl<sub>2</sub>-0.65% Triton X-100-10 mM Tris-hydrochloride (pH 8.0) for 10 min at 4°C. EDTA was added to 10 mM, NaCl to 0.4 M, and sodium dodecyl sulfate to 0.2%, after which the preparations were phenol extracted, chloroform extracted, and precipitated by the addition of 2 volumes of cold ethanol. Initial experiments utilized RNA preparations which had been chromatographed on oligodeoxythymidylate [oligo(dT)]-cellulose; however, this step was omitted in later experiments. The RNA was subjected to electrophoresis in 5% formaldehyde-1% agarose

gels in 5 mM sodium acetate-1 mM EDTA-20 mM morpholinepropanesulfonic acid (MOPS), pH 7 (Sigma Chemical Co., St. Louis, Mo.), as described (52). Cytoplasmic RNA (10  $\mu$ g) or oligo(dT)-selected RNA (1  $\mu$ g) was denatured by heating for 3 min at 95°C in 20  $\mu$ l of loading buffer (50% formamide, 1 $\times$  MOPS buffer, 5% formaldehyde, 0.05% xylene cyanol, bromophenol blue), loaded onto horizontal gels (18 by 25 cm), and electrophoresed for a total of 450 V-h. The gels were stained for 15 min with 10  $\mu$ g of ethidium bromide per ml in MOPS buffer, destained twice with water for 20 min, and visualized by using short-wave UV light. RNA was then immediately transferred to nitrocellulose filters by using 10 $\times$  SSC as blotting buffer (1 $\times$  SSC is 0.15 M NaCl plus 0.015 M sodium citrate). To discern cytoplasmic HBsAg-specific RNA, the nitrocellulose filters were hybridized with <sup>32</sup>P-labeled HBV DNA prepared by nick translation (44) of cloned DNA fragments.

**Construction and screening of cDNA plasmid bank.** COS cells were transfected with plasmid DNA as described, and polyadenylated mRNA was purified by chromatography on oligo(dT)-cellulose (P-L Biochemicals, Milwaukee, Wis.). Construction of a cDNA plasmid bank was performed by using standard techniques (23, 24).

**Radioimmune assays.** HBsAg was assayed by the Ausria II radioimmunoassay kit (Abbott Laboratories, North Chicago, Ill.) and quantitated by serial dilution of the unknown sample and comparison to the positive control supplied as described (12).

**DNA sequence analysis.** Sequence information was obtained by the method of Maxam and Gilbert (30) or by the dideoxynucleotide method (42) with synthetic oligonucleotide primers (11).

## RESULTS

**Construction of vectors.** The 342-bp *PvuII*-*HindIII* fragment of SV40 utilized in this study contains sequences capable of promoting the synthesis of mRNA in both the early and late directions (12, 33, 51) and serves as an origin of replication of the plasmid when introduced into mammalian cells expressing SV40 large T antigen (21, 31). The vectors employed were based on those previously described (12), which replicate and express high levels of HBsAg from the SV40 late promoter within 24 to 36 h after transfection into COS cells. The removal of the *EcoRI* site preceding the SV40 origin by filling in the restriction site with Klenow DNA polymerase (23) of one of these vectors, p342L (12), created pDLRI (Fig. 1). The pHBVterm plasmids were constructed by digesting pDLRI with Bal 31 as described above. Plasmids containing progressive deletions of the 3' terminus of the gene were obtained, and the junction regions were sequenced to precisely determine the extent of the deletion (Fig. 1).

**Transcription of the HBsAg gene in COS cells.** To determine the size of the HBsAg-specific RNA found in COS cells transfected with the surface antigen expression plasmids, we trans-

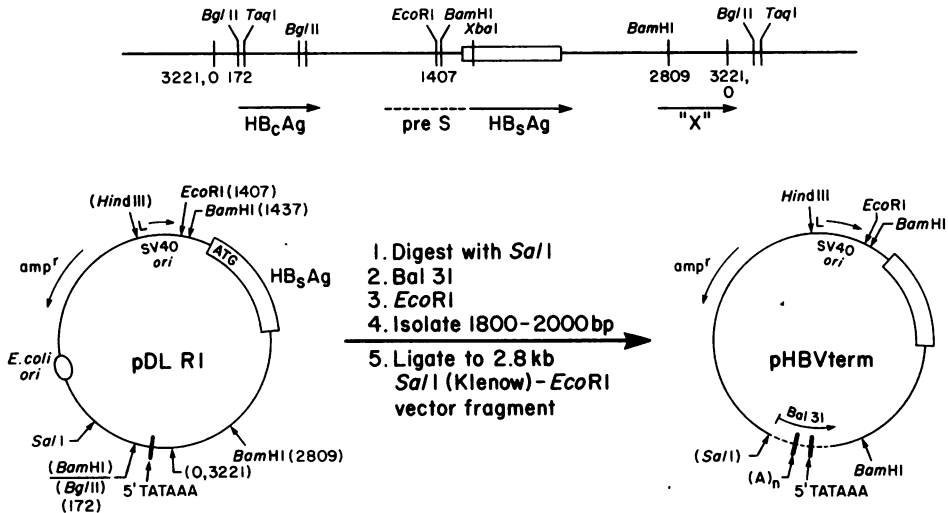


FIG. 1. Plasmid vector constructions. Plasmid maps of pDLRI and the pHBVterm deletion plasmids employed in this study. A linear map of the HBV genome is shown at top. The HBsAg coding region is boxed. The three open reading frames which have been proposed to encode the core antigen, surface antigen, and X protein (19, 36, 54) are indicated by the arrows underneath the line drawing. Map coordinates are those of Valenzuela et al. (54).

fecting COS cells with pDLRI DNA by using the modified DEAE-dextran method (48). At 1, 2, 4, and 6 days posttransfection, cytoplasmic RNA was prepared from transfected cultures. These RNAs were electrophoresed on denaturing agarose gels, transferred to nitrocellulose filters, and hybridized to a 1,400-bp  $^{32}$ P-labeled *Bam*HI fragment containing the coding sequences of the surface antigen gene. Two species of HBsAg-specific RNA were observed within 2 days of transfection: a predominant mRNA species of approximately 2,000 nucleotides (nt) and a minor species of approximately 1,100 nt (Fig. 2). Maximal levels of mRNA were observed on day 4, closely paralleling the time of peak HBsAg expression as detected by radioimmuno assay (12).

The 2,000-nt mRNA species observed in Fig. 2 was of a size consistent with a transcript initiating within the SV40 late promoter and ending at or near the *Bgl*III site of HBV at position 172. The nature of the smaller transcript was unclear, although its size was similar to transcripts spanning the X region of the HBV genome reported in cells containing integrated HBV DNA (22, 38). The generation of different mRNA species from the same gene could be due to the utilization of different polyadenylation sites (44, 45) or differential RNA processing mechanisms (15, 55). It was also possible that an internal promoter within the HBV sequences was being utilized. To identify the sequences contained within the 2,000- and 1,100-nt HBsAg-specific mRNAs, we prepared radiolabeled hybridization probes from various regions of the

HBsAg gene extending from the *Eco*RI site to the *Bgl*III site. These probes were hybridized to filters containing RNA from COS cells transfected with pDLRI. Both the 2,000- and 1,100-nt mRNAs hybridized to a probe prepared from the *Eco*RI-*Bgl*III (Fig. 3, lane A) or *Eco*RI-*Hpa*I (lane B) fragments which span the HBsAg-encoding sequences. When a probe was prepared from two *Rsa*I restriction fragments which span the COOH terminus of the protein and a portion of the 3' nontranslated region (positions 1,979 to 2,178 and 2,334 to 2,540), only the 2,000-nt mRNA hybridized (Fig. 3, lane C), indicating that the 1,100-nt mRNA lacks sequences spanning the termination codon of the HBsAg gene. Finally, lane D shows the pattern of hybridization when the probe was prepared from the *Bam*HI-*Bgl*III fragment spanning positions 2,809 to 172, representing the 3' end of the HBV sequences found in pDLRI. Both the 2,000- and 1,100-nt mRNAs were observed, and in addition a third transcript of approximately 800 nt was seen. Since the 1,100-nt transcript contains sequences corresponding to the 5' and 3' ends of the 2,000-nt transcript, these results suggest that the smaller transcript represents a spliced version of the larger transcript.

**cDNA cloning of HBsAg-specific transcripts.** To examine more closely the nature of these two HBsAg-specific transcripts, we prepared a cDNA library from mRNA obtained from COS cells transfected with pDLRI. Total cytoplasmic RNA was prepared and the polyadenylated fraction was isolated by oligo(dT)-cellulose chroma-

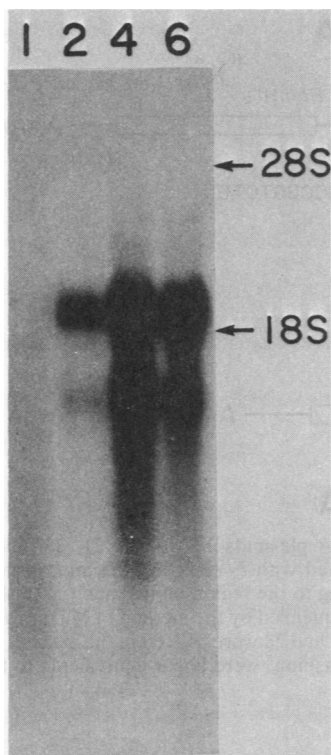


FIG. 2. Hybridization of RNA from transfected COS cells. RNA from COS cells transfected with pDLRI DNA was isolated 1, 2, 4, and 6 days post-transfection (indicated above the lanes) and electrophoresed in formaldehyde-agarose gels (52). The gels were stained with ethidium bromide as described in the text to mark the position of the 18S and 28S rRNAs as well as to verify the quality of the RNA preparations. After transfer of nitrocellulose filters (49), the RNA was hybridized to a  $^{32}\text{P}$ -labeled probe prepared from the 2,000-bp *EcoRI*-*Bgl*II fragment spanning the surface antigen region. Hybridization was performed in 50% formamide-5 $\times$  SSPE (0.9 M NaCl, 50 mM  $\text{NaH}_2\text{PO}_4$ , 5 mM EDTA, pH 7)-0.3% sodium dodecyl sulfate-100  $\mu\text{g}$  of salmon sperm DNA per ml at 42°C as described (13). After washing the filters in 2 $\times$  SSPE, the filters were exposed to Kodak AR2 X-ray film for 1 to 3 days with intensifying screens at -80°C.

tography (23). Complementary double-stranded DNA greater than 500 bp was isolated, incubated with terminal transferase in the presence of dCTP, and ligated into *Pst*I-cleaved pBR322 to which dG residues had been added. Approximately 15,000 colonies were screened for the presence of sequences complementary to the 2,000-bp *EcoRI*-*Bgl*II fragment spanning the coding region of the surface antigen gene. We identified 30 colonies which hybridized to the probe, containing inserts ranging in size from 600 to 1,800 bp. Ten clones were chosen for more extensive analysis. Restriction enzyme digestion with *Bam*HI and *Pst*I released an

approximately 550-bp fragment from all inserts which hybridized to the 580-bp *Bam*HI-*Bgl*II fragment 3' of the HBsAg gene. The sequences 5' of the *Bam*HI site were contained in two classes of cDNA plasmids: one set appeared to be contiguous with the HBV insert in pDLRI, whereas the other class lacked restriction sites upstream of the *Rsa*I site at position 2,790. These results suggested that all of the HBsAg-specific cDNA clones isolated shared the same 3' end, but differed upstream of the *Bam*HI site located at position 2,809. To confirm this and to precisely determine the position at which the two classes differed, DNA from several cDNA plasmids of each class was digested with *Bam*HI, end labeled with [ $\gamma$ - $^{32}\text{P}$ ]ATP in the presence of polynucleotide kinase, and subjected to sequence analysis by the method of Maxam and Gilbert (30). The results are summarized in Fig. 4. One class of cDNA clones was colinear with the HBV DNA. The other class was colinear with the HBV DNA to a point

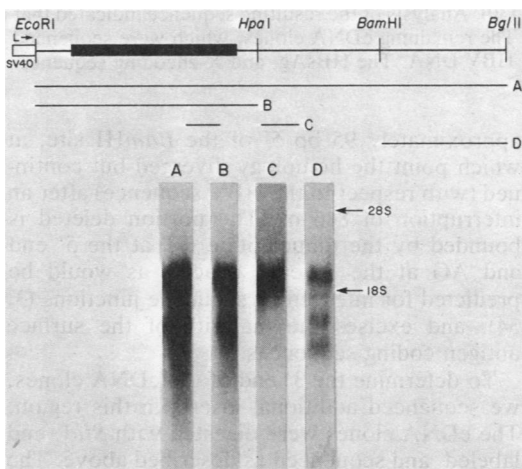


FIG. 3. Northern analysis of HBsAg RNA. RNA was isolated from COS cells transfected 4 days previously with pDLRI plasmid DNA. RNA (2.5  $\mu\text{g}$  per lane) was loaded onto a formaldehyde-agarose gel, electrophoresed, and transferred onto nitrocellulose as described in the legend to Fig. 2. The filter was cut into strips corresponding to the different lanes and hybridized with nick-translated (13, 44) probes prepared from DNA fragments isolated from polyacrylamide gels (46). Lane A, RNA probed with the 2,000-bp *EcoRI*-*Bgl*II fragment (nucleotides 1,407 to 172); lane B, RNA probed with the 900-bp *EcoRI*-*Hpa*I fragment (nucleotides 1,407 to 2,178 and 2,334 to 2,540); lane C, RNA probed with *Rsa*I fragments from areas indicated on line drawing at top of figure (nucleotides 1,979 to 2,178 and 2,334 to 2,540); lane D, RNA probed with the 580-bp *Bam*HI-*Bgl*II fragment (nucleotides 2,809 to 172). All fragments were nick translated to a specific activity of  $0.5 \times 10^8$  to  $1 \times 10^8$  counts per  $\mu\text{g}$ . Filters were hybridized, washed, and exposed as described in the legend to Fig. 2. The HBsAg-encoding region is shown boxed.



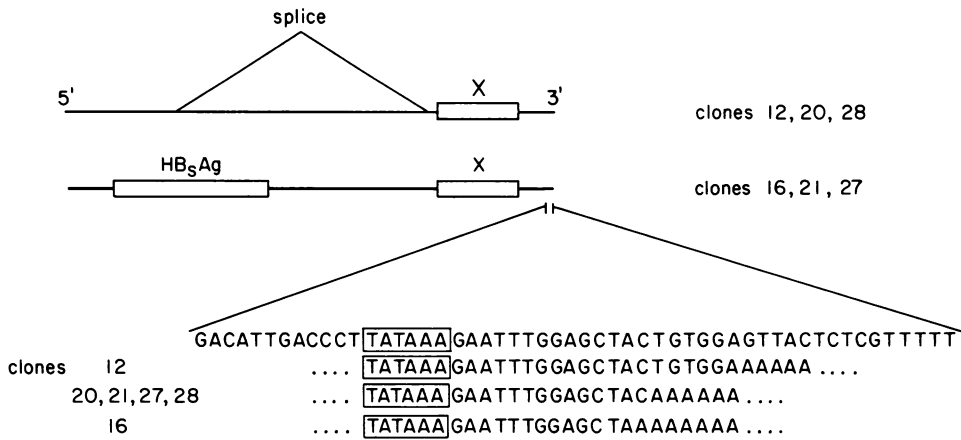


FIG. 5. Analysis of 3' ends of HBV cDNA plasmids. HBV cDNA plasmids 12, 16, 20, 21, 27, and 28 were digested with *Sma*I and labeled with [ $\gamma$ - $^{32}$ P]ATP by using polynucleotide kinase as described in the legend to Fig. 4. After digestion with *Eco*RI, labeled fragments were isolated from polyacrylamide gels and sequenced (30). The sequence of the cloned HBV DNA template and the sites at which the polyadenylate tract begin are indicated. The HBsAg-coding region and open reading frame (X) are boxed.

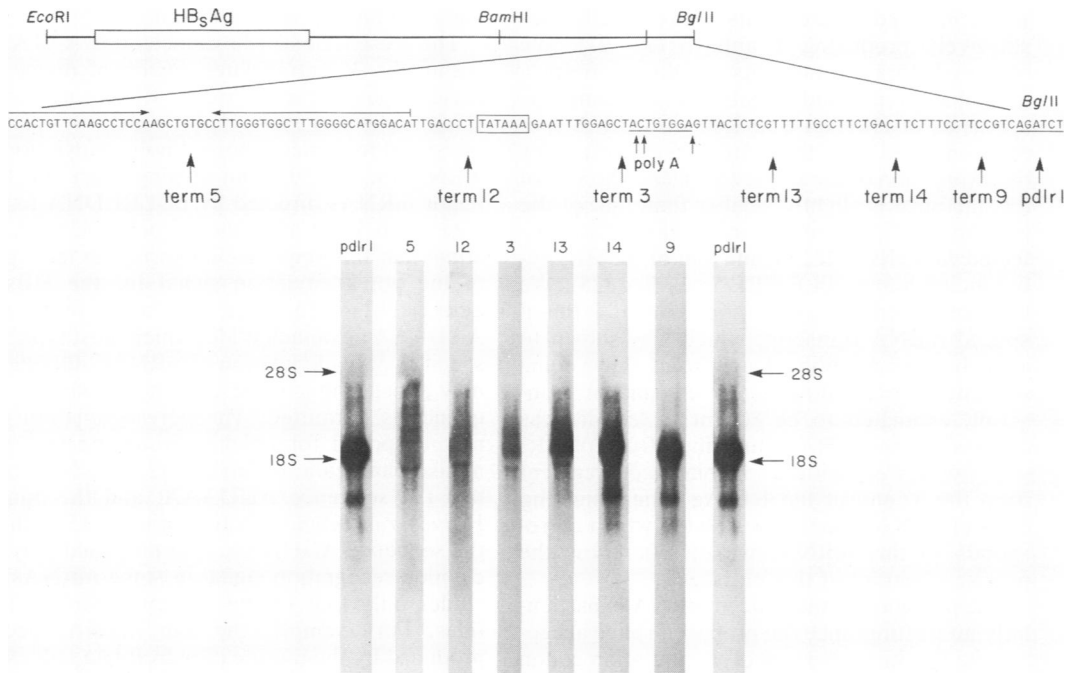


FIG. 6. Northern analysis of RNA expressed from pHBVterm plasmids. Plasmids were constructed by digesting *Sma*I-cleaved pDlrI with *Bal* 31 as described in the text. All plasmids were rejoined to pML (28) at the *Sma*I site, which had been filled in with Klenow DNA polymerase (23). The junction between the HBsAg gene and pML of the various plasmids is indicated at the top of the figure by the arrows. The hexanucleotide 5'-TATAAA, which serves as a polyadenylation signal (position 104), is outlined. The regions where polyadenylation occurs (122 to 129) and the *Bgl*III site at 172 are underlined. The location of an inverted repeat is indicated by the arrows. The HBsAg-encoding region is boxed. COS cells were transfected with each of the plasmids as described in the text, and RNA was harvested at 4 days posttransfection. RNA from each sample was electrophoresed and transferred to nitrocellulose as described in the legend to Fig. 2. The filters were hybridized with nick-translated 2,000-bp *Eco*RI-*Bgl*III fragment, washed, and exposed at  $-80^{\circ}\text{C}$  with an intensifying screen. Lanes marked 5, 12, and 3 were exposed for 72 h. All other lanes were exposed for 24 h. The plasmid used to produce the HBsAg-specific RNA is denoted immediately above each lane.

ther a 1,100- or a 2,000-nt RNA, and as such it behaves identically to pHBVterm 12, which lacks both the polyadenylation signal and site of addition, and pHBVterm 5, which lacks even additional sequences. It is noteworthy that cells transfected with pHBVterm 3, 5, and 12 do express some HBsAg-specific RNA, but it appears to be heterogeneously sized (generally longer than 2,000 nt) and present at very low levels (note that lanes illustrating pHBVterm 3, 5, and 12 transcripts in Fig. 6 were exposed three times longer than the other lanes). We therefore conclude that the processing and polyadenylation of the HBsAg-specific transcripts require, in addition to the UAUAAA signal, sequences encompassing the actual site of polyadenylation addition.

### DISCUSSION

To study transcriptional regulatory signals of the HBsAg gene of HBV, we exploited the ability of recombinant plasmids containing the SV40 origin of replication to replicate and transcribe inserted heterologous genes in cells constitutively producing T antigen (12, 21). We observed three transcripts produced from a recombinant plasmid capable of directing the synthesis of HBsAg. The 2,000-nt mRNA species contained all of the sequences necessary for the expression of the surface antigen. The size of this message is slightly smaller than that of the HBsAg mRNA identified in rodent cells transfected with HBV (22, 38) and that of the HBsAg-producing human PLC/PRF/5 cell line (38). We also observed the generation of a 1,100-nt spliced mRNA transcript which was shown to arise by the deletion of an 846-nt intervening sequence. In addition, an even smaller transcript, estimated to be 800 nt in length, was observed when RNA transcribed from pDLRI was hybridized with a radiolabeled fragment from the 3' end of the HBsAg gene. Since no class of cDNA clones was isolated which corresponds to this mRNA, we do not know the precise structure of this message.

The presence of the spliced mRNA is particularly interesting since the proposed splice acceptor site lies just upstream of the initiation codon of the X gene (19, 36, 54), the open reading frame contained within the 3' nontranslated region of the surface antigen gene. Although the 2,000-nt mRNA encodes the HBsAg secreted from transfected COS cells, the 1,100-nt mRNA interrupts the HBsAg-coding region after 101 amino acids; this reading frame continues into sequences encoding the final 105 amino acids of the putative HBV-encoded polymerase (19, 36). Whether such a fusion protein is synthesized is unclear. Alternatively, it is conceivable the 3' splice site we observe is utilized to express the

HBV X gene product (54), perhaps during natural infections joining to a more distal 5' splice site not present in our constructions. Although we cannot be certain whether this splicing event reflects a bona fide feature of the HBV life cycle, it is clear that a significant fraction of the total HBsAg mRNA found in transfected COS cells is so processed. Two minor mRNA species, estimated to be 1,000 and 900 nt in size, have been observed in human and rodent cell lines transfected with HBV DNA (22, 38), and at least one of these minor RNA species hybridizes to the X region (22). A splicing mechanism could account for the generation of the smaller mRNA observed in HBV-transfected cells which hybridizes to the X region. It must be cautioned, however, that all of the tissue culture systems used to study HBV, like the COS system utilized here, are nonpermissive for viral propagation. Thus, some factor(s) necessary for viral growth is clearly missing, and it is difficult to know whether the splicing pattern observed in this system reflects processing events occurring during a productive viral infection.

The presence of the hexanucleotide AATAAA within the 3' untranslated region of many eucaryotic genes transcribed by RNA polymerase II led to the suggestion that it constitutes part of the signal for processing and polyadenylation of mRNA (3, 17, 39). Our findings that the two major mRNAs directed by pDLRI DNA terminate 12 to 19 nt after the sequence UAUAAA argue that this sequence comprises at least part of the polyadenylation signal for the HBsAg gene. This sequence deviates from the AAUAAA hexanucleotide, which most typically specifies polyadenylation of mRNAs, although it now appears that some deviation from this sequence is permitted. The polyadenylation signals proposed for chicken lysozyme (26) and mouse pancreatic  $\alpha$ -amylase (25) mRNAs contain the sequence AUUAAA, and the human Harvey *ras* cellular proto-oncogene (5) utilizes the sequence AGUAAA. The proposed polynucleotide recognition signal in some mRNAs includes only four of the six consensus nucleotides. For example, the four mRNA species produced by the murine dihydrofolate reductase gene are all preceded by the sequence 5'-AUAA (44, 45), yet the same sequence is found seven other times in the 3' nontranslated region of the gene without producing an associated mRNA species.

It has been proposed from the frequency with which randomly inserted genomic fragments of DNA can provide processing/polyadenylation functions for the herpesvirus thymidine kinase gene that the minimal signal specifying such events is a hexanucleotide (43). This proposal gains some support from the demonstration that

an 88-bp fragment from the middle of the coding region of SV40 large T antigen, containing the sequence AAUAAA which is not normally involved in transcriptional processing, can specify such processing when positioned behind the thymidine kinase gene (10), although it must be stressed that processing promoted in this manner occurs inefficiently. However, it is clear that the AATAAA hexanucleotide can lie unrecognized within the coding region of an mRNA, as in the middle of the early region of SV40 (18, 40) and within the adenovirus type 12 E1a transcriptional unit (37). It therefore appears that sequences in addition to the hexanucleotide are involved in processing/polyadenylation. Whether such sequences are acting as additional recognition signals or are providing a general structure is not known. Our finding that sequences extending 3' of the hexanucleotide signal are required for efficient RNA processing provides direct experimental support for this conclusion. Furthermore, since deletion plasmids having all HBV sequences intact to a point at least 10 nucleotides 3' of the site of polyadenylation still manifest some loss of transcript integrity, we would argue that the efficient cleavage/polyadenylation of HBsAg mRNA involves recognition of sequences extending beyond the actual polyadenylation site. In this regard, Benoist et al. (2) have noted a 10-nucleotide model sequence (5'-TTTTCACTGC-3') at the polyadenylation junction in five of nine mRNAs they examined, perhaps reflecting the structural basis of a processing site which actually extends even further in the 3' direction.

#### ACKNOWLEDGMENTS

We thank Fran Fields and Jill Souder for excellent technical assistance. We also thank Y. Gluzman for providing COS-1 cells used in these studies. The assistance of Dennis Eaton and Ellson Chen in sequencing is gratefully acknowledged.

#### LITERATURE CITED

- Backman, K., M. Ptashne, and W. Gilbert. 1976. Construction of plasmids carrying the  $\text{cl}$  gene of bacteriophage  $\lambda$ . Proc. Natl. Acad. Sci. U.S.A. 73:4174-4178.
- Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon. 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. 8:127-142.
- Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. 50:349-383.
- Burrell, C. J., P. Mackay, P. J. Greenaway, P. H. Hofschneider, and K. Murray. 1979. Expression in *Escherichia coli* of hepatitis B virus DNA sequences cloned in plasmid pBR322. Nature (London) 279:43-47.
- Capon, D. J., E. Y. Chen, A. D. Levinson, P. H. Seeberg, and D. V. Goeddel. 1983. Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature (London) 302:33-37.
- 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 (Eur. Mol. Biol. Org.) J. 2:511-514.
- Charney, P., C. Pourcel, A. Louise, A. Fritsch, and P. Tiollais. 1979. Cloning in *Escherichia coli* and physical structure of hepatitis B virion DNA. Proc. Natl. Acad. Sci. U.S.A. 76:2222-2226.
- Christman, J. K., M. Gerber, P. M. Price, C. Flordellis, J. Edelman, and G. Acs. 1982. Amplification of expression of hepatitis B surface antigen in 3T3 cells cotransfected with a dominant-acting gene and cloned viral DNA. Proc. Natl. Acad. Sci. U.S.A. 79:1815-1819.
- Colbere-Garapin, F., F. Horodniceanu, P. Kourilsky, and A. C. Garapin. 1983. Late transient expression of human hepatitis B virus genes in monkey cells. EMBO (Eur. Mol. Biol. Org.) J. 2:21-25.
- Cole, C. N., and G. M. Santangelo. 1983. Analysis of COS-1 cells of processing and polyadenylation signals by using derivatives of the herpes simplex virus type 1 thymidine kinase gene. Mol. Cell. Biol. 3:267-269.
- Crea, R., A. Kraszewski, T. Hirose, and K. Itakura. 1978. Chemical synthesis of genes for human insulin. Proc. Natl. Acad. Sci. U.S.A. 75:5765-5769.
- Crowley, C., C. C. Liu, and A. D. Levinson. 1983. Plasmid-directed synthesis of hepatitis B surface antigen in monkey cells. Mol. Cell. Biol. 3:44-55.
- Davis, R. W., D. Botstein, and J. R. Roth. 1980. Advanced bacterial genetics. p. 138-158. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
- Dejean, A., C. Brechot, P. Tiollais, and S. Wain-Hobson. 1983. Characterization of integrated hepatitis B viral DNA cloned from a human hepatoma and the hepatoma-derived cell line PLC/PRF/5. Proc. Natl. Acad. Sci. U.S.A. 80:2505-2509.
- DeNoto, F. M., D. D. Moore, and H. M. Goodman. 1981. Human growth hormone DNA sequence and mRNA structure: possible alternative splicing. Nucleic Acids Res. 9:3719-3730.
- 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. 77:4549-4552.
- Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Vande Voode, H. Van Hueverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysehaert. 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) 273:113-120.
- Fitzgerald, M., and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site of polyadenylation of late SV40 mRNAs. Cell 24:251-260.
- Galibert, F., E. Mandart, F. Fitoussi, P. Tiollais, and P. Charnay. 1979. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. Nature (London) 281:646-650.
- Gething, M. J., and J. Sambrook. 1981. Cell-surface expression of influenza haemagglutinin from a cloned DNA copy of the RNA gene. Nature (London) 293:620-625.
- Gluzman, Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23:175-182.
- Gough, N. M. 1983. Core and E antigen synthesis in rodent cells transformed with hepatitis B virus DNA is associated with greater than genome length viral messenger RNAs. J. Mol. Biol. 165:683-699.
- Gray, P. W., D. W. Leung, D. Pennica, E. Yelverton, R. Najarian, C. C. Simonsen, R. Derynck, P. Sherwood, D. M. Wallace, S. L. Berger, A. D. Levinson, and D. V. Goeddel. 1982. Expression of human immune interferon cDNA in *E. coli* and monkey cells. Nature (London) 295:503-508.
- Grunstein, M., and D. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. 72:3961-3965.
- Hagenbuchle, O., R. Bovey, and R. A. Young. 1980. Tissue specific expression of mouse  $\alpha$ -amylase genes: nucleotide sequence of iso-enzyme mRNAs from pancreas and salivary gland. Cell 21:179-187.
- Jung, A., A. E. Sippel, M. Grez, and G. Schutz. 1980.



- Exons encode functional and structural units of chicken lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **77**:5759-5763.
27. Liu, C. C., D. Yansura, and A. D. Levinson. 1982. Direct expression of hepatitis B surface antigen in monkey cells from an SV40 vector. *DNA* **1**:213-221.
  28. Lusky, M., and M. Botchan. 1981. Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. *Nature (London)* **293**:79-81.
  29. Marion, P. L., F. H. Salazar, J. J. Alexander, and W. S. Robinson. 1979. Polypeptides of hepatitis B virus surface antigen produced by a hepatoma cell line. *J. Virol.* **32**:796-802.
  30. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**:560-564.
  31. Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis. 1981. Identification of DNA sequences required for transcription of the human  $\alpha$ -globin gene in a new SV40 host-vector system. *Cell* **27**:279-288.
  32. Miller, R. H., and W. S. Robinson. 1983. Integrated hepatitis B virus DNA sequences specifying the major viral core polypeptide are methylated in PLC/PRF/5 cells. *Proc. Natl. Acad. Sci. U.S.A.* **80**:2534-2538.
  33. Moriarty, A. M., B. H. Hoyer, J. W. Shih, J. L. Gerin, and D. H. Hamer. 1981. Expression of hepatitis B virus surface antigen gene in cell culture by using a simian virus 40 vector. *Proc. Natl. Acad. Sci. U.S.A.* **78**:2606-2610.
  34. Mount, S. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**:459-472.
  35. Mulligan, R. C., B. H. Howard, and P. Berg. 1979. Synthesis of rabbit  $\beta$ -globin in cultured monkey kidney cells following infection with a SV40  $\beta$ -globin recombinant genome. *Nature (London)* **277**:108-114.
  36. Pasek, M., T. Goto, W. Gilbert, B. Zink, H. Schaller, P. Mackay, G. Leadbetter, and K. Murray. 1979. Hepatitis B virus genes and their expression in *E. coli*. *Nature (London)* **282**:575-579.
  37. Perricaudet, M., J.-M. leMoullec, P. Tiollais, and U. Petterson. 1980. Structure of two adenovirus type 12 transforming polypeptides and their evolutionary implications. *Nature (London)* **288**:174-176.
  38. 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.
  39. Proudfoot, N., and G. G. Brownlee. 1976. 3' non-coding region sequences in eucaryotic messenger RNA. *Nature (London)* **263**:211-214.
  40. Reddy, V. B., B. Thimmappaga, R. Dhar, K. N. Subramanian, S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weismann. 1978. The genome of SV40. *Science* **200**:494-502.
  41. Robinson, W. 1979. Viruses of human hepatitis A and B. p. 471-526. *In* H. Frankel-Conrat and R. Wagner (ed.). *Comprehensive virology*, vol. 14: newly characterized vertebrate viruses. Plenum Publishing Corp., New York.
  42. Sanger, F., S. Nicklen, and R. Coulson. 1977. DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5463-5467.
  43. Santangelo, G. M., and C. N. Cole. 1983. Preparation of a "functional library" of African green monkey DNA fragments which substitute for the processing/polyadenylation signal in the herpes simplex virus type 1 thymidine kinase gene. *Mol. Cell. Biol.* **3**:643-653.
  44. Setzer, D. R., M. McGrogan, J. H. Nunberg, and R. T. Schimke. 1980. Size heterogeneity in the 3' end of dihydrofolate reductase messenger RNAs in mouse cells. *Cell* **22**:361-370.
  45. Setzer, D. R., M. McGrogan, and R. T. Schimke. 1982. Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductase gene. *J. Biol. Chem.* **257**:5143-5147.
  46. Siddiqui, A. 1983. Expression of hepatitis B virus surface antigen gene in cultured cells by using recombinant plasmid vectors. *Mol. Cell. Biol.* **3**:143-146.
  47. Sninsky, J. J., A. Siddiqui, W. S. Robinson, and S. N. Cohen. 1979. Cloning and endonuclease mapping of the hepatitis B viral genome. *Nature (London)* **279**:346-348.
  48. Sompayrac, L. M., and K. J. Danna. 1981. Efficient infection of monkey cells with DNA of simian virus 40. *Proc. Natl. Acad. Sci. U.S.A.* **78**:7575-7578.
  49. Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
  50. Szmunn, W., C. E. Stevens, E. J. Harley, E. A. Zang, W. R. Oleszka, D. C. William, R. Sadovsky, J. M. Morrison, and A. Kellner. 1980. Hepatitis B vaccine. Demonstration of efficacy in a controlled clinical trial in a high-risk population in the United States. *N. Engl. J. Med.* **303**:833-841.
  51. Tegtmeyer, P. 1972. Simian virus 40 deoxyribonucleic acid synthesis: the viral replicon. *J. Virol.* **10**:591-598.
  52. Thomas P. 1980. Hybridization of denatured RNA and small DNA fragments to nitrocellulose. *Proc. Natl. Acad. Sci. U.S.A.* **77**:5201-5205.
  53. Tiollais, P., P. Charnay, and G. N. Vyas. 1981. Biology of hepatitis B virus. *Science* **213**:406-411.
  54. Valenzuela, P., P. Gray, M. Quiroga, J. Zaldivar, H. M. Goodman, and W. G. Rutter. 1979. Nucleotide sequence of the gene coding for the major protein of hepatitis B virus surface antigen. *Nature (London)* **280**:815-819.
  55. Young, R. A., O. Hagenbuchle, and U. Schibler. 1981. A single mouse  $\alpha$ -amylase gene specifies two different tissue-specific mRNAs. *Cell* **23**:451-458.