The Transcriptional Program of
Ground Squirrel Hepatitis Virus

by

Gregory H. Enders

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Genetics

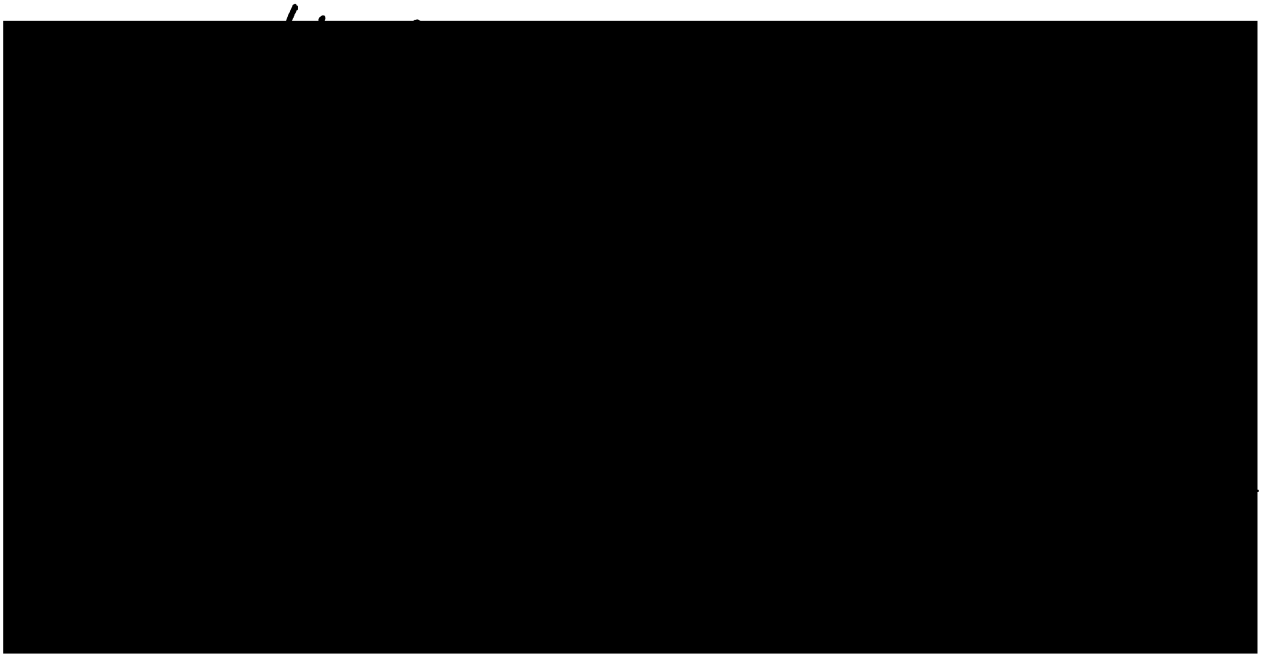
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Acknowledgements

As casual inspection reveals, Chapter 1 is reproduced from Cell and Chapter 2 and the appendix are reproduced from the Journal of Virology.

I'd like to thank the members of the Varmus, Ganem, and Bishop labs and the various members of the Department of Biochemistry and Biophysics for helping to create such a good atmosphere in which to do science.

I'd like to thank the members of my orals committee--Mike Bishop, Keith Yamamoto, Ira Herskowitz, C.C. Wang, and Bill Rutter--for their tolerance of faults and words of encouragement and advice. I'd particularly like to thank Keith Yamamoto, for being, throughout my graduate years, such an invaluable resource on the control of gene expression.

Finally, I'd like to thank Harold Varmus and Don Ganem for five delightful years of science and friendship. I wouldn't trade the experience for anything else I can think of, offhand.

Abstract

Ground Squirrel Hepatitis Virus (GSHV) is a member of a recently defined group of viruses, the hepadnaviruses. The DNA genomes of these viruses are among the smallest and most compact known and are replicated by a novel reverse transcription scheme.

We sought to understand how hepadnavirus RNAs can direct the synthesis of seven distinct proteins and a complete copy of the viral genome. We found that the GSHV genome is transcribed into two major classes of RNAs in the infected liver, 2.3 and 3.5 kb in length (G.H. Enders, D. Ganem, and H.E. Varmus, Cell 42:297-308, 1985). Both classes of RNAs are plus-stranded, unspliced, and polyadenylated at a common site and display heterogeneous 5'-ends that can encode proteins with different amino-termini. The structure of the 3.5 kb RNAs suggested that one or more was also likely to serve as template for synthesis of the viral genome, since each contains an intact copy of the viral genetic information. We found that all of the major viral RNAs are associated with polyribosomes in the infected liver, but only the shortest of the three major 3.5 kb RNAs is found within viral core particles (G.H. Enders, D. Ganem, and H.E. Varmus, J. Virol. 61:35-41, 1987). This finding corroborates evidence implicating this latter RNA in two steps of viral DNA synthesis (C. Seeger, D. Ganem, and H.E. Varmus, Science 232:477-484, 1986) and indicates that the structural homogeneity of the viral genome can be explained by selective RNA packaging. The selectivity in packaging, in which some unpackaged RNAs possess all of the sequences of the packaged RNA and have only 21

additional 5'-terminal nucleotides, is unprecedented. We have proposed four models to explain this observation.

Our analysis of GSHV DNA synthesis confirms the central finding obtained with DHBV (J. Summers and W.S. Mason, Cell 29:403-415, 1982), that the minus strand appears to be synthesized from an RNA template and the plus strand from completed minus strands. Our results also suggest that completion of DNA synthesis is accompanied by association with membranes, consistent with current models of virion morphogenesis.

We designed several strategies for selecting cells that can efficiently use the promoter for the 3.5 kb RNAs. These approaches may improve experimental access to some tissue-specific transcription factors.

Finally, we initiated experiments to address why polyadenylation of the 3.5 kb RNAs occurs only after the second transit of the transcription complex past the viral polyadenylation site. This phenomenon may point to a general requirement for polyadenylation of eucaryotic mRNAs.

Table of Contents

Introduction

The hepatitis B viruses (hepadnaviruses) are undoubtedly among the most efficient parasites of higher eucaryotes. Hepadnaviruses carry the most compact genomes yet identified and are the only DNA viruses of animals known to synthesize their genomes from RNA templates, relying on the host cell RNA polymerase for genome amplification. Viral RNAs containing only 3.3 kilobases of unique sequence direct the synthesis of at least seven distinct proteins and a novel DNA genome. These two functions place fundamentally different, in some respects conflicting, demands on viral RNA, yet viral antigen titers in serum can reach remarkably high levels (Ganem, 1982). How hepadnaviruses so successfully exploit the host cell transcriptional and translational machinery is the subject of this thesis.

Two other groups of viruses are known to employ RNA-directed DNA synthesis (reverse transcription). This method of DNA synthesis was discovered in studies of retroviruses, a large group of RNA viruses of animals, and has recently been found in caulimoviruses, a small group of DNA viruses of plants (Varmus and Swanstrom, 1985). Among these three groups, the hepadnavirus replication strategy appears the most divergent. For example, HBV genomic DNA is covalently linked to a protein, a relationship that appears to result from the protein priming the reverse transcription reaction (Gerlich and Robinson, 1980; Molnar-Kimber et al., 1984). In contrast, retroviruses and caulimoviruses appear to use primers composed exclusively of RNA (Varmus and Swanstrom, 1985). Further differences in the replication schemes of these viruses are described below. These differences have important implications both for the viruses and for their host cells, bearing on

such issues as the types of mutations caused by integration of viral DNA into host DNA and the potential for transduction of cellular genetic information.

In addition to its role in viral replication, reverse transcription has increasingly been implicated in the transposition of genetic elements in diverse eucaryotic species and in the duplication or modification of normal cellular genes (Baltimore, 1985; Fink, 1987). Some 10% of the genetic information in higher eucaryotes appears now to be the product of reverse transcription events (Baltimore, 1985). Some of these events have strictly followed the retroviral model for reverse transcription, others not. For most, no experimental systems are available--we can only attempt to infer what has occurred from the structure of the end products. The hepadnaviruses have presented a major new experimental system for delineating common and uncommon features of reverse transcription schemes and their implications for viral and host cell economies.

Hepadnaviral replication is of interest from several other perspectives, as well. Hepadnaviruses replicate primarily in the liver (cf. Ganem et al., 1982). Although this host cell restriction probably has multiple causes, hepadnaviral DNA appears to be preferentially transcribed in liver and liver-derived cell lines (Babinet et al., 1986; Jameel and Siddiqui, 1986). Hepadnaviruses may therefore be useful experimental tools for studies of tissue-specific gene expression, in a cell type that has become increasingly attractive for such studies (Costa et al., 1986; Gorski et al., 1986). The viruses harbor transcriptional and translational signals selected for efficient function in hepatocytes and offer a means of introducing genetic

information into hepatocytes in vitro and in vivo.

In addition, hepadnaviral proteins have served as useful models in studies of protein topogenesis (Eble et al., 1986). Viral surface antigens aggregate in the endoplasmic reticulum (ER) into particles that are secreted from the cell (Eble et al., 1986; Patzer et al., 1986). The virions are thought to exit by this pathway as well, after budding into the ER. Moreover, evidence has been obtained that structurally distinct viral surface antigens associate in the ER in ways that regulate their secretion (Persing et al., 1986; Standring et al., 1986; Chisari et al., 1986; Cheng et al., 1986). Similar regulatory interactions may occur during synthesis and transport of host constituents. Given the unusually small sizes of the hepadnaviral genomes and the limited number of viral protein species required for virion morphogenesis, hepadnaviruses offer attractive systems for studying viral assembly per se and general properties of protein topogenesis in eucaryotes.

HBV is also a major human pathogen, infecting an estimated 200 million people (Ganem, 1982). Primary infection can go unnoticed or can be fatal. In addition, some 5 to 10 % of primary infections are not resolved, leaving the individual a chronic carrier of the virus or viral antigens, at risk for progressive liver failure and the development of hepatocellular carcinoma (HCC; Beasley et al., 1981). HCC is the most common fatal human malignancy, and persistent infection with HBV appears to be its principal cause. Although vaccines against HBV infection have recently been developed, they are expensive and not uniformly effective. Given the prevalence of the virus and its propensity both for persistence and for vertical transmission, HBV is

certain to remain a major world health problem for years to come. Initial attempts to understand why these viruses are oncogenic have gone largely unrewarded. Because ongoing viral replication appears to be important in oncogenesis, knowledge of how replication occurs may further an understanding of tumor induction.

When this work began, there were no cell lines that supported hepadnaviral replication. Given this fact and the technical difficulties of working with a virus that infects only humans and chimpanzees, we chose to work with a recently discovered close relative of HBV, Ground Squirrel Hepatitis Virus (GSHV; Marion et al., 1980). Chapter one presents the structure of the major GSHV transcripts found in infected liver and discusses the implications of their structures for synthesis of viral proteins and DNA. In chapter two, we analyse the subcellular distribution of these RNAs, identifying those that serve as messenger RNAs and those that serve as templates for reverse transcription of the genome. The results reveal a novel selectivity in packaging of RNA into viral core particles, and we offer several models to explain this selectivity. Chapter three presents data on the structure and subcellular distribution of viral DNAs and discusses their implications for our current understanding of virion morphogenesis. Chapter four describes our efforts to find or generate cell lines that support viral replication. This work may lead to cell lines that better preserve liver-specific gene expression. Chapter 5 presents our initial efforts to understand why polyadenylation of several of the major GSHV RNAs occurs only after the second transit of the transcription complex past the viral polyA addition site. This phenomenon is a common feature of genetic elements that employ reverse

transcription and may point to a general requirement for polyadenylation of eucaryotic mRNAs.

A thorough review of the molecular biology of hepadnaviruses, authored by the two supervisors of this thesis project, has recently been published (Ganem and Varmus, 1987).

References

Babinet, C., Farza, H., Morello, D., Hadchovel, M., and C. Pourcel (1985) Science 230:1160-1163

Baltimore (1985) Cell 40, 481-482

Beasley, R.P., Lin, C.-C., Hwang, L.-Y., Chien, C.-S. (1981) Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22,707 men in Taiwan. Lancet 2, 1129-1133

Cheng, K.-C., Smith, G.L., and B. Moss (1986) J. Virol. 60:337-344

Chisari, F., Filippi, P., Milich, D., McLauchlin, A., Brinster, R., and R. Palmiter (1986) J. Virol. 60:880-887

Costa, R.H., Lai, E., and J.E. Darnell (1986) Moll. Cell. Biol. 6:4697-4708

Eble, B., Lingappa, V., and Ganem, D. (1986) Mol. Cell. Biol. 6, 1454-1463

Fink (1987) Cell 49, 5-6

Ganem, D. (1982) Persistent infection of humans with hepatitis B virus: mechanisms and consequences. Rev. Infect. Dis. 4, 1026-1047

Ganem, D., Weiser, B., Barchuk, A., Brown, J., and H.E. Varmus (1982) J. Virol. 44:366-373

Ganem, D. and Varmus, H.E. (1987) Ann. Rev. Biochem. 56, 651-693

Gorski,K., Carneiro, M., and U. Schibler (1986) Cell 47:767-776

Gerlich, W.H. and Robinson, W.S. (1980) Hepatitis B virus contains protein attached to the 5' terminus of its complete strand. Cell 21, 801-809

Jameel, S. and A. Siddiqui (1986) Moll. Cell. Biol. 6:710-715

Marion, P.L., Oshiro, L., Regnery, D.C., Sculland, G.H., and W.S. Robinson (1980) Proc. Natl. Acad. Sci. USA 77:2941-2945

Molnar-Kimber, K.L., Summers, J., Taylor, J.M., and Mason, W.S. (1984) Mapping of the cohesive overlap of duck hepatitis B virus DNA and of the site of initiation of reverse transcription. J. Virol. 51, 181-191

Patzer, E., Nakamura, G., Simonsen, C., Levinson, A., and Brands, R. (1986) J. Virol. 58, 884-892

Persing, D., Varmus, H., and Ganem, D. (1986) Science 234, 1388-1391

Standring, D., Ou, J., and W. Rutter (1986) Proc. Natl. Acad. Sci. USA 83:9338-9342

Varmus, H.E. and Swanstrom, R. (1985) In Molecular Biology of RNA tumor viruses, Chapter 5. New York: Cold Spring Harbor Press

Chapter 1

Mapping the major transcripts of Ground Squirrel Hepatitis Virus: the presumptive template for reverse transcriptase is terminally redundant

Cell, Vol 42, 297-308, August 1985, Copyright © 1985 by MIT 0092-8674/85/080297-12 $02.00/0

Mapping the Major Transcripts of Ground Squirrel Hepatitis Virus: The Presumptive Template for Reverse Transcriptase Is Terminally Redundant

Greg H. Enders,* Don Ganem,†‡ and Harold Varmus*†
Division of Genetics
* Department of Biochemistry and Biophysics
† Department of Microbiology and Immunology
‡ Department of Medicine
University of California, San Francisco
San Francisco, California 94143

Summary

The ground squirrel hepatitis virus (GSHV) belongs to a recently defined class of viruses with open circular DNA genomes that encode proteins in extensively overlapping reading frames and appear to replicate via RNA intermediates. We have determined the primary structure of the major GSHV transcripts in the livers of infected ground squirrels. Both major classes of transcripts, 2.3 kb and 3.5 kb, are plus-stranded, unspliced, polyadenylated at a common position, and display heterogeneous 5′ ends that can encode proteins with different amino termini. The 2.3 kb transcripts, like their structural analogs transcribed from human hepatitis B virus DNA, are likely mRNAs for products of the major surface antigen and presurface coding domains. The 3.5 kb transcripts are likely mRNAs for one or more products of the core antigen reading frame; these transcripts also encompass the entire genome and contain terminal redundancies of 130–160 nucleotides that include a putative initiation site for reverse transcription.

Introduction

Ground squirrel hepatitis virus (GSHV) is a member of a group of DNA viruses, the hepadnaviruses (Robinson et al., 1981), that infect the livers of vertebrates (Marion et al., 1980; Ganem et al., 1982). Persistent infection with human hepatitis B virus (HBV), the prototypic member of this family, is associated with chronic inflammation of the liver, cirrhosis, and the development of primary hepatocellular carcinoma (Beasley et al., 1981; Ganem, 1982). Strong evidence has been obtained that these viruses synthesize their genomes from RNA (Summers and Mason, 1982; Miller et al., 1984). If this is true, then the hepadnaviruses, like retroviruses and certain other mobile genetic elements of eukaryotes (Varmus and Swanstrom, 1982; Boeke et al., 1985) must employ RNA both as messenger and as genetic material. Because eukaryotic messenger RNA is as a rule monocistronic, protein synthesis and genome replication appear to place divergent demands on the viral transcriptional program. Unlike viruses for which RNA is the sole genetic material, genetic elements that use reverse transcription for replication or transposition appear to rely upon host transcriptional machinery for RNA synthesis from DNA templates. Hence, these elements offer examples of how the eukaryotic transcriptional and translational machinery can be exploited to meet stringent transcriptional and translational requirements.

The genomes of hepadnaviruses, 3.0–3.3 kb in length are the smallest yet identified among DNA viruses that infect animals. One DNA strand (the minus strand) encodes all of the long open reading frames (Tiollais et al., 1981; Seeger et al., 1984a). This strand has a protein covalently bound to its 5′ end (Gerlich and Robinson, 1980; Ganem et al., 1982). The 5′ end of the plus strand of genomic DNA is invariant in position and bridges the ends of the minus strand, forming an open circle, whereas the 3′ end is variable in position. Viruses in serum contain a DNA polymerase that can extend the incomplete plus strands when deoxyribonucleoside triphosphates are provided (Summers et al., 1975; Landers et al., 1977).

Studies of the DNA polymerase activities in core particles isolated from livers infected with the duck hepatitis B virus (DHBV) led Summers and Mason (1982) to propose a model for hepadnavirus replication in which minus strand DNA synthesis uses RNA as template, and plus strand synthesis uses the completed minus strand as template. The fact that even short minus stands in infected liver (Weiser et al., 1983) or in isolated core particles (Molnar-Kimber et al., 1983) are linked to protein suggests that this protein may prime minus-strand synthesis. If so, this mechanism would be unique among elements known to employ reverse transcription.

The following four long open reading frames are present in mammalian hepadnavirus genomes: frames for the core and surface antigens (cAg and sAg), a frame tentatively assigned to the viral DNA polymerase (reverse transcriptase), and one unassigned frame (called B or X). Moreover, the coding sequence for the major sAg (p24) is immediately preceded by in-frame sequences that, in the case of HBV, have been shown to contribute to two additional polypeptides, p45 and p31, initiated at AUG codons in the pre-S region (Heermann et al., 1984; Machida et al., 1984; Persing et al., 1985). Preliminary evidence suggests that the GSHV sAg frame also produces more than one protein (D. Persing, unpublished results). Thus, the mammalian hepadnaviruses appear to contain the coding information for at least six distinct proteins.

The only hepadnaviral transcripts that have been well characterized are the 2.3 kb HBsAg transcripts produced in HBV-transfected L cells (Standring et al., 1984) and HBV-infected chimpanzees (Cattaneo et al., 1984). These transcripts display variable 5′ ends surrounding the p31 AUG codon and are candidate mRNAs for the p31 and p24 proteins. In this work, we have determined the primary structure of the two major classes of viral transcripts in the livers of ground squirrels infected with GSHV. Both classes are plus-stranded, unspliced, initiated at multiple sites, and polyadenylated at a common position. One class of transcripts is 2.3 kb in length and has two major 5′ ends that bracket the AUG codon for the putative GSHV p31 protein; thus, it resembles in detail the HBsAg tran-

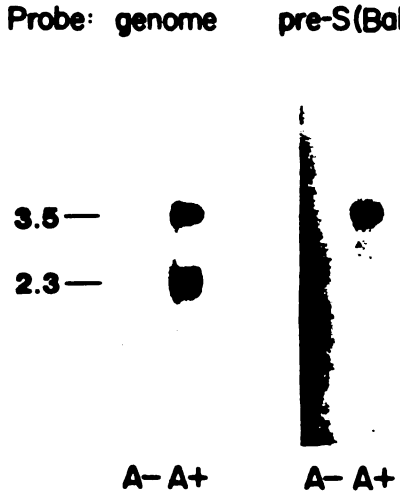


Figure 1. Identification of the Major Size Classes of GSHV RNA in Infected Liver

Poly(A)$^+$ and poly(A)$^-$ RNA was prepared from the liver of a ground squirrel infected with GSHV. Two hundred fifty nanograms of poly(A)$^+$ and 1.5 μg of poly(A)$^-$ RNA were electrophoresed in parallel through a 1% agarose gel in the presence of formaldehyde, transferred to nitrocellulose, and hybridized to the indicated GSHV probes. The genomic probe hybridizes only to plus-stranded sequences; the Bal I-Pst I probe (nucleotides 1110–1516; see Figure 2 for the location of the probe) is not strand-specific. The sizes of the GSHV RNAs were estimated by comparison with the ribosomal RNA bands in the lanes containing poly(A)$^-$ RNA.

Figure 2. S1 Mapping with Unit-Length GSHV DNA Fragments

Unit-length fragments of GSHV DNA linearized at the indicated restriction sites were added to carrier RNA in the presence or absence of poly(A)$^+$ RNA from infected liver. The nucleic acids were heat-denatured in formamide, allowed to reanneal at 50°C for 3 hr, and digested with S1 nuclease. The protected products were electrophoresed through 1% alkaline agarose gels, transferred to nitrocellulose, and probed with plus-stranded GSHV sequences. The exposures of the first six lanes were derived from a single filter. The positions of the restriction sites are as follow: Apa I, 2288; Pvu II, 3166; Eco RI, 972; Xba I, 1783 (Seeger et al., 1984a). The diagram below the data depicts the probes as thin lines, their protected products as thick lines, and the RNA structures inferred from them as wavy lines. The viral long open reading frames are given in boxes at the bottom, with each reading frame in a different shade. (GSHV DNA is 3311 nucleotides long; position 1 is assigned to the first nucleotide in the core antigen reading frame. S and C are in the same frame if one begins reading in the C frame, although they are in different frames if one begins reading in the S frame, since the number of nucleotides in the genome is not a multiple of three.)

scripts. The other major class is composed of transcripts, 3.5 kb in length, that encompass the entire genome and are terminally redundant. These transcripts have three major 5′ ends surrounding the first AUG codon in the cAg reading frame. A short direct repeat sequence (DR1), conserved in all hepadnavirus genomes, is included in the terminal redundancies and is the likely site of initiation of reverse transcription.

Results

Preliminary Characterization of Viral Transcripts

We isolated RNA from the liver of a ground squirrel chronically infected with GSHV of known nucleotide sequence (see Experimental Procedures; Seeger et al., 1984a). This animal developed typical serum markers of viral infection and was viremic until sacrificed for this study, 12 months after inoculation (data not shown). Whole-cell RNA was separated into polyadenylated and nonpolyadenylated fractions by oligo(dT)-cellulose chromatography. Samples of each were then electrophoresed in parallel through a 1% agarose gel in the presence of formaldehyde, transferred to nitrocellulose, and probed with the complete GSHV minus strand cloned in an M13 vector (Figure 1a). Two major species of viral RNA were detected, polyadenylated transcripts of approximately 2.3 kb and 3.5 kb. Several polyadenylated transcripts of between 5 kb and 9 kb can also be seen on longer exposure of the film presented but represent less than 5% of the total viral RNA. No additional species were detected in the poly(A)$^+$ or poly(A)$^-$ fractions upon reprobing the filter for minus-strand sequences (data not shown). The 2.3 kb and 3.5 kb RNAs were also the predominant viral transcripts in RNA prepared from the livers of several other ground squirrels during both acute and chronic phases of infection (data not shown).

The 2.3 kb transcript is similar in size to the HBsAg transcripts seen in L cells transfected with cloned HBV DNA (Standring et al., 1984) and in livers of HBV-infected chimpanzees (Cattaneo et al., 1984). The HBsAg transcripts are initiated at multiple positions near the 3′ end of the pre-S region. As a first step toward defining the primary structure of the GSHV transcripts, we probed a filter similar to that presented in Figure 1a with a GSHV DNA clone span-

Figure 3. 5' Ends of the 2.3 kb Transcripts Determined by S1 Nuclease and Primer-Extension Methods

(a) S1 mapping was performed on poly(A)$^+$ RNA from infected liver using a 5' end-labeled Sfa NI fragment (nucleotides 1227–1997). The denatured nucleic acids were allowed to reanneal at 48.5°C for 20 hr. The S1 protected products were analyzed on a 5% polyacrylamide-urea salt gradient gel. The minus strand was labeled at nucleotide 1991. The major protected products are 770 ± 40, 485 ± 20, and 452 ± 20 nucleotides in length. The diagram below depicts the minus strand of the probe as a thin line and the protected products as thick lines. Also depicted are the translation initiation sites for the major GSHV sAg protein p24 and the putative GSHV p31 protein.

(b) Primer-extension was performed on poly(A)$^+$ RNA from uninfected and infected liver using a Kpn I-Xba I fragment (nucleotides 1610–1783) labeled only at nucleotide 1767 of the minus strand. The extended products were electrophoresed in a 6% polyacrylamide-urea gel. The major extended products are 281 ± 4 and 236 ± 4 nucleotides in length, as depicted below. M, marker DNA fragments of the indicated lengths. Assignment of the 5' end at 1527 was made by taking the mean of the two determinations derived from the above data.

ning the first two-thirds of the pre-S frame, from a Bal I to a Pst I cleavage site (Figure 1b; see Figure 2 for the location of the probe). This probe hybridized only to the 3.5 kb transcript, indicating that, as in the HBV transcripts, few or none of these sequences are present in the GSHV 2.3 kb transcript.

S1 Mapping with Unit-Length Fragments

To determine the approximate locations of the discontinuities (ends or splice junctions) in the two major transcripts, we next performed S1 protection experiments using GSHV DNA fragments of unit length. The fragments were prepared from a plasmid that contains two tandem copies

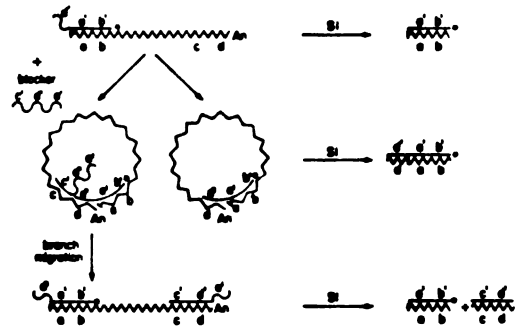


Figure 4. Formation and Disruption of RNA–DNA Circular 'Snapback' Structures during S1 Mapping

See text for discussion.

of the GSHV genome cloned into the Eco RI site of a derivative of pBR322. The plasmid was digested with restriction enzymes that cleave once in each GSHV monomer, and the unit-length fragments were isolated from agarose gels. The fragments were added to carrier RNA in the presence or absence of polyadenylated RNA from infected liver; following heat denaturation, the nucleic acids were incubated at a temperature that favors formation of RNA–DNA hybrids and digested with the single-strand-specific nuclease S1. The protected species were electrophoresed through 1% alkaline agarose gels, transferred to nitrocellulose, and probed with plus-stranded GSHV sequences.

The results demonstrate two classes of protected fragments (Figure 2a). One class consists of fragments 2.4 kb or greater; these will be discussed in a later section. The second class is composed of fragments from 0.8 kb to 2.0 kb. The sizes are consistent with protection by an unspliced transcript with its 5' end late in pre-S and its 3' end early in the core gene, that is, a transcript with the same coordinates as the HBsAg transcripts previously described (Standring et al., 1984; Cattaneo et al., 1984). The proposed RNA should protect approximately 2.0 kb of a unit-length fragment linearized at the Eco RI site and 0.8 kb and 1.2 kb regions of a unit-length fragment linearized at the Apa I site. RNA protection of unit-length fragments linearized at either Pvu II or Xba I should yield fragments of 1.7 kb and 0.3 kb, although the latter could be expected to be inefficiently transferred to nitrocellulose and, hence, not detected in our assay. These predictions were verified by the results in Figure 2; however, the presence of short exons at either end of the transcript cannot be excluded by these data.

The 5' Ends of the 2.3 kb Transcripts

We next used S1 and primer-extension methods to define the 5' ends of the 2.3 kb transcripts with greater precision. We hybridized a 5' end-labeled Sfa NI fragment (nucleotides 1227–1997), spanning the late pre-S region, to RNA from infected liver and subjected the hybrids to S1 nuclease digestion (Figure 3a). Some probe was completely protected, presumably by annealing to the 3.5 kb transcripts (see below). The two other major bands of pro-

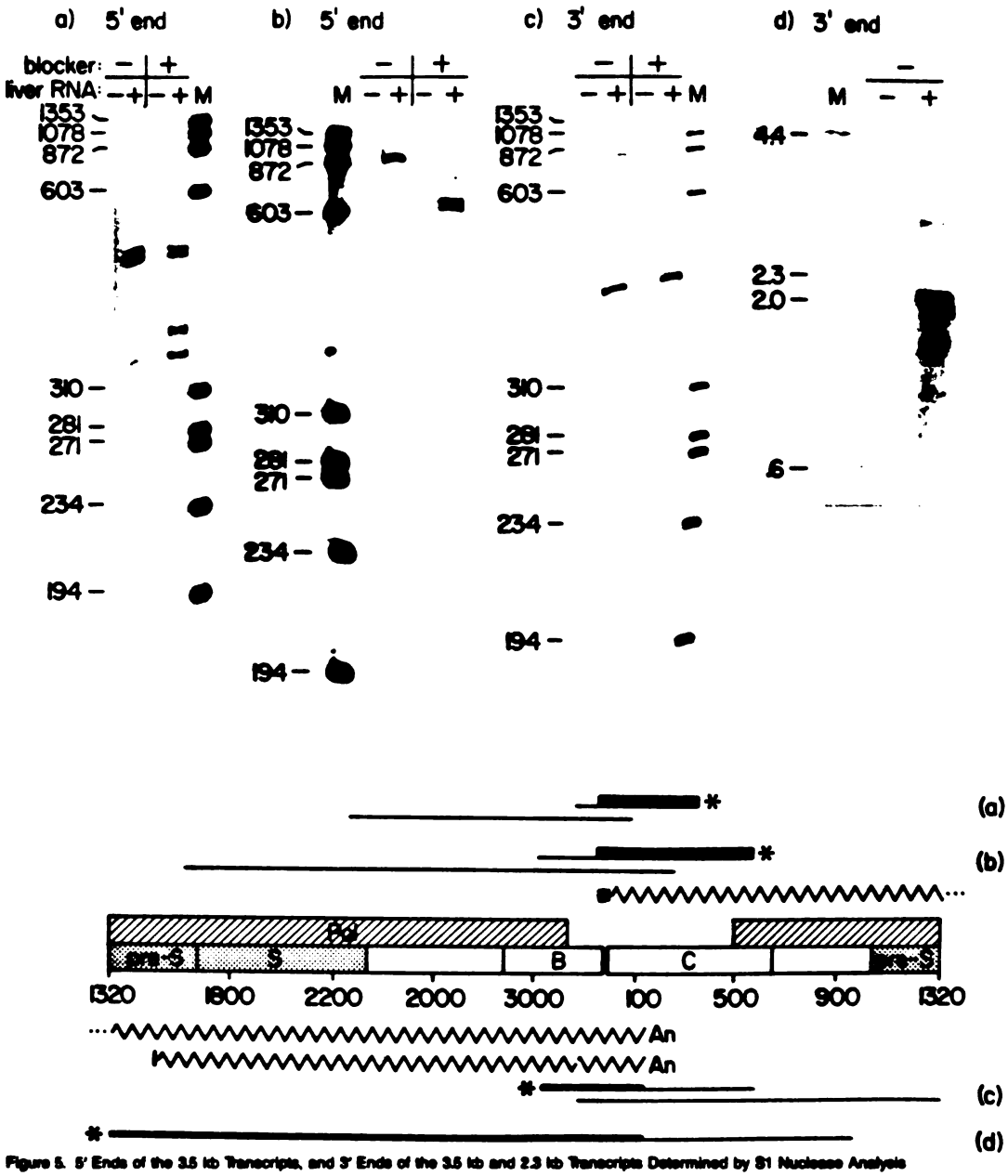


Figure 5. 5' Ends of the 3.5 kb Transcripts, and 3' Ends of the 3.5 kb and 2.3 kb Transcripts Determined by S1 Nuclease Analysis

(a) The 5' end-labeled Sph I-Bcl I fragment (nucleotides 3153–345) was added to carrier RNA in the presence or absence of poly(A)$^+$ RNA from infected liver and the 3' blocker Ape I-Xho 2 (nucleotides 2288–69). Hybridization of the denatured nucleic acids was at 43°C for 4.5 hr. The S1 products were analyzed on a 5% polyacrylamide-urea salt gradient gel. The minus strand of the probe was labeled on nucleotide 348. (A partial cleavage of the fragment with Pvu II made some of these strands 15 nucleotides shorter at their 3' ends.) The major protected products are 606 ± 30, 561 ± 30, 396 ± 15, 376 ± 15, and 351 ± 15 nucleotides in length. The diagram below presents the minus strand of the probe as a thin line, the protected products as thick lines, and the deduced RNAs as wavy lines.

(b) The 5' end-labeled Bst EII-Hga I fragment (nucleotides 3025–575) was added to carrier RNA in the presence or absence of RNA from infected liver and the 3' blocker Kpn I-Stu I (nucleotides 1610–280). Hybridization was at 47°C for 20 hr. The S1 products were analyzed on a 5% polyacrylamide-urea salt gradient gel. The minus strand of the probe was labeled at nucleotide 589. The major protected products are 575 ± 40, 503 ± 20, and 585 ± 20 nucleotides in length.

(c) The 3' end-labeled Bst EII-Hga I fragment (nucleotides 3025–575) was added to carrier RNA in the presence or absence of poly(A)$^+$ RNA from infected liver and the 5' blocker Pvu II-Mst II (nucleotides 3166–1324). Hybridization was at 47°C for 20 hr. The S1 products were analyzed on a 5% polyacrylamide-urea salt gradient gel. The 3' end of the minus strand was at nucleotide 3027. The major protected products are 575 ± 40 and 410 ± 15 nucleotides long.

(d) The 3' end-labeled Mst II-Eco RI fragment (nucleotides 1324–872) was added to carrier RNA in the presence or absence of poly(A)$^+$ RNA from infected liver. Hybridization was at 53°C for 20 hr. The S1 products were analyzed on a 1% alkaline agarose gel, transferred to nitrocellulose, and probed with GSHV plus-strand sequences. The major protected products are 2.9 kb and 2.1 kb long.

tected products, 495 ± 20 and 452 ± 20 nucleotides long, indicated 5′ ends (or splice junctions) of the 2.3 kb RNAs at positions 1496 ± 20 and 1539 ± 20 on the conventional map of GSHV DNA. S1 analysis using a probe 5′ end-labeled at a site nearer these ends (the Eco RI-Xho II fragment, labeled at nucleotide 1587 of the minus strand) more precisely identified the positions of the 5′ ends as nucleotides 1486 ± 3 and 1524 ± 4 (data not shown).

To determine if these discontinuities were true ends, we next performed primer extension using a 5′ end-labeled Kpn I-Xba I fragment (nucleotides 1610–1763) annealed to RNA from uninfected and infected liver (Figure 3b). Extension of the primer on RNA from infected liver yielded two major cDNA products, the lengths of which indicate the presence of 5′ ends at positions 1486 ± 4 and 1529 ± 4. The 2.3 kb transcripts are, therefore, initiated at two major 5′ ends within a 45 nucleotide region. Only one transcript includes the most distal in-phase pre-S AUG codon, at nucleotides 1497–1499. Thus, one transcript could serve as mRNA for only the p24 product of the sAg reading frame, whereas the larger could direct synthesis of either p24 or p31. A more precise determination of the 3′ ends of these RNAs will be presented below.

Prevention of "Snapback" during S1 Analysis

The other class of protected fragments seen in Figure 2 is represented by bands at the position of 3.3 kb fragments in each lane containing RNA from infected liver. (One exception, the band at the position of 2.4 kb fragments in the experiment with Eco RI-cleaved DNA, can be explained, in retrospect, by protection with the 3′ end of the 3.5 kb RNAs; see below.) These 3.3 kb bands result from complete or nearly complete protection of each unit-length GSHV fragment by the liver RNA; their absence from the control lanes shows that they are not caused by reannealing of the two strands of input DNA. These results were surprising because they suggested that GSHV RNA could fully protect unit-length GSHV fragments generated by cleavage at widely spaced sites. The annealing of more than one RNA molecule to each fragment seemed unlikely, as no more than 10% of the input DNA was protected in each experiment. In addition, the relative abundance of the two classes of protected fragments did not vary with the amount of RNA from infected liver (data not shown).

The simplest explanation for protection of the 3.3 kb fragments is the formation of an open circular duplex composed of single 3.5 kb RNAs and unit-length DNA strands, by a "snapback" mechanism (Figure 4). If the 3.5 kb transcripts contain complete, uninterrupted copies of plus-strand sequences, then any GSHV fragment that spans one end of the RNA should also be able to hybridize with the other end of the RNA, forming an open circle. The poly(A) tail and any sequences repeated in the RNA would protrude, and S1 nuclease would be unlikely to cleave the DNA probe at the discontinuity in the RNA (Meyers et al., 1985). To map the 3′ and 5′ ends of the 3.5 kb RNA with labeled probes and S1 nuclease, it was necessary to inhibit the snapback reaction; we chose to do this by including a second, unlabeled "blocker" fragment in the annealing reaction (Figure 4).

For mapping 5′ ends, the blocker fragments contained sequences complementary to the 3′ end of the transcript and sequences (c′ in Figure 4) that would remain free to hybridize to the RNA, even if the labeled probe had already bound to the 3′ end. In this case, random branch migration could then allow the blocker to anneal to the entire 3′ end. When the branch point reaches the 3′ end of the RNA, the circular structure will be converted to a linear form, with RNA:DNA duplexes at both ends. This form should be stable; all of the RNA sequences that can hybridize with either DNA fragment are now duplex. Likewise, the probe should be able to displace any 3′ blocker already bound to the 5′ end of the RNA, by virtue of the sequences unique to the probe (b′ in Figure 4). A similar argument applies to determination of the 3′ end of the transcript.

The 3.5 kb Transcript Is Terminally Redundant

The results of an informative S1 mapping experiment to identify the 5′ end of the 3.5 kb transcript are presented in Figure 5a. A 5′ end-labeled Sph I-Bcl I fragment (nucleotides 3153–345; the genome is 3311 nucleotides long; Seeger et al., 1984a) was hybridized to RNA from infected liver in the presence and absence of an Apa I-Xho II 3′ blocker fragment (nucleotides 2081–99), the hybrids were digested with S1 nuclease, and the products were analyzed on denaturing polyacrylamide gels. In the absence of blocker, the predominant bands correspond to complete protection of the probe. (The probe runs as a doublet because of a partial secondary cleavage at the Pvu II site, 13 nucleotides from the Sph I site; this does not affect interpretation of the other bands.) These bands are not due to self-annealing of the probe, because they are not seen in the control experiment (without added liver RNA) and because the labeled 5′ ends of the probe are protruding and are not protected by self-annealing. In addition to the predominant bands, several minor bands are seen that correspond to fragments which migrate more rapidly in the gel. In the presence of the blocker, complete protection of the probe is reduced about 5-fold, with a concomitant increase in the signal in the more quickly migrating bands. These latter bands represent labeled fragments protected by the 5′ ends of the 3.5 kb RNAs and imply the existence of multiple discontinuities near the beginning of the core frame, at the three major positions 3310 ± 15, 3284 ± 15, and 3260 ± 15.

We confirmed these conclusions by using a 5′ end-labeled Bst EII-Hga I fragment (nucleotides 3025–575) in the presence or absence of a Kpn I-Stu I 3′ blocker fragment (nucleotides 1610–260) (Figure 5b). In this case, the blocker had a more dramatic effect on the size and relative proportions of the protected products, but the position of the 5′ discontinuities remained unchanged: 4 ± 20, 3297 ± 20, and 3295 ± 20. Moreover, the results in Figure 5b indicate that all, or nearly all, the 3.5 kb RNAs possess discontinuities at these sites. This conclusion is further supported by S1 mapping and primer-extension results presented below.

Analogous experiments were performed to map the 3' ends of the 3.5 kb transcripts. We anticipated that the results would be complicated by the presence of the 3' ends of the 2.3 kb transcripts. The Bst EII-Hga I fragment employed above was labeled on its 3' end and hybridized to RNA from infected liver in the presence and absence of a Pvu II-Mst II 5' blocker fragment (nucleotides 3166–1324; Figure 5c). In the absence of blocker, two major species of protected DNA are observed. The longer fragment resulted from complete protection of the probe. The shorter corresponds to protection by RNA with a 3' end at approximately the same position as that identified in the experiments presented in Figure 2. This fragment very likely results, at least in part, from protection by the 2.3 kb transcripts, though some protection could be provided by the 3.5 kb transcripts. In the presence of the blocker, which could compete for annealing only with the 3.5 kb transcript, complete protection of the probe is abolished, but no new band arises. This suggests that the 3.5 kb transcript is 3' coterminal with the 2.3 kb transcript, at position 126 ± 15.

To corroborate this, a 3' end-labeled Mst II-Eco RI fragment (nucleotides 1324–972) was hybridized to RNA from infected liver. The radioactive label on the minus strand of this fragment is at the Mst II site in the pre-S region, a site demonstrated earlier (Figure 1b, Figure 2, and Figure 3) to be present only in the 3.5 kb transcript. Hence, the 3' end of this fragment can be protected only by the 3.5 kb transcript. The major protected species (Figure 5d) corresponds to that predicted by the results in Figure 5c. The position of the 3' end was determined more precisely by using a 3' end-labeled Cla I-Bcl I fragment (nucleotides 3244–345). This experiment placed the 3' end at nucleotide 133 ± 8 (data not shown). The 3' end is, therefore, about 25 nucleotides 3' to the variant polyadenylation signal UAUAAA at nucleotides 106–111, probably at the CA dinucleotide at positions 130–131 (Fitzgerald and Shenk, 1981).

Snapback was apparently less of a problem during the analysis of 3' ends, perhaps because the 3' ends are substantially higher in GC content than are the 5' ends of the 3.5 kb RNAs (Seeger et al., 1984a). This may also provide an explanation for the presence of the protected 2.4 kb Eco RI-cleaved DNA in Figure 2. The Eco RI site divides the 3.5 kb transcript into domains that differ by six degrees in their predicted optimal temperatures for RNA–DNA duplex formation (Thomas et al., 1976).

Confirmation of the 5' Ends of the 3.5 kb Transcripts by Primer Extension

To confirm that the 5' ends of the 3.5 kb transcript identified above were the ends of the RNAs rather than splice junctions, we performed a primer-extension experiment with a synthetic primer composed of minus-strand nucleotides 167–148. The primer was labeled at its 5' end, annealed to RNA from uninfected and infected liver, and extended with AMV reverse transcriptase (Figure 6). Three bands are detectable in the lane containing the products from primer extension on RNA from infected liver. The bands correspond to extended products 161, 183, and 190 nucleotides in length and place the 5' ends of the 3.5 kb transcript at nucleotides 6 ± 3, 3295 ± 4, and 3288 ± 4. These results correspond well to the major 5' ends deduced from the S1 analyses above. Longer exposure of a film similar to that presented in Figure 6 shows an additional primer-extension product that corresponds to the faint band detectable just above the three major bands in Figures 5a. However, we estimate that this end represents less than 5% of the 5' ends, and it is not considered further.

To exclude the possibility that the primer-extension products from infected liver RNA result from unsuspected obstacles to reverse transcriptase in continuous RNA templates (e.g., a high degree of secondary structure), we subcloned the Avr II-Pst I fragment of GSHV DNA (nucleotides 3150–936) into a plasmid vector that contained a promoter for the SP6 polymerase and synthesized plus-strand GSHV RNA in vitro that lacked discontinuities in the region of proposed 5' and 3' ends of the 3.5 kb RNAs. Extension of the primer on the SP6-generated transcripts yielded no major blocks to reverse transcription (Figure 6), and the minor preferred pause sites detected on a longer exposure of this lane did not corespond to the bands generated from infected liver RNA (data not shown). The primer was also used to generate a dideoxy sequencing ladder from the SP6-generated transcript. This ladder confirmed that cDNA synthesis using this primer begins at nucleotide 147 (data not shown). Electrophoresis of the primer-extension products from liver RNA in parallel with fragments of identical base composition in the cDNA sequencing ladder confirmed the size assignments made in Figure 6 (data not shown).

In sum, these results demonstrate that the 3.5 kb transcripts form a class of transcripts with 3 major 5' ends distributed over a 30 nucleotide region. Two of these ends include the first AUG codon in the core open reading frame, at nucleotides 1–3, and all three include the second in-phase AUG codon, at nucleotides 91–93. The 3.5 kb transcripts are polyadenylated at a site copied early in their second transit of the core gene and, therefore, possess terminal redundancies of 130–160 ± 20 nucleotides.

No Other RNA Ends Were Detected

Having carefully examined two regions of the GSHV genome for RNA discontinuities, we sought them elsewhere in the RNAs, using the sensitivity afforded by end-labeled S1 probes. We scanned the rest of the genome using 5' and 3' end-labeled probes less than 1 kb in length, with a minimum overlap of 65 nucleotides between adjacent probes. Figure 7 presents the S1 probes used. No other ends were detected under conditions that allowed ready hybridization of the probes to the already characterized transcripts (data not shown), implying that additional, relatively abundant transcripts did not comigrate on agarose gels with the 2.3 kb and 3.5 kb species described above.

Discussion

We have determined the primary structure of the major GSHV transcripts in livers of infected ground squirrels.

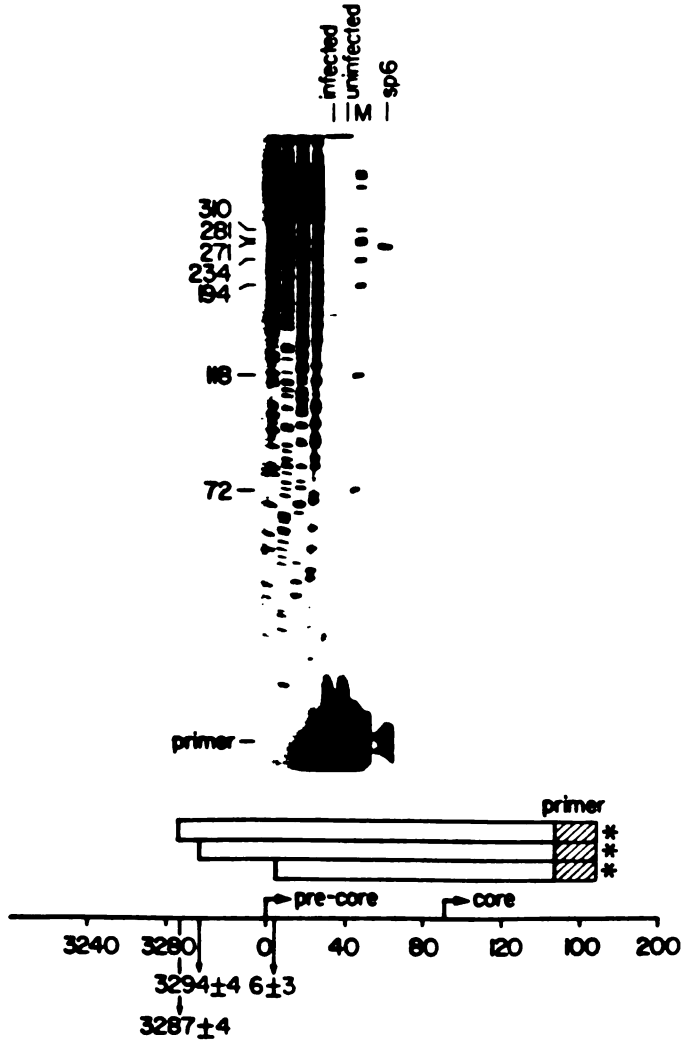


Figure 6. 5' Ends of the 3.5 kb Transcript Determined by Primer-Extension Analysis

A synthetic oligonucleotide composed of minus-strand nucleotides 167–148 was labeled on its 5' end and used to prime cDNA synthesis on RNA from uninfected and infected liver and on plus-stranded RNA generated in vitro from an SP6 vector containing the GSHV genomic sequences from this region (nucleotides 3150–936). The extended products were electrophoresed in 6% polyacrylamide-urea gels (the SP6 lane comes from a separate gel). An unrelated sequence ladder was included in the gel to facilitate calibration of fragment sizes. The sizes of the major extended products are 161 ± 3, 163 ± 4, and 190 ± 4 nucleotides long.

Both major classes of transcripts are plus-stranded, unspliced, polyadenylated at a common position, and display heterogeneous 5' ends that can encode proteins with different amino termini. The 3.5 kb transcripts encompass the entire genome and are terminally redundant. These transcripts are the putative templates for synthesis of the minus strand of genomic DNA.

Our analysis of the 3.5 kb transcripts was complicated by the apparent formation of RNA–DNA circular snapback structures. One could expect this problem to occur during the S1 analysis of other full-length or longer RNAs produced from circular or terminally redundant templates. The formation of such structures has been proposed before to account for aberrant S1 results in studies on cauliflower mosaic virus (Guilley et al., 1982) and the Moloney murine leukemia virus (Honigman et al., 1985), and may help explain the recently reported difficulty in obtaining S1 mapping data on a 3.8 kb HBV transcript (Cattaneo et al., 1984). The technique used here to disrupt the formation of the snapback structures should be applicable in such instances.

The sAg Transcripts as mRNAs

The 2.3 kb GSHV transcripts are structural analogs of the HBsAg transcripts in transfected L cells (Standring et al., 1984) and in infected chimpanzee liver (Cattaneo et al., 1984). Recent work has demonstrated that the HBV pre-S region encodes two polypeptides found in HBsAg particles in addition to the major sAg, p24 (Heermann et al., 1984; Machida et al., 1984; Persing et al., 1985). These pre-S proteins initiate at the first (p45) and last (p31) AUG codons in the pre-S region. Characterization of the HBsAg transcripts in L cells (Standring et al., 1984) and in infected chimpanzee liver (Cattaneo et al., 1984) demonstrated that they possess variable 5' ends surrounding the p31 AUG codon and provide candidate mRNAs for p31 and p24.

The fact that the 5' ends of the GSHsAg transcripts also

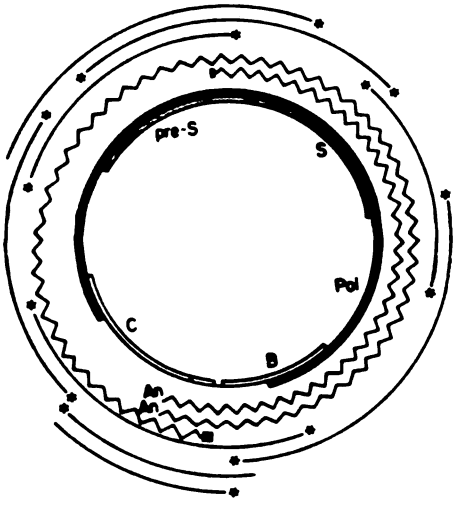


Figure 7. Summary of the Short S1 Probes Used

A summary of the short (less than 1 kb) end-labeled probes used in S1 analysis. Only the minus strands of the probes are depicted. The asterisks mark labeled ends that were used in S1 experiments; for example, most fragments were used, in different experiments, to scan in both 5' and 3' directions.

closely bracket the p31 AUG codon supports the notion that the heterogeneous initiations have functional significance for mammalian hepadnaviral gene expression. Although the 5'-most end of the GSHsAg transcripts has only 11 ± 3 nucleotides preceding the p31 AUG codon, an adenine is present in position −3 relative to this codon, a feature thought to facilitate translation initiation (Kozak, 1984). Furthermore, some eukaryotic mRNAs are known to direct translation initiation from an AUG as close as three nucleotides from the cap site (Kelley et al., 1982). There is an out-of-phase AUG codon between the 5' end of the shorter transcript and the p24 AUG, but it is followed closely by a stop codon in this same phase (Seeger et al., 1984a). (The region between the p31 and p24 AUG codons in HBV is devoid of AUG codons [Valenzuela et al., 1980; Galibert et al., 1979].)

The Genomic Transcripts as mRNAs

The GSHV genomic (3.5 kb) transcripts bear striking similarities to the genomic transcripts of retroviruses in their distribution of open reading frames. Genomic transcripts from both groups of viruses contain the coding information for core (*gag*), polymerase, and surface (*env*) proteins, in that order. (No hepadnaviral polymerase has yet been genetically mapped, but the long open reading frame present in all hepadnaviruses is of the size expected to encode such a protein and possesses homology to retroviral *pol* genes [Toh et al., 1983].) Retroviruses appear to use genomic length RNA as mRNAs for both *gag* and probably *pol* polypeptides. *env*, like sAg, is expressed by subgenomic (albeit spliced) mRNA. To what extent does expression of cAg and the putative *pol* gene in GSHV follow the retroviral model?

A situation analogous to that for expression of p31 and p24 pertains at the 5' end of the GSHV genomic transcript. Our results indicate that the synthesis of two core proteins could occur, from the AUGs at positions 1–3 and 91–93, since separate RNAs exist with each of these codons as their 5'-most in-phase initiator codons. A similar situation may obtain for HBV, which encodes a second in-phase AUG codon at an analogous position in the cAg frame. However, both viruses contain intervening out-of-phase AUG codons that are not followed by stop codons before position 91 is encountered. This may impede the use of the AUG codon at position 91 for translation initiation (Liu et al., 1984).

No transcript has been detected either for retroviruses or for GSHV that contains the first AUG codon in the polymerase frame as its 5'-proximal AUG codon. In the cases that have been investigated, the retroviral *pol* gene is expressed as a fusion protein with *gag*. This is true even for Rous sarcoma virus, where *gag* and *pol* genes are out of frame but overlap (Weiss et al., 1978; Schwartz et al., 1983). The polymerase domain is eventually freed from the polyprotein precursor in a proteolytic step catalyzed by another domain in the polyprotein. Recently, some Ty elements of yeast have been shown to express their second open reading frame as a fusion protein with the product of the preceding, overlapping reading frame; no spliced transcript can be identified that could correct the reading frame (Mellor et al., 1985; Clare and Farabaugh, 1985). An RSV-like organization of core and polymerase coding sequences is found in the GSHV genomic transcripts. Frameshifting during translation must, therefore, be considered as a candidate mechanism for expression of the GSHV polymerase gene. Alternatively, the levels of a GSHV *pol* mRNA may not have been detected by our methods.

Our data do not offer a ready explanation for the mode of synthesis of the putative GSHV p45 protein, encoded by the pre-S and surface domains. HBV transcripts initiating just 5' to the pre-S region have been synthesized in vitro (Rall et al., 1983), and S1 mapping data consistent with the presence of such transcripts have been obtained from COS cells transfected with HBV-SV40 recombinants (Laub et al., 1983). It is possible that the analogous transcripts are present in GSHV-infected livers below the level of detection in these experiments. Such a possibility is consistent with the stoichiometry of HBsAg proteins produced in vivo (Heermann et al., 1984), if they can be taken as a guide for GSHV. Other mechanisms of p45 synthesis are also plausible, however, including reinitiation of translation on the genomic transcript (Liu et al., 1984). These alternatives must also be considered for synthesis of the putative B protein, the coding region for which is present at the 3' ends of both major transcripts.

The Promoters

Why do the abundant GSHV transcripts initiate heterogeneously? Eukaryotic RNA polymerase II is thought to initiate synthesis in relation to a consensus sequence (TATAAA) usually found 30 nucleotides 5' to the start site (Breathnach and Chambon, 1981). The closest consensus or near consensus TATAAA sequence lies greater than 100 nucleotides 5' to the mapped ends of the GSeAg and

HBsAg 2.3 kb transcripts. Although the region immediately 5' to the cAg frame is AT-rich in both GSHV and HBV, neither contains a consensus TATAAA sequence closer than 80 nucleotides from the 5'-most major transcription start site. The major mammalian hepadnavirus promoters may offer examples of how a lack of precision in transcription initiation can be exploited to achieve regulated expression of more than one protein from a single reading frame. Since the cloned genome of GSHV is infectious in the animal, it may be possible to examine the affect of specific DNA sequences on transcription initiation, following site-specific mutagenesis in vitro (Seeger et al., 1984b).

Polyadenylation of the Transcripts

The GSHV sAg and genomic transcripts are 3' coterminal at a site some 25 nucleotides 3' to a variant polyadenylation signal, and it is highly likely that this is indeed the site of polyadenylation of the transcripts (Proudfoot, 1982; Fitzgerald and Shenk, 1981). No other 3' end was detected in our analysis; the signal sequence is conserved at roughly this position in all four hepadnaviruses that have been sequenced, and cDNA clones containing a poly(A) tail at this site have been isolated from HBsAg transcripts in COS cells (Simonsen and Levinson, 1983).

Our data suggest that synthesis of the genomic transcript requires transcription without polyadenylation of sequences identical with those at which the transcript is later polyadenylated. Polyadenylation on the 'first pass' of this site by transcripts initiating at the genomic promoter would result in a transcript too short to be efficiently transferred to nitrocellulose. We have attempted, as yet unsuccessfully, to detect such a transcript by S1 analysis of polyadenylated RNA from infected liver (data not shown). A 'bypass' of polyadenylation sites is also required for synthesis of the genomic transcripts of some of the other genetic elements known to use reverse transcription for replication (Varmus and Swanstrom, 1982; Guilley et al., 1982). A study of this phenomenon may help elucidate the factors, other than the consensus polyadenylation signal, that are required for efficient polyadenylation.

Reverse Transcription of GSHV RNA

The GSHV genomic transcript contains all the sequences needed to serve as template for the synthesis of a complete copy of minus-strand DNA. However, it still must be directly demonstrated that the 3.5 kb RNA serves this function. The manner in which hepadnavirus RNA is used for synthesis of genomic DNA is likely to borrow on principles from the retroviral model but is also likely to possess distinct properties. One unusual feature of hepadnaviral reverse transcription concerns the priming of minus strands. Indirect evidence suggests that priming may occur via linkage of nucleotides to protein; all or nearly all the subgenomic minus strands in GSHV-infected livers are protein-linked (Weiser et al., 1983), and nascent minus strands as short as 30 nucleotides can be found linked to protein in core particles from DHBV-infected livers (Molnar-Kimber et al., 1983).

The protein-linked 5' end of DHBV has been mapped to DR1, one of two direct repeat sequences found conserved in all hepadnaviruses genomes sequenced to date; the 5' end of the plus strand resides near the other repeat sequence, DR2 (Molnar-Kimber et al., 1984). Minus-strand DNA synthesis is likely to be primed at the DR1 site on the RNA. Because the DR1 sequence is present within the terminal redundancies of the genomic transcripts, these transcripts could serve as templates for the synthesis of minus-strand DNA from either site (Figure 8). If the DR1 sequence near the 3' end of the 3.5 kb RNA is used, synthesis of a full-length or greater-than-full-length copy of minus-strand DNA could occur on a single template; if the DR1 sequence near the 5' end is used, synthesis of short "strong stop" species would result, requiring a template switch for completion of synthesis. Particularly in the latter case, the terminal redundancy in the RNA might be used as in retroviral DNA synthesis, to facilitate transfer of minus strands to a 3' end (Varmus and Swanstrom, 1982). DR2, in contrast, is present only once in the 3.5 kb RNA, and is unlikely to be represented more than once in the minus-strand DNA.

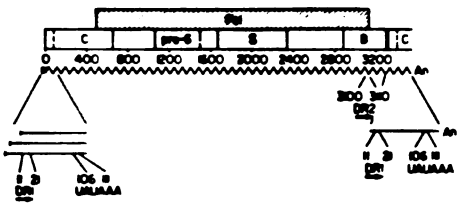


Figure 8. The GSHV Genomic Transcripts

The GSHV genomic transcripts, with significant landmarks identified. The terminal redundancies are enlarged to provide greater detail.

The location of the DR1 sequence suggests that priming occurs at an internal position on the RNA template. Such a mechanism of protein-primed nucleic acid synthesis from a site internal to the template is without precedent. Other viruses with protein primers are known to prime nucleic acid synthesis from the ends of template strands (Wimmer, 1982), and retroviruses use internal priming, but RNAs serve as the primers (Varmus and Swanstrom, 1982). Our results do not yet offer insight into the mechanism of synthesis of the second (plus) strand of hepadnavirus DNA. Events involved in the priming, elongation, and template transfer of this strand remain to be explored.

Experimental Procedures

Animals and Virus Infection

Beechey ground squirrels were trapped in the wild in Palo Alto, California. Surface antigen in serum was assayed using the heterologous solid-phase radioimmunoassay for HBsAg (Ausria 11; Abbot Laboratories, N. Chicago, Ill.) according to the manufacturer's instructions. Viral DNA in serum was assayed by the dot blot method (Thomas, 1980) following proteinase K digestion (400 μg/ml at 50°C for 4 hr) and phenol extraction. The animal used in this study was infected by subcutaneous inoculation with a 10^{-4} dilution of infectious serum (strain 27; Ganem et al., 1982).

RNA Preparation

RNA was prepared by the guanidinium-cesium chloride method (Ghsin et al., 1974). Freshly excised liver tissue (12 g wet wt.) was immediately frozen at −70°C until use. The frozen tissue was homogenized with a

polytron (40s at setting 7) in 40 ml of the denaturing solution: 6 M guanidinium isothiocyanate, 50 mM Tris-HCl (pH 7.5 at 24°C), 2% Sarkosyl, 1% β-mercaptoethanol, and 1% "antifoam" (Sigma). The homogenate was layered onto five 3.5 ml cushions of 5.7 M CsCl and centrifuged for 16 hr at 30,000 rpm, 20 C, in an SW41 rotor. Each pellet was resuspended in 0.5 ml 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (TE) with heating to 65°C for 5 min. An equal volume of 4 M LiCl was added, and the RNA was precipitated overnight at 4°C. The pellet was resuspended in TE and precipitated with NaAc-EtOH. Poly(A)$^+$ and Poly(A)$^-$ fractions were separated by chromatography on oligo(dT) cellulose (Maniatis et al., 1982).

RNA Analysis by Gel Electrophoresis

RNA was electrophoresed through 1% agarose in the presence of 2.2 M formaldehyde and was transferred to nitrocellulose (Maniatis et al., 1982). Preparation of the genomic molecular hybridization probes was as previously described (Weiser et al., 1983). These probes were labeled to a specific activity of 0.5–2 × 10^9 Cerenkov cpm/μg. The Bal I-Pst I probe was generated by annealing the M13 sequencing primer to phage containing the Bal I-Pst I plus-stranded GSHV fragment cloned into the Sma I and Pst I sites in the mp8 polylinker. Extension of the primer with Klenow and dNTPs generates "hairpin" strands, containing sequences of both polarities. This probe was labeled to a specific activity of 0.5–1 × 10^9 Cerenkov cpm/μg. Hybridization was in 50% formamide at 42°C for 24–48 hr. Filters were exposed for 2–24 hr to Kodak XAR-5 film, with an intensifying screen, at −70°C. GSHV RNA sizes were estimated by comparison with the positions of the ethidium-bromide-stained ribosomal RNA bands in the lanes containing poly(A)$^-$ RNA.

S1 Mapping with Unit-Length GSHV DNA Fragments

Unit-length fragments of GSHV were prepared from a plasmid, pBA27.2, that contains two tandem copies of the genome cloned into the Eco RI site of a derivative of pBR322. The plasmid was digested with restriction enzymes Apa I, Pvu II, or Xba I, which cleave once in each GSHV monomer. The unit-length GSHV fragments were purified by agarose gel electrophoresis, chromatography on NACS-52 resin (BRL) according to the specifications of the manufacturer, phenol extraction, and EtOH precipitation. (Cleavage of pBA27.2 with Eco RI yields two unit-length GSHV DNA fragments and a fragment containing only pBR322 sequences; in this case gel purification of the GSHV fragments was omitted.) S1 mapping was performed as described by Favaloro et al. (1980). Aliquots containing 2–10 ng of the purified fragments (20 ng of the Eco RI fragments) were added to 25 μg of yeast tRNA in the presence or absence of 250–500 ng of poly(A)$^+$ RNA from infected liver. The nucleic acids were denatured in 80% formamide at 60°C for 15 min, cooled to 50°C for 3 hr, and digested with 40 U/ml S1 nuclease (PL Biochemicals) at 37°C for 30 min. The protected products were precipitated with isopropanol, electrophoresed through 1% alkaline agarose gels, transferred to nitrocellulose (Maniatis et al., 1982), and hybridized to ^{32}P-labeled probes. Autoradiography was for 15–100 hr. The sizes of the protected products were estimated by comparison with 3′ end-labeled λ Hind III fragments. Dot blots (Thomas, 1980) containing known quantities of the pBA27.2 plasmid served to quantitate the signal intensity.

S1 Mapping with End-Labeled Probes

Five micrograms of pBA27.2 was digested with the appropriate restriction enzymes, treated with bacterial alkaline phosphatase, and electrophoresed through nondenaturing 4% polyacrylamide gels. The desired bands were visualized with ethidium bromide staining and long wavelength UV light, they were then cut out of the gel, minced, and soaked overnight in 300 μl TE. The gel was removed by brief microfuge centrifugation, and end-labeling was performed in 20–40 μl of the supernatant. Fragments were labeled on overhanging 5′ ends with γ-^{32}P-dATP (3000 Ci/mmol; Amersham) and T4 polynucleotide kinase (PL Biochemicals). Fragments were labeled on recessed 3′ ends with the appropriate α-^{32}P-dNTPs, unlabeled dNTPs, and Klenow fragments (Boehringer Mannheim Biochemicals). Free nucleotides were removed by chromatography on G-50. The labeled probes were extracted with phenol-$CHCl_3$ and were precipitated twice with EtOH. The identity and purity of each labeled fragment was assessed by cleavage with the appropriate restriction enzymes. Contaminating fragments were usually undetectable, or, when present, represented less than 5% of the labeled fragments. Aliquots containing 0.5–2.0 × 10^4 Cerenkov cpm of probe were added to 25 μg yeast tRNA in the presence or absence of 250–500 ng RNA from infected liver and a blocking fragment (purified as above) in 2- to 4-fold molar excess over the probe. The nucleic acids were denatured in 80% formamide (10–20 μl) at 85°C for 5 min and were cooled to the appropriate temperature (see figure legends) for 4.5–20 hr. The hybrids were digested with 40 U/ml S1 at 37°C for 30 min, extracted with phenol-$CHCl_3$, precipitated with EtOH, resuspended in 80% formamide, and denatured at 68°C for 5 min immediately prior to loading onto the gel. About half of each sample was electrophoresed through a polyacrylamide-8 M urea gel. Salt gradient gels contained 0.5–2.5× TBE (89 mM Tris-borate, pH 8.0, at 24°C; 89 mM boric acid; 2 mM EDTA) (Biggin et al., 1983). Autoradiography was for 1–7 days. The sizes of the protected products were estimated in comparison with 5′ end-labeled ϕX174 RF Hae III fragments and with fragments generated by restriction enzyme cleavage of the probes.

Primer Extension

For primer-extension analysis of the 2.3 kb transcripts, we cleaved 3 μg of pBA27.2 with Xba I and treated the resulting fragments with bacterial alkaline phosphatase. Phosphatase was removed by three extractions with phenol-$CHCl_3$, and the fragments were chromatographed on G-50, precipitated with EtOH, resuspended in TE, and end-labeled as above. After chromatography on G-50, the fragments were cleaved with Kpn I, extracted with phenol-$CHCl_3$, precipitated with EtOH, and electrophoresed in a polyacrylamide gel. The approximately 150 nucleotide band was identified by autoradiography and was purified as above. Aliquots containing 1 × 10^4 Cerenkov cpm were added to 1 μg of poly(A)$^+$ RNA from uninfected and from infected liver. The nucleic acids were precipitated with EtOH, resuspended in 80% formamide, denatured at 85°C for 5 min, cooled to 45°C for 10 hr, and precipitated with EtOH. The pellet was resuspended in 14 μl dH_2O, to which was added 0.2 μl RNasin (25 U/μl; Promega Biotec), 2 μl reverse transcriptase (RTase) salts (1 M Tris-HCl, pH 8.3, at 42°C; 0.5 M KCl; 100 mM $MgCl_2$; and 100 mM DTT), 2 μl dNTPs (5 mM each), and 2 μl AMV RTase (10 U/μl; Seiki). The solution was incubated at 42°C for 1 hr, 5 μg yeast tRNA was added, and the products were extracted with phenol-$CHCl_3$, precipitated with EtOH, resuspended in 80% formamide, and denatured at 68°C immediately before loading them onto the gel. About half of the sample was electrophoresed through a 6% polyacrylamide-8 M urea gel.

For primer-extension analysis of the 3.5 kb transcript, 0.015 OD units of a synthetic primer composed of minus-strand nucleotides 167–148 (J. Barnett and C. Craik, UCSF Hormone Research Laboratories) was end-labeled, and free nucleotides were removed by chromatography on Sep-pak (Waters Associates) according to the specifications of the manufacturer. Aliquots containing 5 × 10^4 Cerenkov cpm (approximately 0.001 OD units) of the labeled primer were added to 500 ng of poly(A)$^+$ RNA from uninfected and infected liver. The nucleic acids were heated to 95°C for 2 min, chilled on ice, and then used as templates for primer extension using the reaction conditions described above.

For primer extension on the SP6-generated RNA, the GSHV Pst I-Avr II fragment (nucleotides 936–3250) was cloned into the Xba I site of the Salmonella phage pSP65 cloning vector (Promega Biotec). RNA transcribed in vitro was isolated by phenol-$CHCl_3$ extraction and EtOH precipitation. Aliquots containing 2 × 10^5 Cerenkov cpm of primer were added to approximately 20 ng of RNA, and primer extension was performed as described for the 3.5 kb transcript.

Acknowledgments

We thank Dan Trueba and Fred Beck for their help in obtaining liver biopsies and serum from the animals, and Christoph Seeger for providing the M13 clones used in this analysis. This work was supported by NIH grant AI/CA 18782. G. H. E. is a Medical Scientist Training Program Trainee, award #GM07618. D. G. is a fellow of the John L. and George H. Hartford Foundation, and H. E. V. is an American Cancer Society Research Professor.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby

marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 1, 1985

References

Beasley, R. P., Lin, C.-C., Hwang, L.-Y., and Chien, C.-S. (1981). Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22,707 men in Taiwan. Lancet *2*, 1129–1133.

Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA *80*, 3963–3965.

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491–500.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. *50*, 349–383.

Cattaneo, R., Will, H., and Schaller, H. (1984). Hepatitis B virus transcription in the infected liver. EMBO J. *3*, 2191–2196.

Clare, J., and Farabaugh, P. (1985). Nucleotide sequence of a yeast Ty element: evidence for a novel mechanism of gene expression. Proc. Natl. Acad. Sci. USA *82*, 2829–2833.

Favaloro, J., Treisman, R., and Kamen, R. (1980). Transcription maps of polyoma virus-specific RNA: analysis by 2-dimensional nuclease S1 gel mapping. Meth. Enzymol. *65*, 718–749.

Fitzgerald, M., and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell *24*, 251–260.

Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P., and Charnay, P. (1979). Nucleotide sequence of hepatitis B virus genome (subtype ayw) cloned in E. coli. Nature *281*, 646–650.

Ganem, D. (1982). Persistent infection of humans with hepatitis B virus: mechanisms and consequences. Rev. Infect. Dis. *4*, 1026–1047.

Ganem, D., Weiser, B., Barchuk, A., Brown, R. J., and Varmus, H. E. (1982). Biological characterization of acute infection with ground squirrel hepatitis virus. J. Virol. *44*, 366–373.

Gerlich, W. H., and Robinson, W. S. (1980). Hepatitis B virus contains protein attached to the 5' terminus of its complete strand. Cell *21*, 801–809.

Glisin, V. R., Crkvenjakov, R., and Byus, C. (1974). Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry *13*, 2633–2637.

Guilley, H., Dudley, R. K., Jonard, G., Balazs, E., and Richards, K. E. (1982). Transcription of cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. Cell *30*, 763–773.

Heermann, K. H., Goldmann, U., Schwartz, W., Seyffarth, T., Baumgarten, H., and Gerlich, W. H. (1984). Large surface proteins of hepatitis B virus containing the pre-s sequence. J. Virol. *52*, 396–402.

Honigman, A., Bar-shira, A., Silberberg, H., and Panet, A. (1985). Generation of a uniform 3' end RNA of murine leukemia virus. J. Virol. *53*, 330–334.

Kelley, D. E., Coleclough, C., and Perry, R. P. (1982). Functional significance and evolutionary development of the 5'-terminal regions of immunoglobulin variable-region genes. Cell *29*, 681–689.

Kozak, M. (1984). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature *308*, 241–246.

Landers, T. A., Greenburg, H. B., and Robinson, W. S. (1977). Structure of the hepatitis B virus Dane particle DNA and nature of the endogenous DNA polymerase reaction. J. Virol. *23*, 368–376.

Laub, O., Rall, L. B., Truett, M., Shaul, Y., Standring, D. N., Valenzuela, P., and Rutter, W. J. (1983). Synthesis of hepatitis B surface antigen in mammalian cells: expression of the entire gene and the coding region. J. Virol. *48*, 271–280.

Liu, C.-C., Simonsen, C. C., and Levinson, A. D. (1984). Initiation of translation at internal AUG codons in mammalian cells. Nature *309*, 82–84.

Machida, A., Kishimoto, S., Ohnuma, H., Baba, K., Ito, Y., Miyamoto, H., Funatsu, G., and Oda, K. (1984). A polypeptide containing 55 amino acid residues coded by the pre-S region of hepatitis B virus deoxyribonucleic acid bears the receptor for polymerized human as well as chimpanzee albumins. Gastroenterol *86*, 910–918.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Marion, P. L., Oshiro, L., Regnery, D. C., Scullard, G. H., and Robinson, W. S. (1980). A virus in Beechey ground squirrels which is related to hepatitis B virus of humans. Proc. Natl. Acad. Sci. USA *77*, 2941–2945.

Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., and Kingsman, A. J. (1985). A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon TY1. Nature *313*, 243–246.

Meyers, R. M., Lumelsky, N., Lerman, L. S., and Maniatis, T. (1985). Detection of single base substitutions in total genomic DNA. Nature *313*, 495–498.

Miller, R. H., Marion, P. L., and Robinson, W. S. (1984). Hepatitis B viral DNA–RNA hybrid molecules in particles from infected liver are converted to viral DNA molecules during an endogenous DNA polymerase reaction. Virology *139*, 64–72.

Molnar-Kimber, K. L., Summers, J., Taylor, J. M., and Mason, W. S. (1983). Protein covalently bound to minus strand DNA intermediates of duck hepatitis B virus. J. Virol. *45*, 165–172.

Molnar-Kimber, K. L., Summers, J. W., and Mason, W. S. (1984). Mapping of the cohesive overlap of duck hepatitis B virus DNA and of the site of initiation of reverse transcription. J. Virol. *51*, 161–191.

Persing, D. H., Varmus, H. E., and Ganem, D. (1985). A frameshift mutation in the pre-S region of the human hepatitis B virus genome allows production of surface antigen particles but eliminates binding to polymerized albumin. Proc. Natl. Acad. Sci. USA *82*, 3440–3444.

Proudfoot, N. (1982). The end of the message. Nature *298*, 516–517.

Rall, L. B., Standring, D. N., Laub, O., and Rutter, W. J. (1983). Transcription of hepatitis B virus by RNA polymerase II. Mol. Cell. Biol. *3*, 1766–1773.

Robinson, W. S., Marion, P., Feitelson, M., and Siddiqui, A. (1981). The Hepa-DNA-virus group: hepatitis B and related virus. In Viral Hepatitis 1981, W. Szmuness, J. Alter, and J. Maynard, eds. (Philadelphia: International Symposium Franklin Institute Press), pp. 57–68.

Schwartz, D. E., Tizard, R., and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell *32*, 853–869.

Seeger, C., Ganem, D., and Varmus, H. E. (1984a). Nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus. J. Virol. *51*, 367–375.

Seeger, C., Ganem, D., and Varmus, H. E. (1984b). The cloned genome of ground squirrel hepatitis virus is infectious in the animal. Proc. Natl. Acad. Sci. USA *81*, 5849–5852.

Simonsen, C. C., and Levinson, A. D. (1983). Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol. Cell. Biol. *3*, 2250–2258.

Standring, D. N., Rutter, W. J., Varmus, H. E., and Ganem, D. (1984). Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. J. Virol. *50*, 563–571.

Summers, J., and Mason, W. S. (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. Cell *29*, 403–415.

Summers, J., O'Connell, A., and Millman, I. (1975). Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. Proc. Natl. Acad. Sci. USA *72*, 4597–4601.

Thomas, M., White, R., and Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Natl. Acad. Sci. USA *73*, 2294–2298.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments to nitrocellulose. Proc. Natl. Acad. Sci. USA *77*, 5201–5205.

Tiollais, P., Charnay, P., and Vyas, G. (1981). Biology of hepatitis B virus. Science *213*, 406–411.

Toh, H., Hayashida, H., and Miyata, T. (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature *305*, 827–829.

Valenzuela, P., Quiroga, M., Zaldivar, J., Gray, P., and Rutter, W. J. (1980). The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. In Animal Virus Genetics, B. N. Fields and R. Jaenisch, eds. (New York: Academic Press), pp. 369–512.

Varmus, H. E., and Swanstrom, R. (1982). Replication of retroviruses. In RNA Tumor Viruses, R. Weiss, N. Teich, H. Varmus, and J. Coffin, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Press), pp. 369–512.

Weiser, B., Ganem, D., Seeger, C., and Varmus, H. E. (1983). Closed circular viral DNA and asymmetrical heterogeneous forms in livers from animals infected with ground squirrel hepatitis virus. J. Virol. *48*, 1–9.

Weiss, S. R., Hackett, P. B., Oppermann, H., Ulrich, A., Levintow, L., and Bishop, J. M. (1978). Cell-free translation of avian sarcoma virus RNA: suppression of the *gag* termination codon does not augment synthesis of the joint *gag/pol* product. Cell *15*, 607–615.

Wimmer, E. (1982). Genome-linked proteins of viruses. Cell *28*, 199–201.

Note Added in Proof

Using the duck hepatitis B virus, Büscher et al. (Cell *40*, 717–724) have recently reported findings similar to ours.

Chapter 2

5'-terminal sequences influence the segregation of Ground Squirrel Hepatitis Virus RNAs into polyribosomes and viral core particles

JOURNAL OF VIROLOGY, Jan. 1987, p. 35–41
0022-538X/87/010035-07$02.00/0

Vol. 61, No. 1

5′-Terminal Sequences Influence the Segregation of Ground Squirrel Hepatitis Virus RNAs into Polyribosomes and Viral Core Particles

GREG H. ENDERS,[1] DON GANEM,[2,3] AND HAROLD E. VARMUS[1,2]*

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

Received 15 July 1986/Accepted 26 September 1986

To determine which of the major ground squirrel hepatitis virus RNAs serve as mRNAs and which serve as templates for reverse transcription of the genome, we analyzed the subcellular distribution of these RNAs in livers of infected ground squirrels. Both major classes of viral RNA, the 2.3- and 3.5-kilobase (kb) classes, are unspliced, are polyadenylated at a common position, and display heterogeneous 5′ ends that can encode proteins with different amino termini (G. H. Enders, D. Ganem, and H. Varmus, Cell 42:297–308, 1985). Both of the 2.3-kb RNAs, which encode surface antigens, appear to be predominantly associated with polyribosomes. Of the three 3.5-kb RNAs, the two longer, which can encode a protein initiated from the first methionine codon in the core antigen gene, appear to be predominantly associated with polyribosomes, and a minority of the shortest 3.5-kb RNAs, which can encode a protein initiated from the second methionine in the core antigen gene, appears to be associated with polyribosomes. This last RNA is instead found predominantly within viral core particles, consistent with evidence that indirectly implicates it in two steps of viral DNA synthesis (C. Seeger, D. Ganem, and H. E. Varmus, Science 232:477–484, 1986). None of the other viral RNAs is detectably packaged into cores. These findings provide independent evidence that the shortest 3.5-kb RNA is the template for synthesis of the viral genome and reveal a novel selectivity in viral RNA packaging.

The hepatitis B viruses (hepadnaviruses) are the smallest DNA viruses known to infect vertebrates. Viral replication occurs primarily in the liver and is associated with a range of pathology that encompasses the late development of hepatocellular carcinoma (2, 6). The hepadnaviruses synthesize their genomes from RNA templates, using a strategy that differs from the canonical reverse transcription scheme employed by retroviruses and several eucaryotic transposons (21, 23, 25). In this paper we present evidence distinguishing hepadnaviral RNAs that play a role in genome synthesis from those that function solely as mRNAs.

We recently determined the primary structures of the major viral RNAs present in the livers of ground squirrels infected by ground squirrel hepatitis virus (GSHV; 4). Both major classes of viral RNA, the 2.3- and 3.5-kilobase (kb) classes, are plus stranded, unspliced, and polyadenylated at a common position (Fig. 1). The 2.3-kb RNAs are subgenomic in size and display heterogeneous 5′ ends that bracket a methionine codon in the presurface reading frame. This apparently facilitates the expression of two surface antigens (sAgs), designated pre-S2 and S, that differ at their amino termini (4, 18, 22). The 3.5-kb RNAs are slightly longer than genome length and, thus, terminally redundant. Their 5′ ends are clustered at nucleotides −25, −16, and +6, relative to the first methionine codon in the core antigen (cAg) open reading frame. These positions suggest that the 3.5-kb RNAs may serve as mRNAs for cAgs encoded from the first (precore) and second (core) methionine codons in the cAg gene.

Because they include all of the information in viral DNA, the 3.5-kb RNAs are obvious candidates to serve also as templates for synthesis of the viral genome. As noted above, the multiple 5′ ends of the 3.5-kb RNAs confer on these RNAs the potential to direct the synthesis of different viral proteins. Does synthesis of these ends also yield RNAs with different potentials as replicative templates? Two pieces of evidence have emerged that suggest that the shortest 3.5-kb RNA may be the sole functional replicative template. (i) The 3′ end of virion minus-strand DNA is unique and maps to nucleotide +6 (21). This fact suggests that reverse transcription may proceed to the end of the template by using the shortest 3.5-kb RNA. (ii) A capped oligoribonucleotide is attached to the 5′ end of virion plus-strand DNA at position −200 and presumably primes synthesis of the DNA at that site (11, 21). The sequence of this RNA primer corresponds to the 5′-terminal 17 nucleotides of the shortest 3.5-kb RNA. This correspondence suggests that the primer may have been transported to the site of initiation of plus-strand DNA synthesis after cleavage from the 5′ end of the shortest 3.5-kb RNA (11, 21).

These observations, then, suggest that the shortest 3.5-kb RNA is used for synthesis of the DNA structures found in mature virions. It is, however, possible that all three 3.5-kb RNAs are packaged into viral core particles but that only the shortest can be utilized by the replicative machinery or that various DNA forms are synthesized in intracellular core particles with RNAs of different structure but that only cores with the described DNA structure exit the cell.

We sought to clarify these issues by directly examining the disposition of RNAs in cytoplasmic lysates of GSHV-infected hepatocytes to determine which GSHV RNA species are associated with polyribosomes and which are found in core particles (the site of reverse transcription [13, 23]). We present evidence here that most of the 2.3-kb and the two longer 3.5-kb RNAs and a minor fraction of the shortest 3.5-kb RNAs appear to be associated with polyribosomes. This finding provides further evidence that both precore and core proteins are likely to be synthesized in GSHV-infected livers (4). The shortest 3.5-kb RNA is found predominantly within viral core particles. We estimate that the ratio of the

* Corresponding author.

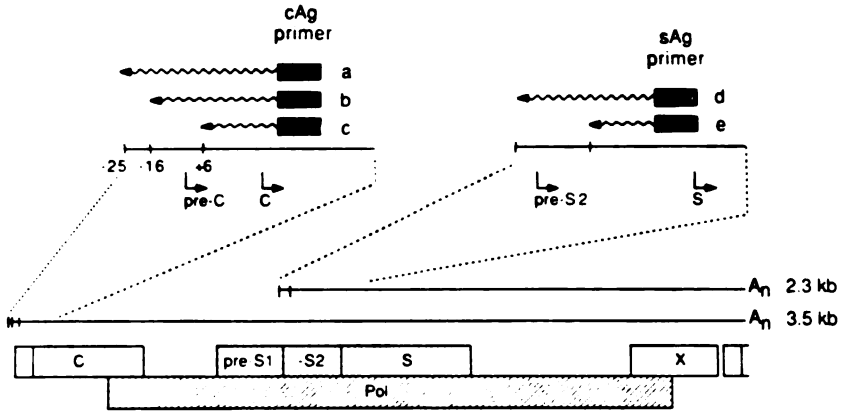


FIG. 1. The major GSHV RNAs and the primers used for primer extension analysis. The two major classes of GSHV RNA are represented by two lines, with short vertical lines marking the positions of the major 5' ends and A_n designating the poly(A) tails. Depicted below them are the viral long open reading frames (20). Depicted above are enlargements of the regions of the RNAs examined by primer extension analysis. The primers (■) and extension products (~~) are not to scale. The cAg primer is composed of minus-strand nucleotides +167 to +148. The sAg primer is composed of minus-strand nucleotides +1643 to +1624. a, b, and c, cAg primer extension products 192, 183, and 162 (±3) nucleotides long, respectively, that terminate at the three major 5' ends of the 3.5-kb RNAs, at nucleotides −25, −16, and +6, respectively (4); d and e, sAg primer extension products 158 and 119 (±3) nucleotides long, respectively, that terminate at the two major 5' ends of the 2.3-kb RNAs (4); ⌊→, translation initiation sites of the putative precore (pre-C) and core (C) proteins (at +1 and +91, respectively) and the pre-S2 and major sAgs (S).

shortest 3.5-kb RNA to all other known viral RNAs is at least 100 to 1 in cores. This finding provides independent evidence that the shortest GSHV 3.5-kb RNA is the template for synthesis of the viral genome and reveals a remarkable selectivity in packaging of viral RNA, in which the extreme 5' end is a critical feature.

MATERIALS AND METHODS

Animals. Beechey ground squirrels were trapped in the wild in Palo Alto, Calif. sAg in serum was assayed by the heterologous solid-phase radioimmunoassay for human hepatitis B sAg (Ausria 11; Abbott Laboratories, North Chicago, Ill.) according to the specifications of the manufacturer. Viral DNA in serum was assayed by the dot-blot method (24) after proteinase K digestion (400 μg/ml at 50°C for 4 h) and phenol extraction. The animal used for this study tested positive for sAg in serum when trapped and contained both sAg and viral DNA in its serum when sacrificed, 4 months later. The second animal, from which data not shown here were derived, was infected by direct liver inoculation with a cloned GSHV genome bearing a G-to-C transversion at nucleotide +18 (21). The animal was sacrificed after remaining positive for sAg in serum for 4 months.

Subcellular fractionation. A freshly excised liver was divided into three portions. One portion was immediately frozen at −70°C and used later for RNA isolation by the guanidium isothiocyanate procedure as described previously (4). A second portion was homogenized at 4°C by 2 strokes of a motor-driven Teflon pestle in 2 ml of a solution containing 10 mM Tris hydrochloride (pH 8.0), 0.15 M KCl, 1.5 mM $MgCl_2$, and 2 mM vanadate complexes per g and 100 μg of cycloheximide, included to inhibit ribosomal runoff, per ml. Nuclei and cell debris were removed by centrifugation at 10,600 × *g* at 4°C for 10 min. The supernatant was brought to 1% Nonidet P-40 by the addition of 20% Nonidet P-40, generating the solution referred to as the cytoplasmic lysate.

The sucrose gradient sedimentation protocol was adopted, in modified form, from Katze et al. (10). Lysate (1.3 ml) was loaded onto linear 10 to 50% sucrose gradients containing 20 mM Tris hydrochloride (pH 8.0), 50 mM KCl, and either 5 mM magnesium acetate or 10 mM EDTA (all solutions except the Tris hydrochloride were treated with diethylpyrocarbonate). Centrifugation was carried out at 27,000 rpm (12,400 × *g*) at 4°C for 4 h in an SW41 rotor. Fractions were collected from the top with an Auto Densi-Flo 11-C probe and a monostatic pump (both from Buchler Instruments Div., Nuclear-Chicago Corp., Fort Lee, N.J.). The pellet was suspended in gradient buffer without sucrose. Fractions from the gradient containing Mg^{2+} were brought to 20 mM EDTA by the addition of 0.5 M EDTA, and all fractions were brought to 1% sodium dodecyl sulfate (SDS) by the addition of 20% SDS immediately before the isolation of nucleic acids.

RNA was isolated by extraction with phenol-chloroform-isoamyl alcohol (25:24:1) that had been equilibrated with 0.1 M Tris hydrochloride (pH 8.0). Extractions were repeated until the phenol-water interface cleared, with additions of 10 mM Tris hydrochloride (pH 8.0)–1 mM EDTA as needed to maintain the volume of the aqueous phase. The nucleic acids were then precipitated twice with sodium acetate and ethanol. To degrade RNases in the lysate prior to RNA isolation or to isolate DNA, samples were pretreated with proteinase K (proteolysis removes the protein that is covalently attached to the 5' end of viral minus-strand DNA, allowing retention in the aqueous phase of minus strands and any plus strands that may be extensively hydrogen bonded to them (7, 8). Yeast tRNA (5 μg) was added as a carrier, and the samples were brought to 10 mM vanadate complexes and 200 μg of proteinase K per ml and were incubated at 37°C for 30 min. The samples were then brought to 1% SDS by adding 20% SDS and 500 μg of proteinase K per ml and were incubated at 37°C for 1 h (14).

Equilibrium density centrifugation was performed as described by Feitelson et al. (5) except that the cesium was buffered with 10 mM Tris hydrochloride (pH 7.5), and the samples were centrifuged at 24,500 rpm (10,000 × *g*) at 10°C

for 70 h in an SW41 rotor. Lysate (1.3 ml) was adjusted to 1.19 g/ml with CsCl and loaded onto a 10-ml CsCl step gradient from 1.25 to 1.40 g/ml. Fractions were collected as described above. The pellet was suspended in gradient buffer. The density of each fraction was determined by weighing duplicate 50-μl aliquots (12 fractions were collected; after their densities were determined, the fractions were pooled by twos). The fractions were then dialyzed exhaustively at 4°C against 10 mM Tris hydrochloride (pH 8.0)–0.15 M NaCl, and nucleic acids were isolated as described above.

The third portion of the liver was used for puromycin treatment, as described by Adelman et al. (1), followed by sucrose gradient sedimentation.

Nucleic acid analysis. Primer extensions and agarose gel electrophoresis were performed as described previously (4). DNA dot-blots of extract-derived nucleic acids were performed as described by Thomas (24). RNA was hydrolyzed by incubating each sample, plus a control sample containing 1 ng of GSHV RNA synthesized in vitro with the SP6 polymerase, in 0.2 M NaOH at 65°C for 20 min. For analysis of DNA from the second animal, the nucleic acids were applied to nitrocellulose without prior alkaline hydrolysis and allowed to hybridize with a probe specific for viral minus strands.

RESULTS

Experimental design. Cytoplasmic lysates were prepared from the livers of ground squirrels chronically infected with GSHV and subjected to rate zonal sedimentation in sucrose, nuclease digestion, and equilibrium density centrifugation in CsCl. After each of these procedures, nucleic acids were isolated and tested for the presence of viral RNA and DNA species. Because the major GSHV RNAs differ in the positions of their 5′ ends, they are distinguishable by primer extension analysis (Fig. 1). Extension of a synthetic primer composed of minus-strand nucleotides +167 to +148 (the cAg primer) on RNA from infected liver yielded major products of 192, 183, and 162 (±3) nucleotides (labeled a, b, and c in the figures), corresponding to the three major 5′ ends of the 3.5-kb RNAs (4). Extension of a synthetic primer composed of minus-strand nucleotides +1643 to +1624 (the sAg primer) yielded major products of 158 and 119 (±3) nucleotides (labeled d and e), corresponding to the two major 5′ ends of the sAg RNAs. We have not detected RNAs in infected liver with ends outside these two regions of the genome (4). Furthermore, the use of 5′-proximal primers to identify the RNAs should yield signals that are relatively unaffected by moderate amounts of nonspecific degradation that may occur during fractionation or by degradation of the body of the RNAs that may occur during their use as replicative templates (23). (The cAg primer anneals to sequences just outside of the terminal redundancies in the 3.5-kb RNAs; hence, it anneals only to the 5′ ends of these RNAs.) All the data presented in this paper are derived from experiments performed in parallel on material from a single animal; where noted below, similar experiments were performed on material from a second animal, with qualitatively similar results.

Sucrose gradient sedimentation. As a first step toward distinguishing the RNAs in intracellular viral core particles from those associated with polyribosomes and other ribonucleoprotein particles, we subjected a Nonidet P-40-treated cytoplasmic lysate to sedimentation on linear 10 to 40% sucrose gradients in the presence and absence of

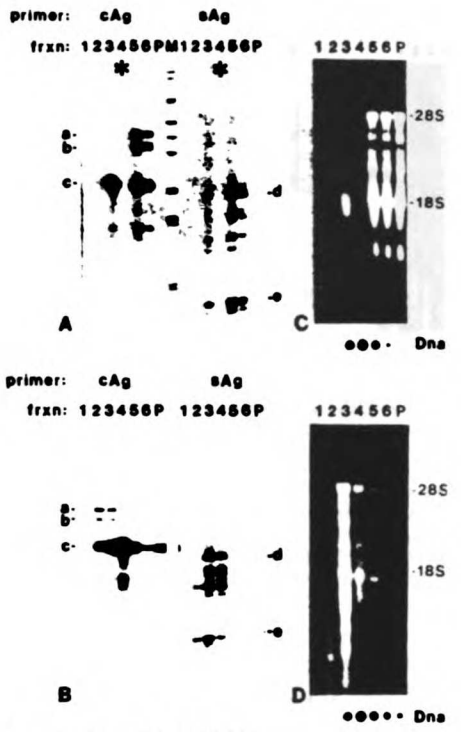


FIG. 2. Sucrose gradient sedimentation of cytoplasmic lysates. A cytoplasmic lysate prepared from a GSHV-infected liver was layered onto linear 10 to 50% sucrose gradients containing either Mg^{2+} (A and C) or EDTA (B and D). Six fractions were collected, beginning at the top of each gradient, plus the pellet (P). Nucleic acids were isolated from each fraction, and equal portions were subjected to three treatments: (i) primer extension analysis with the cAg and sAg primers (Fig. 1), followed by electrophoresis of the products through a 6% polyacrylamide-urea gel (A and B); (ii) electrophoresis in an agarose gel containing ethidium bromide (C and D, top); or (iii) alkaline hydrolysis, followed by dotting onto nitrocellulose and hybridization with a probe containing the GSHV DNA genome cloned in a derivative of pBR322 (C and D, bottom [DNA]). The nucleic acids isolated for primer extension analysis and agarose gel electrophoresis from fraction 4 of the gradient containing Mg^{2+} (asterisks) were lost. However, nucleic acids were also isolated, with prior proteinase K treatment, for the DNA dot-blot analysis presented in C, and a small portion of this sample was included in the ethidium-stained gel. M, Msp fragments of pBR322 (molecular weights, 110, 123, 147, 160, 180, 190, 201, 217, 238, and 242); a, b, c, d, and e, primer extension products diagrammed in Fig. 1. Some sAg primer extension products likely resulted from the primer annealing to fragments of 3.5-kb RNAs.

EDTA. EDTA dissociates polyribosomes, whereas GSHV core particles remain intact (13). We divided each gradient into six fractions, resuspended the material that pelleted during sedimentation, and isolated nucleic acids by extraction with phenol and precipitation with ethanol. To confirm that EDTA had achieved the desired disruption of polyribosomes, we subjected nucleic acid from each fraction to electrophoresis in an agarose gel in the presence of formaldehyde and stained the gel with ethidium bromide. The results demonstrate the expected shift of 28S and 18S rRNAs toward the top of the gradient containing EDTA (Fig. 2D), relative to their positions in the gradient containing Mg^{2+}

(Fig. 2C). To confirm our expectation that EDTA would not affect the sedimentation of viral core particles in this experiment, we examined fractions for viral DNA (see Materials and Methods). RNA in these samples was hydrolyzed with base, and the DNA was applied in dots to a nitrocellulose filter. The filter was then incubated with a radioactive probe containing the GSHV genome cloned in a derivative of pBR322. The results demonstrate that, in the presence (Fig. 2D, bottom) or absence (Fig. 2C, bottom) of EDTA, viral DNA sedimented to the middle of the gradients (fractions 3, 4, and 5), the region expected from previous work to contain viral cores (23; unpublished results). Identical distributions of DNA were seen in comparable gradients prepared from the second animal (data not shown).

We then subjected nucleic acids isolated from each gradient fraction to primer extension with the cAg and sAg primers. All the viral RNAs appeared to cosediment in the gradient containing Mg^{2+} (Fig. 2A). (The nucleic acids prepared for this primer extension analysis from fraction 4 were lost; however, we had also prepared nucleic acids from this fraction under slightly different conditions [see Materials and Methods], and primer extensions on this material yielded the amounts of product expected for the peak fraction [data not shown].) In the presence of EDTA, however, most of two longer 3.5-kb RNAs were shifted to the top of the gradient (Fig. 2B, fractions 2 and 3, bands a and b), whereas most of the shortest 3.5-kb RNAs again sedimented to the middle of the gradient (Fig. 2B, band c). Most of the sAg RNAs, like the two longer 3.5-kb RNAs, were shifted toward the top of the gradient (Fig. 2B, bands d and e). The peak of sAg RNAs appeared in fraction 3, as opposed to fraction 2 for the longer 3.5-kb RNAs, for reasons that are unclear. Results similar to those depicted in Fig. 2A and B were also obtained with material from the second animal (data not shown).

Taken together, these data suggested that the RNAs shifted toward the top of the gradient containing EDTA, i.e., most of the 2.3-kb and the two longer 3.5-kb RNAs but only a small fraction of the shortest 3.5-kb RNA, were released from structures, presumably largely polyribosomes, that sediment farther in the gradient containing Mg^{2+}. To be certain that the differences in recovery of the extension products using the cAg primer were not due to the variable copurification of a diffusible inhibitor in the different samples, we mixed samples from fractions 1 and 4 of the gradient containing EDTA (Fig. 2B) and repeated the primer extension reaction. The recovery of products from this reaction was not inhibited by the sample from fraction 1, thereby ruling out this potential artifact (data not shown). We then serially diluted the products from fractions 2 and 4 (Fig. 2B) prior to electrophoresis to quantify the ratios of the 5′ ends in these samples. The results demonstrate that the shortest 3.5-kb RNA is at least 15-fold enriched relative to the two longer 3.5-kb RNAs in fraction 4 relative to fraction 2 (data not shown).

RNase sensitivity. From these data, we predicted that the sAg and the two longer 3.5-kb RNAs would be preferentially sensitive to digestion by endogenous and exogenous RNases, whereas the shortest 3.5-kb RNA would be relatively resistant. Strong indirect evidence has been obtained previously that RNA in hepadnavirus core particles is resistant to digestion with 0.1 to 0.5 μg of RNase A per ml (13, 23). We therefore isolated nucleic acids from the total lysate with and without prior incubation with proteinase K and exposure to increasing amounts of RNase A. Nucleic acids isolated from the lysate simply by phenol extraction in the presence

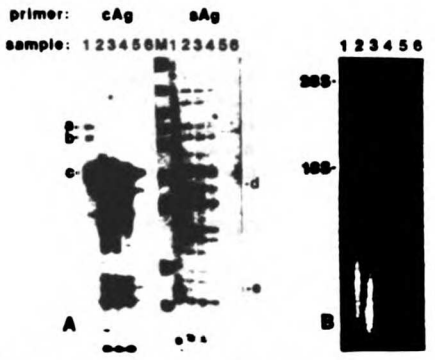


FIG. 3. RNase sensitivity of GSHV RNAs in the cytoplasmic lysate. Six samples taken from the cytoplasmic lysate were used for isolation of nucleic acids by phenol-chloroform extraction. Sample 1, no prior treatment; sample 2, prior proteinase K digestion; samples 3 through 6, prior incubation with 0.05, 0.5, 5, and 25 μg of RNase A per ml, respectively, followed by proteinase K digestion. Isolated nucleic acids were subjected to primer extension analysis with the cAg and sAg primers (A) or were electrophoresed through an agarose gel containing ethidium bromide (B). M, Msp fragments of pBR322; a, b, c, d, and e, primer extension products diagrammed in Fig. 1.

of 1% SDS followed by ethanol precipitation yielded primer extension products corresponding to all of the major viral RNAs (Fig. 3A, sample 1). The rRNAs in this preparation appeared largely intact (Fig. 3B, sample 1). After incubation of the lysate at 37°C for 30 min with proteinase K and then for another 90 min with proteinase K and 1% SDS, recovery of the sAg and the two longer 3.5-kb RNAs was virtually abolished, whereas recovery of the shortest 3.5-kb RNA was actually enhanced ca. threefold (Fig. 3A, sample 2 and other exposures not shown). The loss of the sAg and the two longer 3.5-kb RNAs was due to endogenous RNase activity in the lysate; under these conditions, there was substantial degradation of rRNA (Fig. 3B, sample 2), despite the presence of carrier tRNA and concentrations of vanadate complex RNase inhibitor higher than those included in the sucrose gradients. (The degradation was not due to RNase activity in the proteinase K, because identical treatments of both RNA synthesized in vitro and material from the middle of the sucrose gradient containing Mg^{2+} resulted in undiminished recovery of all the 5′ ends [data not shown].) Note that the primer extension products migrating more rapidly than the major products in Fig. 3 are absent in lane 1 and were not observed after identical treatment of fractions from both the sucrose and cesium gradients (data not shown), indicating that the corresponding RNAs were generated in vitro in this experiment.

These data indicate that the shortest 3.5-kb RNA is preferentially resistant to nuclease attack. We found further that recovery of the shortest 3.5-kb RNA was only slightly diminished by prior incubations with 0.05 and 0.5 μg of RNase per ml (Fig. 3A, samples 3 and 4 and other exposures not shown). Substantial loss of the shortest 3.5-kb RNA was not seen until the concentration of RNase A reached 5 μg/ml, a concentration at which we were no longer able to detect RNA by ethidium bromide staining (Fig. 3A and B, sample 5). To confirm these findings, we treated a portion of fraction 5 from the sucrose gradient containing Mg^{2+} (Fig. 2A) with 0.1 μg of RNase A per ml. This treatment resulted

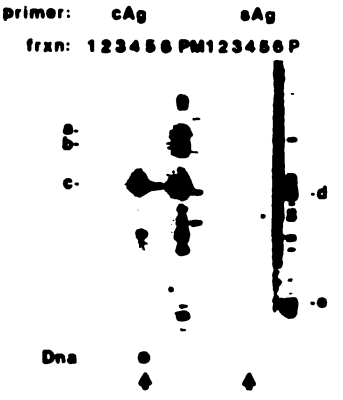


FIG. 4. Equilibrium density sedimentation of the cytoplasmic lysate. The cytoplasmic lysate was loaded onto a cesium chloride step gradient from 1.25 to 1.40 g/ml and subjected to equilibrium density sedimentation. Six fractions were collected, beginning at the top of the gradient, plus the pellet (P). After dialysis, nucleic acids were isolated and subjected to primer extension with the cAg and sAg primers. A portion of the nucleic acids from each fraction was also subjected to alkaline hydrolysis, followed by dotting onto nitrocellulose and hybridization with a probe containing the GSHV DNA genome cloned in a derivative of pBR322 (bottom [DNA]). M, Msp fragments of pBR322; a, b, c, d, and e, primer extension products diagrammed in Fig. 1; arrows, fraction with a density of 1.31 to 1.36 g/ml, where GSHV core particles are expected to band (5).

in the preferential loss of the sAg and the two longer 3.5-kb RNAs (data not shown).

Equilibrium density sedimentation. To provide further evidence that the shortest 3.5-kb RNA is in particles with the properties expected of viral cores, we subjected the cytoplasmic lysate to equilibrium density sedimentation in cesium chloride. GSHV core particles are known to band in cesium gradients at a density of 1.34 g/ml (5), whereas free RNA, with a density of ca. 1.7 g/ml, pellets in such gradients. We divided the gradient into six fractions and suspended the pellet; after dialysis to remove the cesium, we isolated nucleic acids and performed primer extension analyses as outlined above for the sucrose gradients. The results demonstrate that, whereas all the viral RNA species were recovered in the pellet fraction, the only viral RNA in the gradient was the shortest 3.5-kb RNA (Fig. 4). This RNA displayed a sharp peak in signal in the fraction with a density range from 1.31 to 1.36 (Fig. 4, fraction 4 [arrow]). A DNA dot-blot, performed as described for Fig. 2, revealed a coincident peak of viral DNA (Fig. 4, bottom, fraction 4 [arrow on left]).

We sought evidence that the viral DNA in fraction 4 was characteristic of the forms expected for viral core particles. We used primer extension to examine nucleic acids isolated from the gradient after treatment with proteinase K; proteolysis permits recovery of nucleic acids that are extensively hydrogen bonded to protein-linked viral minus-strand DNA (7, 8). Use of the cAg primer with these nucleic acids yielded a sharp peak in fraction 4 of an extension product of 370 nucleotides (data not shown). This band corresponds precisely to that expected from extension of the cAg primer to the 5' end of plus-strand DNA (including its terminal oligoribonucleotide). In addition, prior proteinase K treatment increased recovery of the shortest 3.5-kb RNA from fraction 4 about twofold relative to recovery of this RNA from the pellet. These data demonstrate that the shortest 3.5-kb RNA cosediments on cesium density gradients with viral core particles and suggest that some of this RNA may be hydrogen bonded to protein-linked viral minus strands. Equilibrium density sedimentation of a lysate from the second animal yielded similar results.

DISCUSSION

Evidence for selective packaging. We have presented evidence that, of the major viral RNAs present in the livers of GSHV-infected ground squirrels, only the shortest 3.5-kb RNA is efficiently packaged into viral core particles. Thus, viral RNAs with only 21 additional or 1,480 fewer nucleotides at their 5' termini are not detectably packaged. These conclusions were derived by determining which RNAs were associated with viral core particles by three independent criteria. We also examined the ability of the viral RNAs to sediment through sucrose after treatment with puromycin under conditions known to dissociate polyribosomes (1). We found that the sedimentation of the shortest 3.5-kb RNA was preferentially resistant to such treatment (data not shown).

Because the major GSHV RNAs differ in their subcellular localizations, the relative recovery of these RNAs can differ among isolation procedures. For example, hepadnaviral cores are known to be somewhat unstable in cesium (16), and we saw an overall decrease in yield of shortest 3.5-kb RNAs from the cesium gradients relative to the yields achieved by other isolation procedures (compare Fig. 4 to Fig. 2 and 3; data not shown). The relative yield of RNA from fraction 4 and from the pellet from the cesium gradient (Fig. 4), therefore, probably underestimates the ratio of RNA in cores to RNA associated with other structures. Conversely, disruption of polyribosomes with puromycin reduced the overall recovery of polyribosome-associated RNAs (data not shown). We found, however, that the ratios of the RNA ends recovered directly from the lysate or from the sucrose gradients were similar to the ratios of ends in RNA recovered by the guanidinium isothiocyanate procedure from another portion of the same liver (data not shown). This finding indicates that these latter procedures did not result in significant preferential loss of RNAs.

To be able to assess the relative abundances of the 2.3- and 3.5-kb RNAs, we compared the efficiency of extension of the cAg and sAg primers by using an RNA preparation with which we had independently established the ratio of these two RNA classes (4). We found that extension of the two primers was about equally efficient (data not shown).

These findings allow us to make some quantitative statements about the approximate relative subcellular distribution of the major viral RNAs. At steady state, most of the sAg and the two longer 3.5-kb RNAs are associated with polyribosomes. In contrast, most of the shortest 3.5-kb RNA, which is more abundant than either of the former sets of RNAs, is in core particles. We could not detect bands corresponding to sAg or the two longer 3.5-kb RNAs in primer extensions from the fraction of the cesium gradient containing cores, even with overexposures of the gel presented in Fig. 4. We estimate, therefore, that the ratio of the shortest 3.5-kb RNA to all other known viral RNAs is at least 100 to 1 in cores.

Two earlier pieces of evidence had indirectly implicated the shortest 3.5-kb RNA as the replicative template. (i) The 3' end of virion minus-strand DNA is unique, and it is the

precise complement (±3 nucleotides) of the 5' end of the shortest 3.5-kb RNA (21). (ii) The sequence of the oligoribonucleotide attached to the 5' end of virion plus-strand DNA matches, within a couple bases of the ends, that of the 5'-terminal 17 nucleotides of the shortest 3.5-kb RNA (21). A similar correspondence in sequence has been detected between an oligoribonucleotide attached to the duck hepatitis B virus (DHBV) plus-strand DNA and the 5' end of a greater-than-genome-length DHBV RNA (11). The data presented in this paper provide independent evidence that the GSHV 3.5-kb RNA with its 5' end at +6 is the sole template employed in synthesis of the viral genome. The homogeneity of the relevant features of the virion DNA can thus be explained by selective RNA packaging, without invoking selection at other steps of viral maturation.

We anticipated that our primer extension assays might reveal novel RNA 5' ends generated by cleavage of the packaged RNAs. This could occur if the plus-strand primer was cleaved from viral RNA prior to degradation of 3' sequences. In addition, synthesis of minus-strand DNA apparently begins at nucleotide +15, which is present at both ends of the template (21). If initiation occurs at the 5' end of the template, it could be accompanied by RNase H degradation of the extreme 5' terminus of the RNA. We did not observe such cleavages, which suggests that, if they occur, the resultant RNAs must be short-lived.

Our results do not rule out processive degradation of the RNAs from a site 3' of the cAg primer. Consistent with this possibility is our observation that the 5' end of the shortest 3.5-kb RNA is enriched at least twofold, relative to the other RNA ends, in the nonpolyadenylated fraction of liver RNA (data not shown). Our finding that prior proteinase K treatment of the lysate (Fig. 3) or isolated core particles (data not shown) preferentially increased recovery of the 5' end of the shortest 3.5-kb RNA two- to threefold, relative to that achieved by SDS and phenol alone, suggests that the majority of the packaged RNAs may be associated with protein-linked viral minus strands and hence undergoing reverse transcription.

Messenger function. Although most of the shortest 3.5-kb RNAs are packaged into core particles, the data suggest that a fraction of these RNAs may be associated with polyribosomes (compare fraction 2 in Fig. 2A and B). In contrast, most of the two longer 3.5-kb RNAs were associated with polyribosomes, as judged by their release from rapidly sedimenting structures upon treatment with EDTA (Fig. 2) and puromycin (data not shown). Furthermore, the exclusion of the two longer 3.5-kb RNAs from core particles suggests that their function is confined to translation. If the first in-frame methionine codon in each of the three 3.5-kb RNAs is used for translation initiation, then both precore and core proteins are presumably synthesized in the infected liver.

The amino termini of the cAgs produced in the infected liver have not been identified for any of the hepadnaviruses. The homologous human hepatitis B virus (HBV) precore protein contains a signal peptide that can be cleaved after synthesis of the protein in vitro (P. Garcia, J.-H. Ou, W. J. Rutter, and P. Walter, personal communication), and constructs expressing this protein in transfected cells result in the secretion of immunoreactive products of the cAg frame (17, 19). The homologous HBV core protein synthesized under the same in vitro conditions is not detectably cleaved (Garcia et al., personal communication), and its immunoreactive products remain largely within the transfected cell (17, 19). These data indicate that the precore region confers distinct properties on the resultant protein. We have initiated genetic studies to address whether both proteins are required for viral infectivity.

Thus far, RNAs containing the precore region at their 5' ends have been detected only in livers infected by GSHV (4) and woodchuck hepatitis virus (15; unpublished results), whereas transcripts analogous to the shortest GSHV 3.5-kb RNAs have been detected in livers infected by woodchuck hepatitis virus (15; unpublished results), DHBV (3), and HBV (H. Will, personal communication). However, given that the precore reading frame is preserved in all infectious hepadnaviral genomes that have been sequenced and is well conserved among mammalian isolates (17), it seems likely that the precore region is expressed via mRNAs that have yet to be detected in DHBV- and HBV-infected livers. Such mRNAs, like their GSHV counterparts, may be more susceptible than the detected RNAs to degradation by endogenous RNases and, thus, may be preferentially lost during isolation.

Packaging models. The GSHV sAg RNAs, like retroviral *env* mRNAs, appear not to be efficiently packaged into core particles, presumably because they both lack the *cis*-acting sequences required for packaging (12, 26). This explanation cannot, however, account for the absence of the longer 3.5-kb RNAs from core particles, because these RNAs possess all of the primary sequences present in the packaged RNA. Several potential models to explain this observation can be envisioned. The models we offer invoke (i) recognition of a 5'-terminal sequence in conjunction with a cap structure, (ii) recognition of a 5'-proximal secondary structure, (iii) *cis*-packaging, and (iv) translational interference.

(i) Sequence plus cap. The packaging apparatus may recognize the precise sequence of the 5' end of the shortest 3.5-kb RNA, perhaps in conjunction with the cap structure putatively present on this RNA. The apparatus performing this recognition could be that which chooses the site of initiation of reverse transcription. In this model, formation of the initiation complex for minus-strand DNA synthesis dictates packaging and can occur only on the 5' end of the shortest 3.5-kb RNA. Polymerization on the 5' end of the template would then yield an AT-rich deoxyribonucleotide that could be transferred to the 3' end of the template to allow completion of minus-strand DNA synthesis (4, 21). Alternatively, the recognition apparatus could be one that also recognizes the oligoribonucleotide primer for cleavage or transfer of the primer to the site of initiation of plus-strand DNA synthesis (11, 21).

(ii) Secondary structure. The additional bases found on the longer 3.5-kb RNAs may alter the secondary structure of the 5' end of the RNA in a manner that prevents recognition of a packaging signal by *trans*-acting factors. We have examined predicted secondary structures for these 5' ends but can detect no obvious basis for such an effect.

(iii) *Cis* packaging. In the *cis* packaging model, each RNA that is packaged is recognized by a protein translated from the same molecule. The precore protein would be incapable of initiating the packaging of its mRNA, perhaps as a result of translocation of the protein into the endoplasmic reticulum. The core protein, synthesized only from the shortest 3.5-kb RNA, would be able to initiate packaging, perhaps by virtue of remaining cytoplasmic. Not every translation event would result in packaging of the RNA, allowing a number of core proteins to be made from a single RNA prior to packaging of the RNA into a core particle. In a variation of this model, a core-pol fusion protein, putatively generated by ribosomal frameshifting near the carboxy

terminus of core (4, 9), would initiate the packaging of its mRNA.

(iv) **Translational interference.** The 5' ends of the longer 3.5-kb RNAs may be bound by translation machinery in a manner that renders a packaging signal in the precore region inaccessible to packaging factors. Similarly, translation of the precore sequence may target the RNAs to a location in the cell where packaging does not occur efficiently.

These models have different implications for the potential of hepadnaviruses to package cellular RNAs and to support the replication of defective genomes. Discrimination between the models will likely require analysis of the behavior of viral mutants in tissue culture cells that permit normal packaging; partially defective phenotypes could be then examined biochemically.

ACKNOWLEDGMENTS

We thank Dan Trueba for help with the animal work, Robin Colgrove for RNA secondary-structure predictions, Peter Walter for advice on characterization of polyribosome-associated RNA, Titia de Lange and David Persing for comments on the manuscript, and members of the Rutter and Walter laboratories for communication of unpublished results.

This work was supported by Public Health Service grants from the National Institutes of Health. G.H.E. is a Medical Scientist Training Program trainee, award GM07618; D.G. is a fellow of the John L. and George H. Hartford Foundation; and H.E.V. is an American Cancer Society Research Professor.

LITERATURE CITED

1. **Adelman, M. R., D. D. Sabatini, and G. Blobel.** 1973. Ribosome-membrane interaction. J. Cell. Biol. **56**:206–229.
2. **Beasley, R. P., C.-C. Lin, L.-Y. Hwang, and C.-S. Chien.** 1981. Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22,707 men in Taiwan. Lancet **ii**:1129–1133.
3. **Buscher, M., W. Reiser, H. Will, and H. Schaller.** 1985. Transcripts and the putative RNA pregenome of duck hepatitis B virus: implications for reverse transcription. Cell **40**:717–724.
4. **Enders, G. H., D. Ganem, and H. Varmus.** 1985. Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. Cell **42**:297–308.
5. **Feitelson, M. A., P. L. Marion, and W. S. Robinson.** 1982. Core particles of hepatitis B virus and ground squirrel hepatitis B virus. I. Relationship between hepatitis B core antigen- and ground squirrel hepatitis core antigen-associated polypeptides by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and tryptic peptide mapping. J. Virol. **43**:687–696.
6. **Ganem, D.** 1982. Persistent infection of humans with hepatitis B virus: mechanisms and consequences. Rev. Infect. Dis. **4**:1026–1047.
7. **Ganem, D., L. Greenbaum, and H. E. Varmus.** 1982. Virion DNA of ground squirrel hepatitis virus: structural analysis and molecular cloning. J. Virol. **44**:374–383.
8. **Gerlich, W. H., and W. S. Robinson.** 1980. Hepatitis B virus contains protein attached to the 5' terminus of its complete strand. Cell **21**:801–809.
9. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230**:1237–1242.
10. **Katze, M. G., Y.-T. Chen, and R. M. Krug.** 1984. Nuclear-cytoplasmic transport and VA1-independent translation of influenza viral messenger RNAs in late adenovirus-infected cells. Cell **37**:483–490.
11. **Lien, J.-M., C. E. Aldrich, and W. S. Mason.** 1986. Evidence that a capped oligoribonucleotide is the primer for duck hepatitis B virus plus-strand DNA synthesis. J. Virol. **57**:229–236.
12. **Mann, R., and D. Baltimore.** 1985. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs. J. Virol. **54**:401–407.
13. **Miller, R. H., P. L. Marion, and W. S. Robinson.** 1984. Hepatitis B viral DNA-RNA hybrid molecules in particles from infected liver are converted to viral DNA molecules during an endogenous DNA polymerase reaction. Virology **139**:64–72.
14. **Miller, R. H., C.-T. Tran, and W. S. Robinson.** 1984. Hepatitis B virus particles from plasma and liver contain viral DNA-RNA hybrid molecules. Virology **139**:53–63.
15. **Moroy, T., J. Etiemble, C. Trepo, P. Tiollais, and M.-A. Buendia.** 1985. Transcription of woodchuck hepatitis virus in the chronically infected liver. EMBO J. **4**:1507–1514.
16. **Ohori, H., M. Tamaki, S. Onodera, E. Yamada, and N. Ishida.** 1980. Antigenic conversion from HBcAg to HBeAg by degradation of hepatitis B core particles. Intervirology **13**:74–82.
17. **Ou, J.-H., O. Laub, and W. J. Rutter.** 1986. Hepatitis B gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. Proc. Natl. Acad. Sci. USA **83**:1578–1582.
18. **Persing, D. H., H. E. Varmus, and D. Ganem.** 1986. Antibodies to pre-S and X determinants arise during natural infection with ground squirrel hepatitis virus. J. Virol. **60**:177–184.
19. **Roossinck, M. J., S. Jameel, S. H. Loukin, and A. Siddiqui.** 1986. Expression of hepatitis B viral core region in mammalian cells. Mol. Cell. Biol. **6**:1393–1400.
20. **Seeger, C., D. Ganem, and H. E. Varmus.** 1984. Nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus. J. Virol. **51**:367–375.
21. **Seeger, C., D. Ganem, and H. E. Varmus.** 1986. Biochemical and genetic evidence for the hepatitis B virus replication strategy. Science **232**:477–484.
22. **Standring, D. N., W. J. Rutter, H. E. Varmus, and D. Ganem.** 1984. Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. J. Virol. **50**:563–571.
23. **Summers, J., and W. S. Mason.** 1982. Replication of the genome of a hepatitis B virus by reverse transcription of an RNA intermediate. Cell **29**:403–415.
24. **Thomas, P. S.** 1980. Hybridization of denatured RNA and DNA fragments to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
25. **Varmus, H., and R. Swanstrom.** 1985. Replication of retroviruses. p. 75–134. *In* R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (ed.), The molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Watanabe, S., and H. Temin.** 1982. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5' long terminal repeat and the start of the *gag* gene. Proc. Natl. Acad. Sci. USA **79**:5986–5990.

Chapter 3

GSHV DNA synthesis: implications for virion morphogenesis

It has been known for a decade that the nucleocapsid core of HBV contains a DNA polymerase: when virions are incubated with radiolabeled deoxyribonucleotides, label is incorporated into growing 3' ends of plus strand viral DNA (Kaplan et al., 1973; Landers et al., 1977). In 1982 Summers and Mason isolated viral core particles from the cytoplasm of DHBV-infected hepatocytes and found that these cores could incorporate deoxyribonucleotides into both viral DNA strands (Summers and Mason, 1982). Their observation that synthesis of the minus strand in such cores was resistant to actinomycin D, an agent that inhibits DNA-dependent DNA synthesis, provided the first strong evidence that hepadnaviral minus strands are synthesized from RNA templates. Minus strands as short as thirty nucleotides can be radiolabeled by the endogenous polymerase in DHBV cytoplasmic cores. These data imply that at least a fraction of such cores are competant to carry out most steps of genome synthesis.

We sought to confirm the central finding obtained with DHBV, that minus strand DNA synthesis is resistant to actinomycin D, using cores isolated from the cytoplasm of GSHV-infected liver. In addition, we sought to address whether, in fact, cytoplasmic cores contain most of the viral DNA intermediates observed in the liver and whether cores containing viral DNAs at other stages of synthesis could be isolated from membranous compartments of the liver. This information is needed to begin to construct models of virion morphogenesis.

We isolated cytoplasmic cores from the liver of a GSHV-infected Beechey ground squirrel, using the procedure of Summers and Mason (1982), and subjected them to sedimentation through a 10 to 40% sucrose

gradient. Portions of each fraction were incubated with radiolabeled deoxyribonucleotides in the presence and absence of actinomycin D, and DNA was isolated by Proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation. Isolated DNA was allowed to anneal to M13 clones of viral minus and plus strand sequences immobilized on nitrocellulose filters. The filters were then washed and counted.

We observed (figure 1) a peak of incorporation into both viral DNA strands at the expected position in the gradient, based on the studies of DHBV (Summers and Mason, 1982). As predicted, synthesis of the plus strand but not the minus strand was sensitive to actinomycin D. Cores isolated from two other ground squirrels yielded the same result (data not shown). For reasons that are unclear, we obtained only about 50% inhibition of plus strand DNA synthesis, as judged by this assay, compared to the 80 to 90% inhibition reported for DHBV. The results, however, support the notion that the hepadnaviruses follow a common replication scheme, in which minus strand DNA is synthesized from an RNA template. Other data now make this conclusion beyond doubt (Seeger et al., 1986).

To examine the distribution of viral DNA in these gradients, we isolated unlabeled DNA from each gradient fraction, applied the samples in dots to nitrocellulose filters, and probed for GSHV sequences. The results (not shown) demonstrated single broad peaks of minus and plus strand GSHV DNA, coincident with the peak of endogenous DNA synthesis. Thus, by these criteria, the gradients yield a single major species of cytoplasmic core particle.

The relative signal strengths obtained with plus and minus strand

Figure 1. Minus and plus strand GSHV DNA synthesis in cytoplasmic core particles sedimented through sucrose.

A cytoplasmic extract of a GSHV-infected liver was subjected to sucrose gradient sedimentation as described by Summers and Mason (1982). Fractions from the gradient were incubated with 32P-TTP and unlabeled deoxyribonucleotide triphosphates in the absence (A) and presence (B) of actinomycin D (0.5 mg/ml). Labeled DNA was isolated by Proteinase K digestion, phenol extraction, and ethanol precipitation. The DNA was resuspended in hybridization buffer (Maniatis et al., 1982) and allowed to anneal to M13 clones of GSHV plus and minus strands immobilized on nitrocellulose discs (ca. 3mm in diameter) with shaking at 41 C for 12h. The discs were washed and counted. X---X: incorporation into minus strand DNA (counts annealed to an M13 clone of the GSHV plus strand); O---O: incorporation into plus strand DNA; 1---1: counts annealing to discs bearing wild-type M13.

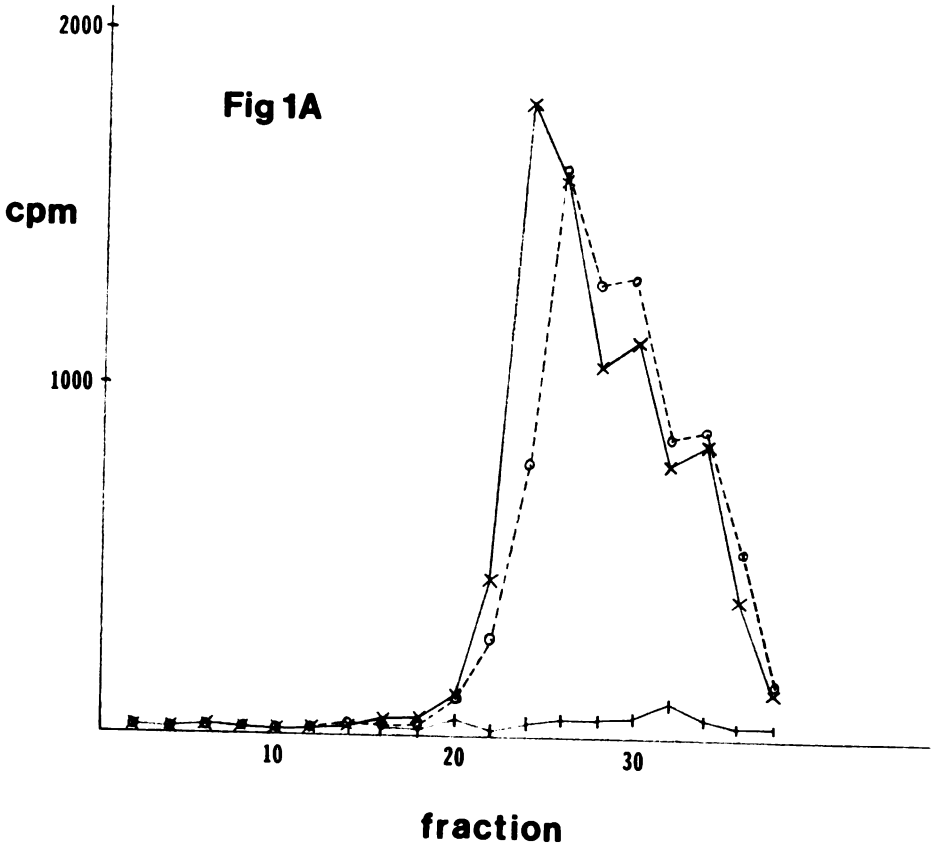


Fig 1A

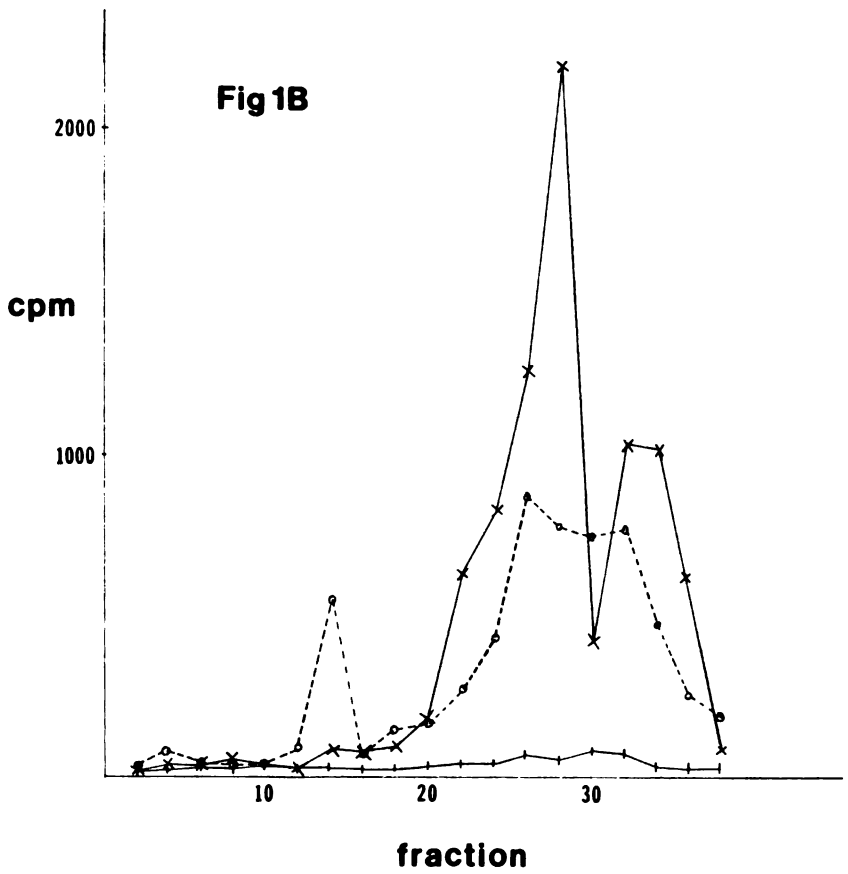


Fig 1B
2000
1000
cpm
10
20
30
fraction

DNA probes suggested a preponderance of viral minus strands in the cores. To more carefully determine the DNA content of cores, in relation to the DNA species seen in the liver as a whole, we prepared DNA from unfractionated liver and from cytoplasmic cores of an infected animal. In addition, we treated a crude cell pellet obtained from this liver with 0.5% NP40. This cell pellet, obtained in the absence of detergent, is expected to contain some undisrupted cells, all the major cell organelles, and membranous vesicles derived from various compartments of the secretory pathway. We reasoned that washing the pellet with NP40 might liberate cores sequestered within the membanous vesicles, while leaving nuclei intact. DNA prepared from whole liver, cytoplasmic cores, and the NP40 wash was subjected to denaturing alkaline gel electrophoresis, transferred to nitrocellulose, and probed for viral plus and minus strand sequences.

The results (Figure 2) support several qualitative statements about the subcellular distribution of viral DNA, although they do not permit precise quantification of the DNAs present in each cellular compartment (see figure legend). The cytoplasmic cores display a large molar excess (roughly 5-fold) of minus strands over plus strands, with the caveat that short DNA strands (eg. plus strand "strong stop" (Seeger et al., 1986); see below) may not have been detected in this assay. The minus strands in cytoplasmic cores are predominantly full-length, whereas the plus strands are mainly shorter that those seen in mature virions. In constrast, viral DNA isolated from the NP40 wash of the cell pellet has the ratio and sizes of minus and plus strand DNAs seen in mature virions (Figure 2; data not shown; Ganem et al., 1982). Consistent with this observation, we find that cores from the NP40 wash

Figure 2. GSHV DNAs isolated from infected liver

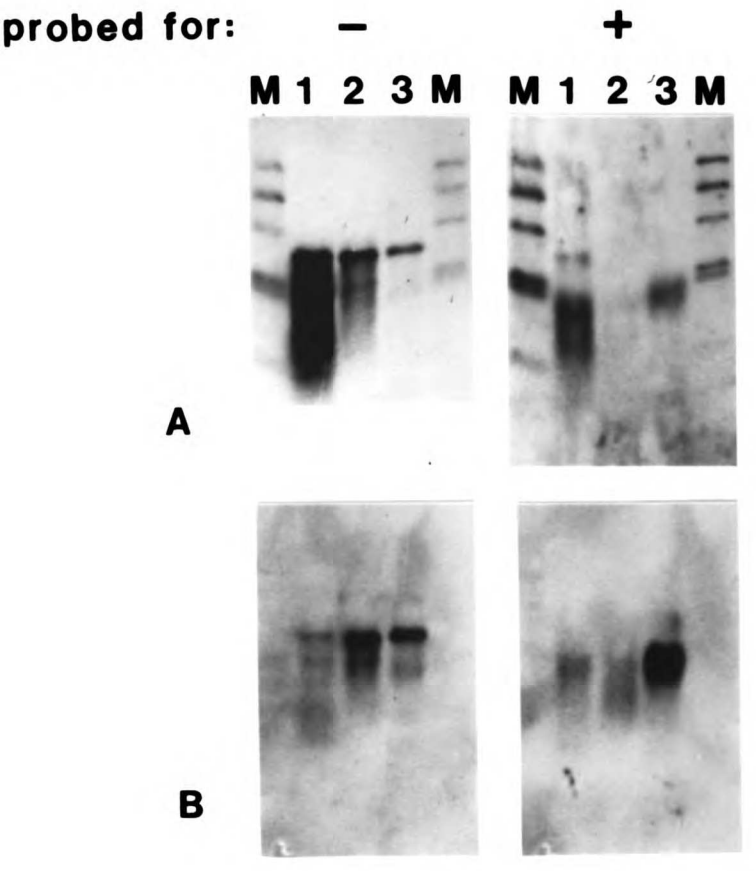
DNA was prepared from whole liver and cytoplasmic cores as described by Summers and Mason (1982). The crude cell pellet obtained during this procedure was treated with 0.5% NP40 and recentrifuged at 10,000 rpm for 20 min at 4 C in a Sorvall HB-4 rotor. DNA was then isolated from the supernatant by SDS-proteinase K digestion as described. The samples were subjected to alkaline gel electrophoresis (Maniatis et al., 1982) and probed for the presence of minus and plus strand sequences as described by Weiser et al. (1983). Lane 1-whole liver DNA. Lane 2-cytoplasmic core DNA. Lane 3-NP40 wash DNA. M-32P-labelled lambda hindlll fragments. A) DNAs from (Lanes 1-3) ca 15, 25, and 50 mg, respectively, of a single liver. B) samples from a second liver (relative proportions of samples not quantitated).

Figure 2

1-whole liver

2-cytoplasmic cores

3-NP40 wash



cosediment with cytoplasmic cores on sucrose gradients but incorporate deoxyribonucleotides exclusively into plus strand DNA; cytoplasmic cores treated with NP40 still synthesize both viral strands (data not shown). Thus, few cytoplasmic cores but most membrane-associated cores have completed plus strand synthesis.

The results in Figure 2 appear to suggest that the liver may contain more short plus strands than are accounted for in the cores. However, this could represent an artefact in preparation of whole liver DNA, due to either degradation of plus strands by nucleases in the liver homogenate or contamination by residual RNA. (No precautions were taken to remove RNA from any of these fractions, although most RNA is likely to be degraded under the purification and electrophoresis conditions employed here (see chapter 2).) Note also that the liver appears to contain short minus strands that are not accounted for in the cores (see also Weiser, et al., 1983). Again, this discrepancy may result from degradation of DNA during its isolation from the crude liver homogenate. (This issue could be resolved by determining whether the sequences of the short minus strands obtained from whole liver represent random fragments of the genome or are polarized towards the site of initiation of reverse transcription.) Alternatively, some minus strand DNA synthesis may occur outside of the cores isolated in these experiments.

The above data, in conjunction with that obtained in chapters one and two, establish several important features of viral maturation: 1) Viral proteins selectively package the shortest genomic RNA into core particles. 2) Minus strands DNA synthesis is completed in the cytoplasm. 3) There is a block to early plus strand DNA synthesis,

which results in an preponderance of minus strands at steady state. 4) Synthesis of mature-length plus strands is accompanied by association with membranes.

These features raise a number of questions pertaining to mechanism. It is not yet clear when, where, or how the template RNAs are segregated from the pool of messenger RNAs. Several models that can explain the selective packaging of the shortest genomic RNA were presented in chapter 2 and will not be reiterated here. The data in this chapter suggesting the presence of short minus strands outside of cores imply that reverse transcription may have initiated on some RNA templates prior to their packaging; if this is true, however, it appears not to be a prerequisite for packaging, because roughly a third of the genomic RNAs in cytoplasmic cores behave as if they are not annealed to viral minus strands (see chapter 2). The block to plus strand synthesis described above is distinct from the block that leads to termination of most plus strand synthesis when the chains reach 2.0 - 2.2 kb in length. The early block could be kinetic or could result from the accumulation of particles defective in subsequent steps. Two steps that are obvious candidates for causing delay or abortion of plus strand DNA synthesis are the transfer of the RNA primer to the site of initiation of plus strand DNA synthesis and the transfer of the nascent plus strand 3'-end beyond the 5'-end of the minus strand DNA template (Seeger et al., 1986). Finally, the observation that completion of plus strand DNA synthesis in cores is accompanied by their association with membranes may simply reflect a temporal association or the two events may be linked mechanistically.

From work on the export of hepadnaviral surface antigens, one can

infer that the membranous compartment(s) with which the mature cores associate is probably one involved in normal protein export. Three surface antigens are synthesized by the mammalian hepadnaviruses from a single long open reading frame. Because they are initiated at different methionine codons, the proteins share carboxy termini but differ at their amino termini (Standring, et al., 1984). The smallest of the three, p24, is the major protein of lipaceous surface antigen particles secreted into the serum of infected animals; these particles do not contain core components and are far more numerous than true virions. p24 can be secreted from cultured cells that do not synthesize any other viral proteins. In contrast, the largest of the three surface antigens, p45, is not secreted from cultured cells that synthesize just this viral component. p45 is also highly enriched in serum virions, compared to surface antigen particles (Heermann et al., 1984), and can specifically inhibit in trans the secretion of p24 from cultured cells (Persing et al., 1986; Standring et al., 1986; Chisari et al., 1986; Cheng et al., 1986). p45 is myristylated at its amino terminus, a feature that may play an important role in the protein's behavior (Persing et al., 1987).

An attractive model emerging from these findings is that p45 may normally require association with core components for export. p24 (and probably all three surface antigens) is initially synthesized as a transmembrane protein of the endoplasmic reticulum (ER; Patzer et al., 1986; Eble et al., 1986). This protein then traverses the ER and the golgi, acquiring endoglycosidase-H (endo-H)-sensitive and then endo-H-resistent carbohydrates, and is secreted from the cell (Patzer et al., 1984). Thus, p24 apparently can bud into the ER and complete this

pathway without association with other viral components (Eble et al., 1986). By requiring association with viral core components in order to enter this pathway, p45, on the other hand, may act as a docking protein, uniting cores with surface coats and funnelling cores into the export pathway. If completion of plus strand DNA synthesis is required for association of cores with the ER, this might, in turn, represent a strategy for selecting mature cores for export. The selection must not be perfect in all hepadnaviruses, however, because some particles found in the serum of HBV carriers appear to contain viral RNA (Miller et al., 1984). Furthermore, some single-stranded minus strand DNA (without associated plus strand DNA) appears to be associated with membranes in DHBV-infected duck embryos, suggesting that the data obtained from chronically infected ground squirrels may not apply universally to the hepadnavirus group (J. Tuttleman, PhD dissertation; Fox Chase Cancer Center, 1986).

Studies of DHBV suggest that some cores may donate their DNA genomes to the nucleus in an intracellular process that regenerates covalently closed circular DNA, the presumptive template for viral transcription (Tuttleman et al., 1986). Another potential role for p45, then, is to determine the fractions of cores sent to the nucleus or exported. Consistent with this notion is the finding that several hepatoma cells bearing transfected viral DNA appear to overexpress the transcript for p45, relative to the levels seen in vivo; these cells also lack CCC viral DNA (Sureau et al., 1986; Yaginuma et al., 1987; Tsurimoto et al., 1987).

The availability of cultured cells that appear to support otherwise normal viral maturation (Sureau et al., 1986; Yaginuma et

al., 1987; Tsurimoto et al., 1987) permits tests of most of these issues. For example, the phenotype of a mutant deficient in p45 synthesis is of obvious importance in establishing the role of this protein in viral maturation. To address whether completion of plus strand DNA synthesis is required in cis for membrane association, a lesion can be made in the binding site (direct repeat 2 (Seeger et al., 1986)) for the RNA that primes plus strand DNA synthesis. This construct can then be transfected into cells that harbor helper viral DNA, to ensure the production of essential trans-acting factors. Viral DNA in each subcellular compartment can then be examined for the presence of the lesion, with the expectation that templates defective in plus strand synthesis might not generate DNAs associated with membranes. Cis-acting sequences required for RNA packaging can also be identified in this manner; informative mutations are directly suggested by the models outlined in chapter 2.

References

Cheng, K.-C., Smith, G.L., and B. Moss (1986) J. Virol. 60:337-344

Chisari, F., Filippi, P., Milich, D., McLachlin, A., Brinster, R., and R. Palmiter (1986) J. Virol. 60:880-887

Eble, B. Lingappa, V., and D. Ganem (1986) Mol. Cell. Biol. 6:1454-1463

Ganem, D., Weiser, B., Barchuk, R., Brown, J., and H.E. Varmus (1982) J. Virol. 44:366-373

Ganem, D. and H.E. Varmus (1987) Ann. Rev. Biochem. 56:651-93

Heermann, K.H., Goldmann, U., Schwartz, W., Seyffarth, T., Baumgarten, H., and W.H. Gerlich (1984) J. Virol. 52:396-402

Kaplan, P.M., Greenman, R.L., Gerin, J.L., Purcell, R.H., and W.S. Robinson (1973) J. Virol. 12: 995-1005

Landers, T.A., Greenberg, H.B., and W.S. Robinson (1977) J. Virol. 23: 368-376

Maniatis, T., Fritsch, E.F., and J. Sambrook (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Miller, R.H., Tran, C.-T., and W.S. Robinson (1984) Virology 139:53-63

Patzer, E.J., Nakamura, G.R., and A. Yaffe (1984) J. Virol. 51:346-353

Patzer, E., Nakamura, G., Simonsen, C., Levinson, A. and R. Brands (1986) J. Virol. 58:884-892

Persing, D., Varmus, H., and D. Ganem (1986) Science 234:1388-1392

Persing, D.H., Varmus, H.E., and D. Ganem (1987) J. Virol. 61:1672-1677

Seeger, C., Ganem, D., and H.E. Varmus (1986) Science 232:477-83

Standring, D.N., Rutter, W.J., Varmus, H.E., and D. Ganem (1984) J. Virol. 50:563-571

Standring, D., Ou, J., and W. Rutter (1986) Proc. Natl. Acad. Sci USA 83:9338-42

Summers, J. and W.S. Mason (1982) Cell 29:403-415

Sureau, C, Romet-Lemonne, J.-L., Mullins, J.I., and M. Essex (1986) Cell 47:37-47

Tsurimoto, T., Fujiyama, A., and K. Matsubara (1987) Proc. Natl. Acad. Sci. USA

Tuttleman, J.S., Pugh, J.C., and J.W. Summers (1986) J. Virol. 58:17-25

Weiser, B., Ganem, D., Seeger, C., and H.E. Varmus (1983) J. Virol. 48:1-9

Yaginuma, K., Shirakata, Y., Kobayashi, M., and K. Koike (1987) Proc. Natl. Acad. Sci. USA

Chapter 4

Strategies for generating cell lines that efficiently use the genomic promoter of GSHV

For years the glaring need in hepadnavirus research has been a tissue culture system that permits viral growth. Toward this end, several groups have transfected HBV DNA into established cell lines. This approach has commonly achieved faithful and efficient synthesis of surface antigen transcripts and proteins (Dubois et al., 1980; Gough et al., 1983; Standring et al., 1984), but synthesis of core antigens has been at best inefficient. Early experiments yielded neither genomic transcripts of the correct structure nor detectable virus. These results are consistent with the notion that efficient function of the genomic promoter, which directs both the expression of core antigens and the synthesis of the pregenomic RNA, may require factors specific to the liver. Furthermore, this problem is not easily circumvented by replacing the genomic promoter with a heterologous promoter, because, among other reasons, synthesis of a functional pregenomic RNA likely necessitates exact positioning of the 5'-end (see chapters 2 and 5).

We chose two approaches to the problem. Given that most of the experiments described above employed non-hepatic cells, our first approach was to begin examining well-differentiated hepatoma cell lines, to identify any that can accurately initiate transcription from the genomic promoter of GSHV. We know of only a single cell line established from ground squirrel liver (D. Ganem, unpublished results), and this line appears not to have preserved a well-differentiated hepatocyte phenotype. Some human and rat hepatoma cell lines are available that maintain transcription of many liver-specific genes (Nakabayashi et al., 1982). However, careful studies by Darnell and colleagues, using many different culture conditions, have demonstrated

that in vitro culture of hepatocytes results in greatly reduced transcription of most liver-specific genes (Clayton et al., 1985). Thus, attempts to find a human hepatoma cell line that can produce GSHV may be thwarted either by species-specific effects or because the cells may have lost essential transcriptional factors present in vivo. If the available hepatoma cell lines do not permit use of the promoter, an alternative approach is to try to select variants or mutants of these cells that can use the promoter. It is not clear, however, whether a single mutation could overcome the transcriptional block in the parental cells; there may be multiple species-specific barriers, missing factors, or suppressors preventing promoter function (cf. Killary and Fournier, 1984).

Our second approach was therefore designed to generate directly a cell line with the desired characteristics--a ground squirrel liver line that maintains the factors required for use of the genomic promoter of GSHV. To do this, we sought to generate tumors from ground squirrel liver in such a way that tumor outgrowth is dependent upon use of the GSHV genomic promoter. Our experimental design was based on the observation that injection of hepadnaviral DNA directly into the liver of susceptible hosts readily elicits viral infection (Seeger et al., 1984). At some frequency, hepatocytes must be capable of assimilating the injected DNA into a transcriptionally competant form. Thus, by injecting constructs in which viral oncogenes are expressed from the GSHV genomic promoter, we hoped to generate tumors that maintain use of this promoter. We would then try to culture cells from the tumor and examine them for the ability to support transcription of GSHV DNA. In addition to the potential utility of such cells in analysing GSHV

transcription and replication, this approach, if successful, may provide improved experimental access to tissue-specific transcriptional factors and may afford simple and rigorous assays for the ability of genes to produce tumors in specified tissues.

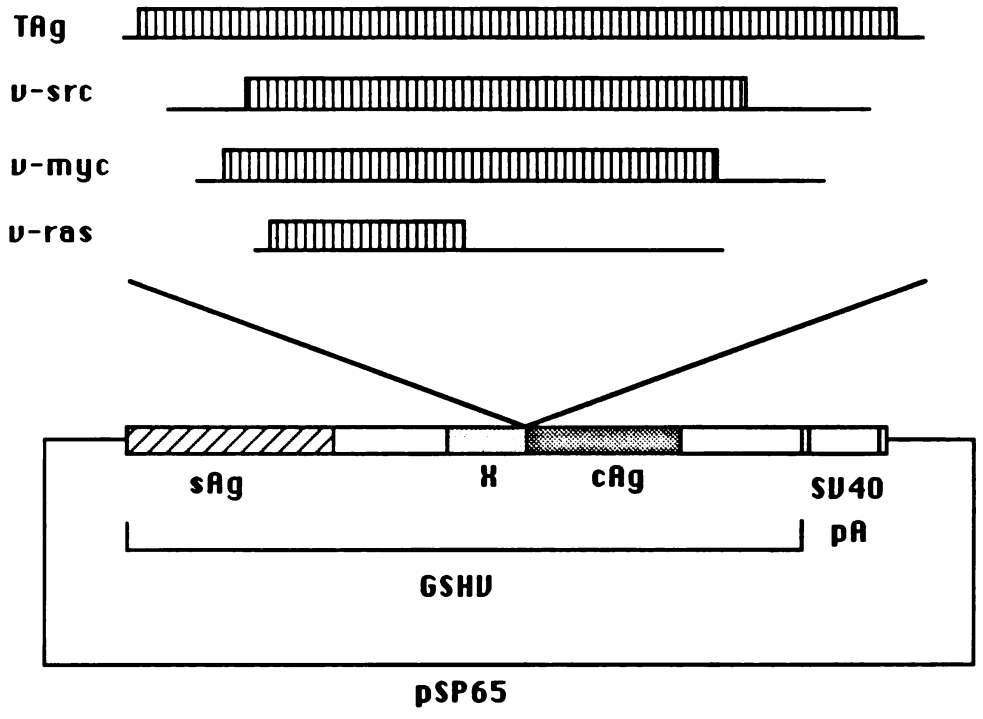
The experiment outlined here is, however, vitiated by two uncertainties. First, the components of the GSHV genomic promoter have not been identified. As a result, any construct made incorporates assumptions about what is required for promoter function. We assumed that the required sequences lie between the transcription initiation sites and a site 1.5 kb 'upstream'. To facilitate insertion of the viral oncogenes, we introduced a single base change at position -7, thus generating an EcoRl restriction site. Cloning into the new site conveniently places the inserted genes upstream of the precore and core translation initiation codons (positions +1 and +91, respectively). This site is also just downstream of two of the major transcription initiation sites (-25 and -17) and just upstream of the third (+6).

The mutation was generated using an M13 subclone of the region and the double-primer method (Norris et al., 1983). The presence of a single base change was confirmed by dideoxy nucleotide sequencing and the relevant region was cloned back into the viral genome. A BamHl linker was inserted into the new EcoRl site in such a way that an EcoRl site was regenerated on both sides of the linker. Then a Pstl fragment with ends 1.5 kb upstream and 1.0 kb downstream of the linker was cloned into the polylinker of pSP65, generating the parental plasmid shown in Figure 1. Note that the major surface antigen promoter of GSHV (see chapter 1 and Standring et al., 1984) has been excluded from the plasmid. This was done for two reasons--to prevent expression of

Figure 1. Structure of the gp-onc constructs.

A G to T tranversion was generated at nucleotide 3303 in the GSHV genome (between the X and cAg genes (Seeger et. al., 1984)), downstream of the putative GSHV genomic promoter (see text), using the double-primer method (Norris et al., 1983). The resultant EcoRl site was used to introduce a single BamHl linker (sequence CGCGGATCCGCG), regenerating flanking EcoRl sites. The large GSHV DNA Pstl fragment (nucleotides 1518-936) bearing this alteration was cloned into the polylinker of pSP65 (Promega) upstream of an SV40 DNA fragment (BamHl-Bcll (nucleotides 2517-2753)) bearing the early region polyadenylation signal (Lanoix et al., 1986). DNA fragments encoding the following oncogenes were then cloned into the novel restrictions sites in GSHV DNA: SV40 TAg (Bgll-BamHl (nucleotides 5218-2516)), v-src (Bgll-EcoRl (nucleotides ca. 7000-9238); gift of J. Kaplan), v-myc (EcoRV-Pstl (nucleotides -35 in Alitalo et. al. (1983)-1673; gift of A. Bruskin), and v-Ha-ras (Smal-EcoRl (nucleotides 1009-2540); gift of M. Scott).

Fig 1



the inserted oncogenes from this promoter following illegitimate recombination during transfection and to reduce the chance that expression of viral surface antigens may generate an immune response against the transfected cells. However, there is no obvious way to test whether the parental construct preserves a functional genomic promoter.

A second important uncertainty in the design of this experiment concerns the choice of oncogenes. Hepatomas have recently been found in ground squirrels, but the responsible gene(s) has not been identified, and no tumors have been experimentally induced in these animals. Two genes are known to induce hepatomas in other experimental settings. The v-myc gene expressed by an infecting MC29 retrovirus is associated with hepatoma development in birds, and rearrangements and enhanced expression of c-myc has been observed in some woodchuck hepatomas (Moroy et al., 1986). The SV40 TAg gene can induce hepatomas in transgenic mice when expressed from a promoter active in the liver (Messing et al., 1985), and TAg has been used to establish cell lines from primary cultures of adult and fetal rat hepatocytes (Isom and Georgoff, 1984; Schlegel-Hauter et al., 1980). It the latter instance, it was shown that continued expression of TAg was required to maintain the transformed phenotype. This characteristic is particularly attractive, given our interest in selecting cells that maintain use of the promoter driving expression of the oncogene. We therefore cloned these two genes separately into the novel EcoRl-BamHl sites of the parental plasmid. We also cloned into these sites two other oncogenes that have relatively broad oncogenic potentials, v-Ha-ras and v-src. Activated c-Ha-ras genes have been isolated from some mouse hepatomas

(Reynolds et al., 1986; Wiseman et al., 1986; Reynolds et al., 1987).

We injected the four constructs (termed gp-oncs) into the livers of a total of nine ground squirrels and two mice, as outlined in table 1. The DNA was injected in the form of calcium-phosphate precipitates into livers exposed by sterile surgical technique (Seeger et al;, 1984). All four gp-oncs were co-precipitated for injection into four ground squirrels, one BALB/c mouse, and one nude (nu/nu) mouse. The mice were chosen as targets because of their ready availability and their proven susceptibility to SV40 TAg-induced hepatoma. A nude mouse was included because of the increased susceptibility of nude mice to papovavirus-induced neoplasms (Allison et al., 1974). In addition, two ground squirrels were injected with a mixture of gp-ras and gp-myc and two others received both gp-TAg and gp-src. These sub-pools were used to speed identification of any construct that might prove oncogenic and to mitigate the possibility that a particular gp-onc might be lethal to the transfected cell. Finally, all four gp-oncs were injected into a ground squirrel that has repeatedly tested positive for serum surface antigen. This animal was chosen because of the increased susceptibility of human (Beasley et al., 1981), woodchuck (Chen et al., 1986), and perhaps ground squirrel (Marion et al., 1986) surface antigen carriers to hepatoma development and because this may lessen the chance of immune rejection of transfected cells due to expression of viral antigens.

The animals were followed by weekly visual inspections, looking for obvious masses, behavioral changes, or other signs of illness. One ground squirrel died a few weeks after the injections of an apparently unrelated cause. No unusual findings were noted in any of the other

Table 1. Animals injected with the gp-onc constructs.

	constructs:			
animals injected	gp-TAg	gp-src	gp-ras	gp-myc
uninfected ground squirrel 1	+	+	+	+
" " " 2	+	+	+	+
" " " 3	+	+	+	+
" " " 4	+	+	+	+
" " " 5	+	+	-	-
" " " 6	+	+	-	-
" " " 7	-	-	+	+
" " " 8	-	-	+	+
infected ground squirrel	+	+	+	+
BALB/c mouse	+	+	+	+
nu/nu mouse	+	+	+	+

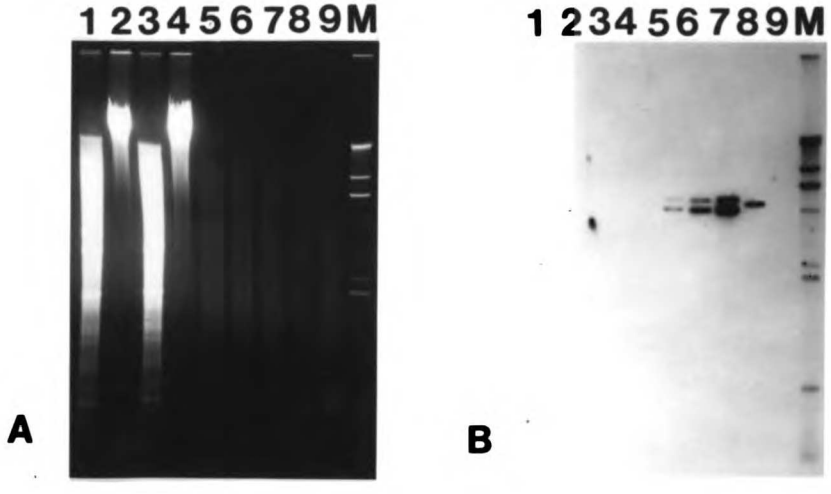
animals over the ensuing four months. We therefore opened the abdomens of the ground squirrels under sterile conditions (laparotomy) and inspected their livers. Six livers appeared grossly normal; one appeared unusually large, though otherwise normal; and one, from an animal that had recieved gp-TAg and gp-src, had a firm, yellowish mass bulging from its ventral aspect. Histological examination of the mass revealed it to be either a well-differentiated hepatoma or, less likely, a hepatic adenoma.

We prepared high molecular weight DNA from a portion of the tumor and subjected it to Southern blotting, using GSHV and SV40 probes. The GSHV probe yielded only a faint low molecular weight smear, in lanes containing cut or uncut tumor DNA (data not shown). The signal strength was not accurately calibrated in this experiment but, as judged from control DNA samples, appeared sufficient to have detected the presence in high molecular weight tumor DNA of a single copy per cell of the injected constructs. The faint low molecular weight signal may have resulted from a lowgrade GSHV infection of the tumor. (The negative results from assays for surface antigen in this animal's serum cannot exclude the presence of a low-grade infection; further experiments on the remaining tissue are needed to decide this point.) The SV40 probe failed to yield a signal with tumor DNA, under conditions adequate to detect a single copy per cell of the gp-TAg construct (Fig 2). Since the SV40 probe also contained vector sequences present in both gp-TAg and gp-src, this result argues strongly against generation of the tumor by stable integration of one of the injected constructs. We also probed this Southern blot for v-src sequences and assayed src kinase activity in tumor material. Both

Figure 2. Probing tumor DNA for the presence of the injected constructs.

High MW DNA was prepared from the hepatic tumor found in one of the animals injected intrahepatically with gp-TAg and gp-src. The DNA was subjected to electrophoresis through 1% agarose, transferred to nitrocellulose, and allowed to anneal with 32P-labelled SV40 DNA cloned into a pBR322 vector (pHR401; gift of D. Ganem). Lanes 1, 2-normal ground squirrel liver DNA (10 ug) cut with EcoRl and uncut, respectively. Lanes 3, 4-tumor DNA (10 ug) cut with EcoRl or uncut. Lanes 5,6,7-salmon sperm DNA (5 ug) supplemented with 50, 200, and 1000 pg of plasmid DNA containing total SV40 sequences (pSVRl; gift of M. Verderame), cut with EcoRl. Lane 8-salmon sperm DNA (1 ug) supplemented with ca 500 pg gp-src. Lane 9-same as lane 7, but uncut. M-32P-labelled lambda Hindlll DNA fragments. Left-ethidium bromide-stained gel. Right-autoradiograph of probed nitrocellulose filter.

Fig 2



assays were negative, although positive controls were not included in these two studies. Thus, the data suggest that the tumor was not induced by the injected DNA, unless this DNA was never integrated or was lost during tumor outgrowth. Efforts to culture cells from this tumor were unsuccessful; these efforts were hampered by the untimely death of the animal two weeks after discovery of the tumor.

The remaining animals were followed for varying amounts of time. The ground squirrel carrying GSHV surface antigen in its serum died eight months after the injection. Histological examination of the liver found no hepatoma tissue. Laparotomies performed on the remaining five ground squirrels 12 months after the injections revealed no obvious pathology. The euthymic mouse lived in apparent good health for more than 13 months after the injection. The nude mouse lived for six months after the injection before dying from what appeared to be an infectious disease. No abdominal masses or other signs of neoplastic disease were observed during weekly visual inspections of the animal.

Our initial attempts to identify hepatoma cell lines that suport accurate hepadnaviral transcription focused on the Fao line. This line, derived from a rat hepatoma, is the best characterized hepatoma line available. The cells are relatively well-differentiated. Furthermore, phenotypic variants and revertants have been identified that have, respectively, lost and regained expression of some liver-specific genes. We therefore considered this line an attractive one in which to examine GSHV transcription. Our first effort to detect transcription of GSHV DNA in transiently transfected cells of the Fao parental line was unsuccessful (data not shown).

While this work was in progress, DHBV was found to replicate in

primary cultures of duck hepatocytes (Tuttlemen et al., 1986). The cells are infectable and yield the expected products. However, the cells are only transiently susceptible to infection, and it is not yet clear whether replication can be reproducibly achieved following DNA transfections. Several human hepatoma cell lines have recently been identified that support most steps of viral replication following transfection with HBV DNA, including faithful synthesis of the major viral RNAs and replicative form DNA (Sureau et al., 1986; Yaginuma et al., 1987; Tsurimoto et al., 1987). Three issues pertaining to the use of these cell lines remain unresolved: 1) Can the cells be infected? 2) Do the cells produce infectious virus? 3) Why don't these cells produce covalently closed circular viral DNA? CCC DNA is the likely transcriptional template in hepadnaviral infections. It is generated from virion and probably from intracellular replicative form DNA, and its absence suggests that these cells do not support the full viral life cycle. Despite these reservations, the cells will no doubt be quite useful in studies of hepadnavirus biology.

We are examining whether these human hepatoma cell lines permit transcription and replication of GSHV; the availability of such a cell line would unite the full advantages of tissue culture with an animal system. We have not yet been able to convincingly demonstrate GSHV transcription following transient transfection with GSHV DNA (data not shown); parallel transfections with HBV DNA demonstrated production of both major classes of viral RNA (data not shown). We have also attempted to select clones that support use of the GSHV genomic promoter. To do this, we cloned a gene conferring resistence to the antibiotic G418 into the novel EcoRl-BamHl sites described above,

transfected this construct into the hepatoma cells, and selected for resistant colonies. Several such attempts using two of the hepatoma cell lines known to support HBV transciption failed to yield any resistant colonies, whereas control transfections, using constructs in which the G418-resistance gene is expressed from other promoters, were successful (data not shown). These reults suggest either that these hepatoma cell lines do not support efficient use of the GSHV genomic promoter or that the genomic promoter constructs do not preserve adequate promoter function.

Even if these human hepatoma cell lines can be shown to support authentic GSHV transcription, it would be preferable to have a line derived from a ground squirrel. Furthermore, the cell lines described above, like other cultured hepatocytes currently available, do not maintain in vivo levels of transcription of liver-specific genes. Levels of hepadnaviral RNA in these lines are low compared to many infected livers (chapter 1), perhaps for multiple reasons. Thus, the tumor-induction experiments described in this chapter remain germane. We are considering the use of constructs containing well-characterized promoters (such as the SV40 early promoter and the mouse albumin promoter), to try to better assess whether the direct DNA injection technique is practicable. In addition, we are attempting to generate transgenic mice bearing the above gp-onc constructs, with the hope of establishing cell lines from tumors that may arise in this setting.

References

Alitalo, K., Bishop, J.M., Smith, D.H., Chen, E.Y., Colby, W.W., and A.D. Levinson (1983) Proc. Natl. Acad. Sci. USA 80:100-104

Allison, A.C., Monga, J.N., and V. Hammond (1974) Nature 252: 746-747

Beasley, R.P., Lin, C.-C., Hwang, L.-Y., and C.-S. Chien (1981) Lancet 2:1129-33

Chen, P., Kalpana, G., Golberg, J., Mason, W., and B. Werner (1986) Proc. Natl. Acad. Sci. USA 83:8774-78

Clayton, D.F., Weiss, M., and Darnell, J,.E. (1985) Mol. Cell. Biol. 5:2623-2632

Dubois, M.-F., Pourcel. C., Rouset, S., Chany, C., and Tiollais, P. (1980) Proc. Natl. Acad. Sci. USA 77:4549-53

Gough, N. (1983) J. Mol. Biol. 165:683-99

Isom, H.C. and I. Georgoff (1984) Proc. Natl. Acad. Sci. USA 81:6378-6382

Killary, A.M. and R.E.K. Fournier (1984) Cell 38: 523-534

Lanoix, J., Tseng, R.W., and N.H. Acheson (1986) J. Virol. 58:733-742

Marion, P., Van Davelaar, M., Knight, S., Salazar, F., Garcia, G., Popper, H., and W.S. Robinson (1986) Proc. Natl. Acad. Sci. USA 83:4543-46

Messing, A., Chen, H.Y., Palmiter, R.D., and R.L. Brinster (1985) Nature 316:461-463

Moroy, T., Marchio, A., Etiemble, J., Trepo, C., Tiollais, P., and M.-A. Buendia (1986) Nature 324:276-279

Nakabayashi, H., Taketa, K., Miyano, K., Yamane, T., and J. Sato (1982) Cancer Res. 42:3858-3863

Norris, K., Norris, F., Christiansen, L., and N. Fiil (1983) Nuc. Acids Res. 11:5103-5112

Reynolds, S.H., Stowers, S.J., Maronpot, R.R., Anderson, M.W., and S.A. Aaronson (1986) 83:33-37

Reynolds, S.H., Stowers, S.J., Patterson, R.M., Maronpot, R.R., Aaronson, S.A., and M.W. Anderson (1987) Science 237:1309-1316

Schlegel-Hauter, S., Schlegel, W., and J.Y. Chou (1980) Proc. Natl. Acad. Sci. USA 77:2731-2734

Seeger, C., Ganem, D., and H. Varmus (1984) Proc. Natl. Acad. Sci. USA 81:5849-52

Standring, D.N., Rutter, W.J., Varmus, H.E., and D. Ganem (1984) J. Virol. 50:563-71

Sureau, C., Romet-Lemonne, J.-L., Mullins, J.I., and M. Essex (1986) Cell 47: 37-47

Tsurimoto, T., Fujiyama, A., and K. Matsubara (1987) Proc. Natl. Acad. Sci. USA

Tuttleman, J.S., Pugh, J.C., and J.W. Summers (1986) J. Virol. 58:17-25

Wiseman, R.,W., Stowers, S.J., Miller, E.C., Anderson, M.W., and J.A. Miller (1986) Proc. Natl. Acad. Sci. USA 83:5825-5829

Yaginuma, K., Shirakata, Y., Kobayashi, M, and Koike, K. (1987) Proc. Natl. Acad. Sci. USA

Chapter 5

Requirements for polyadenylation of GSHV RNAs

Each GSHV genomic RNA contains an intact copy of the viral genome plus a region copied twice, generating terminal redundancies (chapter 1). Synthesis of these RNAs requires that they be polyadenylated only after the second transit of the transcription complex past the viral polyA addition site. The terminal redundancies so generated in the shortest of the three genomic RNAs appear to be important for reverse transcription of the viral genome (Seeger et al., 1986). In addition, 'readthrough' of the polyA site is essential for expression of viral core proteins, since the polyA site lies early in the core gene. PolyA site readthrough occurs in each of the hepadnaviruses (Ganem and Varmus, 1987), in many other elements that use reverse transcription for replication or transposition (Varmus and Swanstrom, 1985), and in the regulation of expression of some cellular genes (Early et al., 1980; Rosenfeld et al., 1984; Birnsteil et al., 1985).

The molecular basis of polyA site readthrough is not understood in any of these systems. The readthrough phenomenon in GSHV (as in most of the other systems) has not been unequivocally distinguished from simple 'leakiness' of polyadenylation. There is evidence that the GSHV polyA site, like many others, may be intrinsically slightly leaky: RNAs roughly a genome-equivalent larger than the two major size classes are faintly visible on Northern blots of RNA from GSHV-infected liver (chapter 1). Thus, it appears likely that 5% or so of the RNAs initiated at the surface antigen promoter escape polyadenylation on their first pass of the polyA site and that an equal fraction of RNAs initiated at the genomic promoter escape polyadenylation on their second pass of the site. If polyadenylation occurs at this same

efficiency following the first pass of the polyA site by transcription complexes initiated at the genomic promoter, then short (130-160 bases of GSHV sequence), apparently non-functional, polyadenylated RNAs must be synthesized at about 20-fold the rate of the genomic RNAs. Such short RNAs have not been detected, but they might not transfer well to the nitrocellulose or nylon filters that have been used for Northern blotting. We have sought to detect such RNAs, using uniformly-labeled S1-nuclease probes of the relevant region, but the experiments have thus far failed for technical reasons. Even if these RNAs are absent from our preparation of RNA from infected liver, it remains possible that the short species are made but are unstable in the cell or not efficiently recovered by the isolation procedure we used. Pulse-labeling studies are required to exclude these possibilities.

If not explicable simply by leaky polyadenylation, the hepadnavirus polyadenylation readthrough presents a particularly telling example of the phenomenon, because the template for transcription is apparently covalently closed circular DNA (cccDNA) and the resultant RNAs are unspliced (see chapter 3). Thus, the local RNA sequences surrounding the polyA site, or 'contexts', are identical on each pass and are not modified by subsequent splicing events. This fact suggests that transcription of remote sequences may affect in cis the efficiency of polyadenylation at a 'downstream' site. One can imagine such transcription altering the secondary structure in the vicinity of the polyadenylation site, the functional potential of the transcriptional complex, or the ability of other trans-acting factors to interact with the nascent RNA. Study of this phenomenon may therefore offer insight into dynamic properties of transcription.

The requirements for polyadenylation of eucaryotic pre-mRNAs remain incompletely specified. A consensus polyadenylation signal, AAUAAA, is almost invariably present 10-30 bases upstream of the polyadenylation site (Birnsteil et al., 1985). This 'primary' signal is, however, insufficient to induce polyadenylation: it appears within some transcription units in the absence of detectable polyadenylation. Efficient polyadenylation at many sites also requires a GU-rich sequence 10-40 bases downstream of the primary signal (Simonsen and Levinson, 1983a; McDevitt et al., 1984; Birnsteil et al., 1985). Recent studies of small nuclear RNA (snRNA) transcription have demonstrated that untranscribed sequences in the promoter region can influence the efficiency of polyadenylation at a downstream site (Hernandez and Weiner, 1986; Neuman de Vegvar et al., 1986). Finally, the relationship between transcription termination and polyadenylation remains obscure. In yeast, the two events appear to be tightly coupled (Zaret and Sherman, 1982), but this appears not to be the case in higher eucaryotes (Birnsteil et al., 1985). There is now convincing evidence that polyadenylation in higher eucaryotes usually occurs by endonucleolytic cleavage of a precursor, followed by polyA addition (Moore and Sharp, 1985), and that this can occur prior to transcription termination (Nevins and Darnell, 1978). Evidence has been obtained that ribonucleoprotein particles containing snRNAs are involved in recognition of the polyadenylation site (Hashimoto and Steitz, 1986).

The GSHV primary polyadenylation signal is a variant of normal (UAUAAA) and is not followed by a GU-rich sequence. However, the HBV primary signal matches the consensus and was the first to reveal the importance of a downstream GU-rich sequence. Thus, the HBV

polyadenylation site appears typical in every respect, yet it also displays the readthrough phenomenon described above for GSHV. Furthermore, a DNA fragment spanning the 300 bases upstream and 200 downstream of the HBV polyadenylation site functions efficiently in a number of heterologous systems (Simonsen and Levinson, 1983b; Miller et al., 1985). This latter finding suggests either that the sequences necesary to specifiy polyadenylation are included in this fragment or that whatever signals are missing from the fragment are supplied by many transcription units. Thus, there is no evidence at present that the polyA sites that display readthrough are intrinsically unusual, in terms of the characteristics that we can presently identify.

The studies of snRNA transcription are the first to show that upstream sequences can be important in polyadenylation. However, this phenomenon--a promoter effect--appears distinct from the readthrough phenomenon seen in hepadnavirus transcription, since the issue here is why transcription initiated at a single promoter does not permit polyadenlyation on the first pass of a given site but permits it on the second. It remains possible, of course, that the readthrough phenomenon also depends in some way upon the nature of the transcription complex assembled at the promoter.

We have formulated three working models to explain the polyadenylation readthrough in GSHV. The first posits that the key variable is simply the distance between the transcription initiation site and the polyA site; too short a distance may preclude polyadenylation. In the second model transcription from the genomic promoter forms a secondary structure on first pass of the polyA site that prevents polyadenylation; on the second pass, the presence of

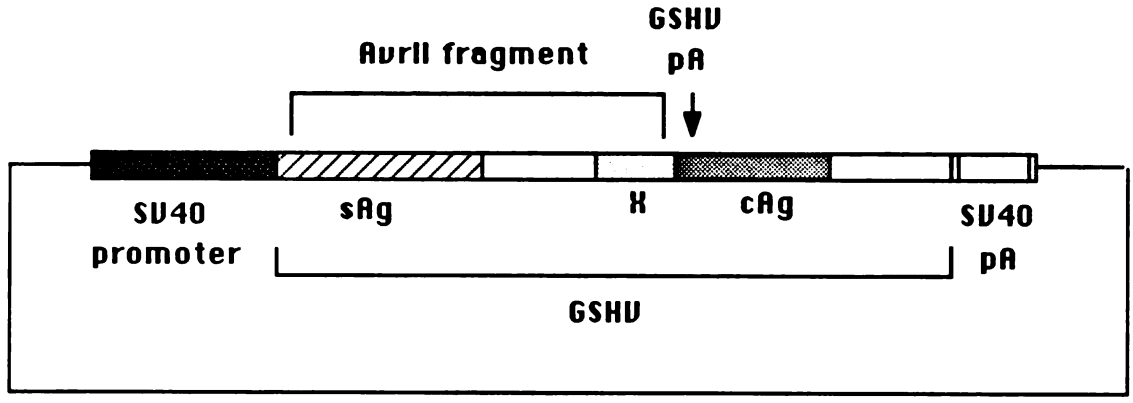
contiguous upstream sequences induces formation of a secondary structure compatible with polyadenylation. In the third, transcription of a specific sequence upstream of the genomic promoter (and downstream of the surface antigen promoter) is independently required for efficient polyadenylation. This sequence would not affect the secondary structure surrounding the polyadenylation site; rather, it would affect the interaction of the transcription complex or other factors with the nascent RNA. Given the absence of an obvious consensus polyadenylation sequence upstream of the primary signal in most genes, it seems likely that such a facilitating upstream sequence may have subtle sequence requirements. An example of such requirements might be the lack of a stable secondary structure, a feature that is essential for rho-dependent transcription termination in prokaryotes (Chen et al., 1986). Other explanations of the readthrough phenomenon are possible; these three appear to be several of the more plausible and testable.

To test these models, we sought a system in which efficient transcription of the relevant GSHV sequences does not require viral replication. We cloned a GSHV DNA fragment bearing the polyA site into a SV40-based vector (figure 1). This vector contains the SV40 origin, allowing replication of the plasmid after transient transfection into COS cells. Because the GSHV promoters have not yet been demonstrated to function in tissue culture cells (see chapter 4), the GSHV fragment was positioned to permit transcription of plus-strand GSHV sequences from the SV40 early promoter. The GSHV insert includes the 1800 bases upstream (within 100 bases of the surface antigen promoter) and 900 bases downstream of the polyA site. In addition, to facilitate

Figure 1. Structure of the polyadenylation constructs.

The native large GSHV DNA KpnI-PstI fragment (nucleotides 1610-936) was cloned into the pSP65 vector (Promega) upstream of an SV40 DNA fragment bearing the early region polyadenylation site (BamHI-BclI fragment (nucleotides 2517-2753); Lanoix et al., 1986). An SV40 DNA fragment bearing the early region promoter and the origin of replication (PvuII-HindIII; nucleotides 270-5171) was then cloned upstream of the GSHV DNA fragment, generating the parental plasmid. Derivatives were then made in which the GSHV AvrII fragment (nucleotides 1691-3250), containing nearly all of the GSHV sequences between the GSHV sAg and genomic transcript initiation sites (see text), was, respectively, inverted and deleted. Also marked are the GSHV surface antigen (sAg), X, and core antigen (cAg) genes and the GSHV polyadenylation site.

Fig 1


AvrII fragment
GSHV
pA
SV40
promoter
sAg
X
cAg
SV40
pA
GSHV
pSP65

polyadenylation of transcripts not processed at the GSHV polyA site, we cloned the SV40 early region polyadenylation signal (Lanoix et al., 1986) downstream of the GSHV sequences. However, it is no more possible to predict in advance the conditions under which the SV40 polyA site will be used than it is for the GSHV polyA site. It is also important to bear in mind that, in establishing a more tractable experimental system, we have made alterations that may affect the phenomenon we seek to study. These changes include the use of a new cell type (that lacks liver-specific and viral-encoded factors), a new promoter, and a transcription template with different remote 5′ and 3′ sequences.

To begin to discriminate between our models, we derived two more constructs, in which most of the GSHV sequences originally present between the two major GSHV promoters are inverted or deleted (fig 1). In the parental construct, transcripts are expected to be initiated from the SV40 early promoter and polyadenylated at the GSHV polyA site, since this transcription event mimics normal transcription from the GSHV surface antigen promoter (with the caveats listed above). In the first model ('distance'), inversion of the intervening sequences should have little effect on polyadenylation, since the distance between the SV40 promoter and the GSHV polyA site is unchanged. Deletion of the intervening sequences may or may not abrogate use of the GVHV polyA site, depending upon how much distance is required between the initiation and the polyadenlylation sites in order to maintain efficient use of the latter. Following the deletion there would be ca. 330 bases between the transcription initiation and the GSHV polyA sites (more if the 'late-early' SV40 transcription initiation site is used,

rather than the 'early-early' site (Buchman et al., 1984). In comparison, there are 127-157 bases between the normal GSHV genomic promoter transcription initiation sites and the GSHV polyA site. In this model, if the GSHV polyA site is not used, the SV40 polA site will be, since it is presumably an adequate distance from the promoter (ca 1.2 kb).

The second model ('secondary structure') does not lead to firm predictions about the transcriptional behavior of the mutants. This model states that nascent RNA on the first pass of the polyA site forms a secondary structure incompatible with polyadenylation and that this structure is not formed on the second pass of the site. The model tends to predict that specific upstream sequences are required to form an alternative secondary structure that permits polyadenylation. In this case, both inversion and deletion of the bulk of the intervening sequences will probably eliminate efficient use of the GSHV polyA site. If the requirements for upstream sequences that will disrupt the secondary structure surrounding the polyA site are not stringent, then both the inversion mutation and (less likely) the deletion mutation may still permit efficient polyadenylation at the GSHV site.

The third model ('independent sequence') states that transcription of an upstream sequence facilitates polyadenylation downstream without affecting the local secondary structure encompassing the polyA site. Here again, if the requirements for this independent sequence are loose (eg. the simple absence of secondary structure), then the inverted fragment may also supply such a sequence. However, deletion of most of the upstream region would likely eliminate efficient use of the GSHV polyA site. Hence, the third model predicts substantially decreased

use of the GSHV polyA site with the deletion and probably also the inversion mutations. Whether the downstream SV40 polyA site will be efficiently used in these constructs is not clear.

Construction of more selective mutations should be simple, given the abundance of convenient restriction sites in the mutated region.

Our results using the three constructs described above remain preliminary. We transfected the constructs into COS cells, harvested RNA 48 h later, and analysed the RNA by Northern blotting. When the blot was probed for the presence of plus-stranded GSHV sequences, a major transcript of the expected size (ca 1.8 kb) was obtained from the parental construct (figure 2), consistent with a transcript initiated at the SV40 promoter and polyadenylated at the GSHV site. This transcript is not detected in RNA isolated from COS cells transfected with the mutant constructs, nor are there major new transcripts. The absence of bands in lanes derived from use of the mutant constructs cannot be simply explained by differences in transfection efficiency or in recovery of RNA, because equal amounts of RNA were loaded in each lane, and each lane yielded sharp, moderately dark autoradiographic bands when probed for RNAs derived from other sequences in the vector plasmid (data not shown). Note, however, that in the inversion construct only ca. 200 bases of plus-stranded GSHV sequence would be present in the RNA initiated at the SV40 promoter and polyadenylated at the GSHV site; to better test for the presence of this RNA, the blot should be probed for minus-stranded GSHV sequences. With this caveat, it appears that the mutations prevented the synthesis of stable RNAs containing GSHV plus-strand sequences. The mutations may have prevented polyadenylation at both the GSHV and SV40 sites, or they may

Figure 2. Transient transfection of COS cells with the polyadenylation constructs.

Cos cells were transfected in paralell with each of the polyadenylation constructs, using the DEAE-dextran technique (Maniatis et al., 1982). RNA was prepared by the guanidinium-isothyocyanate procedure and pelleted through CsCl. RNA samples were subjected to electrophoresis through 1% agarose gels in formaldehyde, transferred to nitrocellulose, and probed for the presence of GSHV plus strand sequences (Weiser et al., 1983). Lane 1-control sample of RNA from a GSHV-infected ground squirrel liver (chapter 1). Lane 2,3,4-RNA (ca. 20 ug) prepared from COS cells transfected for 48 h with the parental, AvrII inversion, and AvrII deletion constructs, respectively.

Fig 2

have yielded efficiently polyadenlyated RNAs that are unstable. Other explanations are also possible and difficult to exclude. Our initial attempt to determine the structure of the major transcript generated from the parental construct, using long S1 probes, demonstrated that this will likely be difficult, given the amounts of the plasmid-derived RNA recovered in our conditions.

While this work was in progress, Dougherty and Temin (1987) reported that a small deletion upstream of the spleen necrosis virus (SNV) polyA site abolished efficient polyadenylation at that site. The deletion encompassed 400 nucleotides of SNV sequences ending 60 bases upstream of the AAUAAA sequence. Transcripts initiated from the SV40 early promoter, positioned ca. 1.5 kb upstream, were no longer polyadenylated at that site but were polyadenylated further downstream, at the SV40 late polyadenylation site. These data argue against the 'distance' model and favor transcription of a specific sequence or generation of local secondary structure as a further requirement for polyadenylation at the SNV polyA site.

References

Birnsteil, M.L., Busslinger, M., and K. Strub (1985) Cell 41:349-359

Buchman, A.R., Fromm, M., and P. Berg (1984) Moll. Cell. Biol. 4:1900-1914

Chen, A.C.-Y., Gallupi, G.R., and J.P. Richardson (1986) Cell 46:1023-1028

Dougherty, J.P. and H.M. Temin (1987) Proc. Natl. Acad. Sci. USA 84:1197-1201

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and L. Hood (1980) Cell 20:313-319

Ganem, D. and H.E. Varmus (1987) Ann. Rev. Biochem. 56:651-693

Hashimoto, C. and J.A. Steitz (1986) Cell 45:581-591

Hernandez, N.M. and A.M. Weiner (1986) Cell 47:249-258

Lanoix, J., Tseng, R.W., and N.H. Acheson (1986) J. Virol. 58:733-742

Maniatis, T., Fritsch, E.F., and J. Sambrook (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

McDevitt, M.A., Imperiale, M.J., Ali, H., and J.R. Nevins (1984) Cell 37:993-999

Moore, C.L. and P.A. Sharp (1985) Cell 41:845-855

Neuman de Vegvar, H.E., Lund, E., and J.E. Dahlberg (1986) Cell 47:259-266

Nevins, J.R., and J.E. Darnell, Jr. (1978) Cell 15:1477-1493

Rosenfeld, M.G., Amara, S.G., and R.M. Evans (1984) Science 225:1315-1320

Seeger, C., Ganem, D., and H.E. Varmus (1986) Science 232:477-484

Simonsen, C.C. and A.D. Levinson (1983a) Mol. Cell. Biol. 3:2250-2258

Simonsen, C.C. and A.D. Levinson (1983b) Proc. Natl. Acad. Sci. USA 80:2495-2499

Varmus, H.E. and R. Swanstrom (1985) In Molecular Biology of Tumor Viruses, Chapter 5S. New York: Cold Spring Harbor Press

Weiser, B., Ganem, D., Seeger, C., and H.E. Varmus (1983) J. Virol. 48:1-9

Appendix

Expression of the precore region of an avian hepatitis B virus is not required for viral replication

JOURNAL OF VIROLOGY, Oct. 1987, p. 3322–3325
0022-538X/87/103322-04$02.00/0

Vol. 61, No. 10

Expression of the Precore Region of an Avian Hepatitis B Virus Is Not Required for Viral Replication

CHRISTOPHER CHANG,[1] GREG ENDERS,[1] ROLF SPRENGEL,[2†] NICK PETERS,[3] HAROLD E. VARMUS,[1,2] AND DON GANEM[2,4*]

Departments of Microbiology and Immunology,[2] Biochemistry and Biophysics,[1] and Medicine,[4] University of California Medical Center, and Department of Biology, University of San Francisco, San Francisco, California 94143[3]

Received 21 May 1987/Accepted 26 June 1987

The core-antigen-coding region of all hepadnaviruses is preceded by a short, in-phase open reading frame termed precore whose expression can give rise to core-antigen-related polypeptides. To explore the functional significance of precore expression in vivo, we introduced a frameshift mutation into this region of the duck hepatitis B virus (DHBV) genome and examined the phenotype of this mutant DNA by intrahepatic inoculation into newborn ducklings. Animals receiving mutant DNA developed DHBV infection, as judged by the presence in hepatocytes of characteristic viral replicative intermediates; molecular cloning and DNA sequencing confirmed that the original mutation was present in the progeny genomes. Infection could be efficiently transmitted to susceptible ducklings by percutaneous inoculation with serum from mutant-infected animals, indicating that infectious progeny virus was generated. These findings indicate that expression of the precore region of DHBV is not essential for genomic replication, core particle morphogenesis, or intrahepatic viral spread.

The hepadnaviruses are a group of small, enveloped DNA viruses that produce acute and chronic infections of hepatocytes and that replicate their DNA by reverse transcription of RNA intermediates (5, 21). To date, members of this group have been recovered from hosts of several species, including humans, woodchucks, ground squirrels, and ducks (19). All of these viruses share a similar genomic organization, which includes at least three open reading frames (5, 19, 21). One of these encodes the multiple surface proteins of the envelope, another encodes the reverse transcriptase required for replication, and a third encodes the major structural protein of the viral nucleocapsid or core (core antigen). The mammalian viruses harbor an additional coding region (termed X) of unknown function.

In all hepadnaviral genomes, the coding region for core antigen is preceded by an in-phase, contiguous open reading frame of 29 to 43 codons known as precore (Fig. 1A). The expression of this region is predicted to generate core-related polypeptides with additional amino acids at their N termini. There is good evidence that precore expression does indeed occur in vivo. All hepadnavirus-infected liver cells produce at least two sets of transcripts containing core sequences. One RNA initiates within the precore region and therefore can encode only the core antigen. The other set of transcripts initiates upstream of the precore ATG codon and hence could encode the larger precore polypeptide (1, 3, 11, 24; R. Sprengel and H. Will, unpublished data).

Extensive work with human hepatitis B virus (HBV) indicates that precore polypeptides can be generated from such RNAs and that their properties differ from those of core antigen. Several studies (9, 12, 13) have used heterologous promoters to drive the expression of either precore or core gene products in cultured cells transfected with subgenomic fragments of cloned HBV DNA. When protein synthesis is initiated at the core ATG, the resulting protein is cytoplasmic in location, while initiation of translation at the precore ATG results in the targeting of the gene product to a membranous compartment (12); after further proteolytic processing (9; D. Standring, J. Ou, and W. Rutter, personal communication), the products are secreted into the medium. Similar products (collectively termed hepatitis B e antigen) are observed in the serum of HBV-infected patients, arguing strongly that this pathway is functional in vivo (5, 21). These results suggest that the HBV precore region encodes a signal sequence that can function to direct core antigen determinants to novel subcellular locales. Similarly, we have recently demonstrated that the serum of ducks infected with duck hepatitis B virus (DHBV) also contains secreted polypeptides derived from the expression of the DHBV precore region (R. Sprengel and H. Will, manuscript in preparation). However, the functional role of precore gene products in authentic viral replication in productively infected cells has not yet been examined.

Progress in dissecting the roles of the individual viral gene products in vivo has been hampered by the absence of convenient cell culture systems for viral growth. However, the observation that cloned viral DNA can initiate productive infection of susceptible hosts following intrahepatic inoculation (16, 17, 22) has allowed the development of a simple strategy for the mutational analysis of viral functions. Briefly, lesions in the region of interest are generated by site-directed mutagenesis in vitro, and the resulting mutant genomes are tested for viability in vivo by intrahepatic transfection. We have previously used this strategy to examine the role of short direct repeats (DRs) of viral DNA in genomic replication (15). In this study, we present a similar analysis of the hepadnaviral precore region. Although the limited host range of HBV (which includes only humans and higher primates) makes this approach difficult for the human virus, the animal hepadnavirus models are ideally suited to

* Corresponding author.
† Present address: Max Planck Institute for Biochemistry, Munich, Federal Republic of Germany.

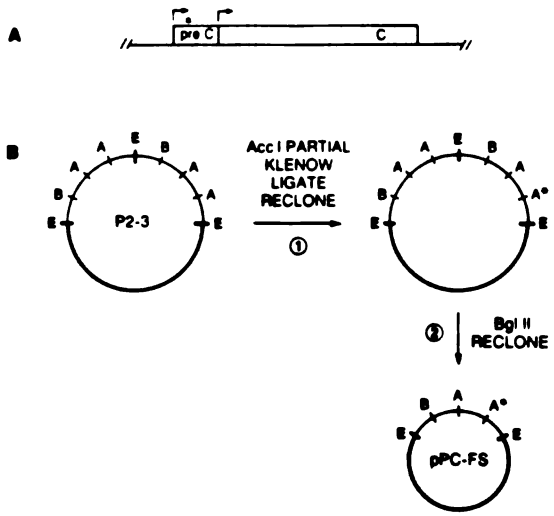


FIG. 1. Construction of precore frameshift mutation in DHBV DNA. (A) The organization of the precore and core-antigen-coding regions of DHBV. □, Coding regions; →, ATG codons; PreC, precore region; C, core-antigen-coding frame; *, site of precore frameshift mutation. (B) Construction of precore frameshift lesion. Plasmid p2-3 is a tandem dimer of DHBV DNA, with restriction sites as indicated. E, *Eco*RI; A, *Acc*I; B, *Bgl*II. The manipulations at each step of the construction are described in detail in the text and are indicated next to the arrow depicting each step. A*, Precore *Acc*I site inactivated by the 2-base-pair insertion.

this form of analysis. The avian system is particularly attractive, since susceptible ducklings are readily available and the incubation period of DHBV infection is only 2 to 3 weeks (versus 2 to 3 months for the mammalian viruses). Accordingly, we used DHBV as the experimental system for these studies.

To address the role of precore sequences in vivo, we constructed a frameshift mutation in the DHBV precore region and tested the mutant DNA for infectivity. The scheme for construction of the mutant is detailed in Fig. 1B. Briefly, plasmid p2-3, bearing a tandem dimer of the DHBV genome (generous gift of W. Mason), was linearized by partial digestion with *Acc*I; one of the two *Acc*I cleavage sites in DHBV DNA lies within the precore region (8). The termini of gel-purified linear molecules were then repaired with DNA polymerase I (Klenow fragment), religated, and cloned. Clones which had lost the appropriate *Acc*I site in the precore region were identified by restriction mapping. The wild-type copy of the DHBV genome was then excised with *Bgl*II to generate clones bearing only the mutant genome. In one of these (pPC-fs), the presence of the expected two-base insertion was verified by DNA sequencing (data not shown).

For infectivity testing, mutant DHBV genomes were liberated from the plasmid vector by *Eco*RI digestion, and the released molecules were self-annealed and ligated at low DNA concentrations (1 to 10 μg/ml) to promote recircularization. A 5-μg sample of this DNA (in 0.2 ml of 500 μg of DEAE-dextran per ml) was percutaneously injected into the livers of each of six virus-free ducklings (SPAFAS, Inc., Norwalk, Conn.) 1 day posthatching. Four control ducklings were similarly exposed to wild-type DNA. Three weeks later, all ducks were sacrificed and their liver DNA was examined for progeny DHBV sequences by Southern blotting with ^{32}P-labeled DHBV DNA as a probe. Five of the six mutant-infected samples (and one of the four controls) revealed the characteristic array of replicative intermediates typical of productively infected hepatocytes (20), i.e., asymmetric, protein-linked, heterogeneous forms and fully duplex monomeric circles lacking covalently attached protein (two representative examples are shown in Fig. 2C). To verify that the replicative forms seen in mutant-infected livers were of mutant origin, two experiments were carried out. First we examined the restriction pattern of the intrahepatic circular duplex viral DNA (Fig. 2A). Liver DNA from all five infected animals was prepared by phenol extraction in the absence of proteinase K; this procedure eliminates the abundant protein-linked heterogeneous viral DNA forms (6), which would otherwise obscure the cleavage pattern. *Acc*I cleavage of the remaining viral DNA in all cases revealed the presence of only one *Acc*I site; mapping of this site relative to the unique *Eco*RI and *Bgl*II sites of DHBV confirmed that the missing *Acc*I site was that in the precore region (representative results for one of the five mutants are shown in Fig. 2B).

This result rules out contamination of the input DNA with wild-type sequences or reversion of the mutant to the wild type during replication but does not exclude the occurrence of a second-site (compensatory) frameshift mutation in precore, which could restore the reading frame to the wild type. To examine this possibility, we digested the liver DNA of one mutant-infected animal with *Eco*RI. We recovered fragments of 3.0 to 3.5 kilobases (kb) from agarose gels and cloned these into lambda gtWES. Progeny plaques annealing to ^{32}P-labeled DHBV DNA sequences were picked, and a 1.3-kb *Xba*I fragment spanning the precore region was subcloned from one of these into pSP65 (10). In this plasmid vector, the sequence of the entire precore region was determined by the chain termination method. The original two-base frameshift mutation was still present in the recovered mutant DNA (Fig. 3); further sequencing (data not shown) revealed no additional changes between the precore ATG and the first out-of-phase terminator (at position 2631) following the frameshift lesion.

The presence of replicative forms in intrahepatic DNA strongly implies that infectious virus generated in the initially transfected cells successfully spread to uninfected hepatocytes (16, 17). To confirm this inference, we prepared a cell-free homogenate from one mutant-infected liver and used this material to inoculate four uninfected 1-day-old ducklings by subcutaneous inoculation in the thigh. Examination of liver DNA from these animals 21 days later revealed the presence of typical DHBV replicative forms in all four samples; restriction analysis of two of these samples performed as outlined in Fig. 2 again confirmed the presence of the mutation (data not shown). We then tested for the presence of infectious virus in the serum of these recipients. A pool of serum from three of these mutant-infected animals was inoculated subcutaneously into three ducklings (0.1 ml per recipient). All three developed viremia, as judged by dot hybridization of recipient serum.

These experiments demonstrate that the expression of the precore region of DHBV is not essential for the uptake, replication, assembly, or intrahepatic spread of the virus in experimental infection in vivo. However, it is possible that precore expression might affect the efficiency of titer of virus infection or contribute to more specialized biological functions, such as tropism for extrahepatic viscera (7) or competence for vertical transmission; these possibilities are

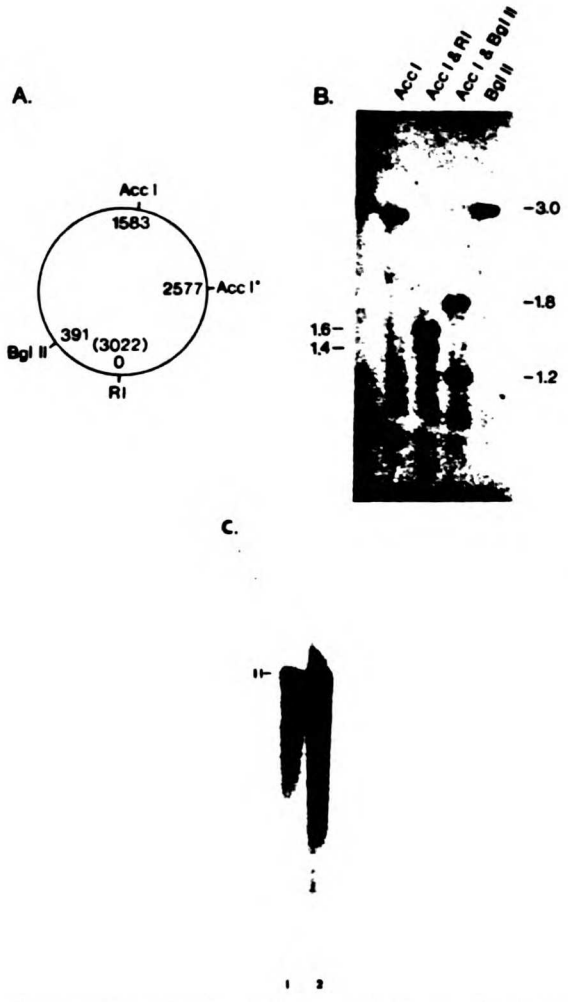


FIG. 2. Viral DNA forms in the livers of ducks transfected with mutant genomes. (A) Restriction map of relevant sites within DHBV DNA. Nucleotide positions are numbered with respect to the unique *Eco*RI site (8). *, *Acc*I site eliminated by the precore frameshift lesion. (B) Cleavage analysis of duplex DHBV DNA in the liver of animal 5 at 3 weeks posttransfection with pPc-fs DNA. Liver homogenates in 1% sodium dodecyl sulfate were extracted with phenol chloroform (1:1) without prior proteinase K treatment, as previously described (16). A 10-μg portion of the resulting relaxed circular DNA was digested with the indicated enzymes; the sample in the left lane was undigested. Product DNA was electrophoresed through 1% agarose, transferred to nitrocellulose, hybridized to ^{32}P-labeled DHBV DNA, and autoradiographed as previously described (16). The sizes of the fragments (in kilobases) are indicated. (C) Liver DNA prepared from animals 5 (lane 1) and 6 (lane 2) 3 weeks posttransfection with pPC-fs DNA. Liver homogenates were treated with 1% sodium dodecyl sulfate and 500 μg of proteinase K per ml for 3 h prior to phenol extraction; 10 μg of the resulting DNA was digested with *Pvu*II, which does not cleave within DHBV. Product DNA was electrophoresed through 1% agarose, transferred to nitrocellulose, and hybridized to ^{32}P-labeled DHBV DNA as in panel B. II, Position of fully duplex open circles of DHBV DNA.

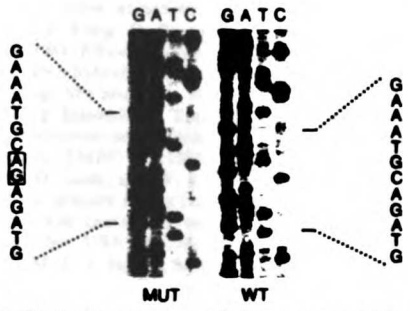


FIG. 3. Nucleotide sequence of the precore region of cloned DHBV DNA recovered from liver of a mutant-infected duck. A 1.3-kb *Xba*I fragment of DHBV DNA was excised from a lambda gt WES clone of the mutant viral DNA and subcloned into the *Xba*I site of pSP65 (10). A clone in which the precore region was adjacent to the SP6 promoter region was identified. The sequence of the entire precore region was determined by the dideoxy method by using a primer complementary to the SP6 promoter sequence. Left panel, a portion of the sequence showing the region spanning the lesion in recovered mutant (MUT) DNA; right panel, corresponding sequence of wild-type (WT) DHBV DNA. The sequences depicted are of minus-strand polarity.

currently being explored. The availability of viable precore mutant viruses will also make possible an assessment of the impact of precore proteins on viral persistence and on the pathogenic potential of DHBV.

We do not know at present whether these findings can be extrapolated to the mammalian viruses, including HBV. As noted previously, DHBV was chosen for these studies because of experimental tractability. However, DHBV is the most diverged of the hepadnaviruses at the sequence level (8, 18). Its 35-kilodalton core antigen is distinctly larger than those of the mammalian viruses (20 to 22 kilodaltons) and within the DHBV precore/core open reading frame, there is little amino acid homology to the cognate region of HBV (18). On the other hand, the DHBV precore region does contain a consensus signal peptidase recognition sequence (R. Colgrove, personal communication), and the virus does generate secretory precore products analogous to the e antigen of HBV. Additional experiments will be required to directly test whether the mammalian viruses require functional precore polypeptides for replication; we are currently constructing analogous mutations in the ground squirrel hepatitis virus precore region to examine this issue further.

Our findings with the DHBV precore gene recall an interesting parallel in murine retrovirology. The *gag* gene of Moloney murine leukemia virus encodes the nucleocapsid or core proteins of the virus, and its ATG initiator is preceded by an upstream initiator codon that is also expressed in vivo. Translation from the upstream initiator gives rise to *gag*-related proteins which enter the secretory pathway, are glycosylated, and are expressed on the cell surface and in the medium (2). Interestingly, as for DHBV, mutational ablation of these upstream sequences does not impair viral replication or infectivity (4, 14).

The experiments reported here represent the first demonstration of a nonessential region within the genome of a hepadnavirus. The existence of such a region is surprising given the small size (3.0 kb) and extremely compact coding organization of the genome; in DHBV (as in HBV), every nucleotide in the genome is in at least one coding region and 50% of the sequence is read in more than one frame. The fact

that the precore region is nonessential also suggests that insertions of exogenous DNA sequences into this region may be tolerated and raises the intriguing possibility that hepadnaviruses could be developed as genetic vectors for the delivery of foreign DNA to the liver.

These experiments were supported by Public Health Service grants (AI18782 and AM26743) from the National Institutes of Health to the authors and to the University of California at San Francisco Liver Center. H.E.V. is American Cancer Society Professor of Molecular Virology.

We thank Effie Meredith for outstanding assistance with the preparation of the manuscript.

LITERATURE CITED

1. **Buscher, M., W. Reiser, H. Will, and H. Schaffer.** 1985. Transcripts and the putative RNA pregenome of duck hepatitis B virus: implications for reverse transcription. Cell **40**:717–724.
2. **Edwards, S., and H. Fan.** 1980. Sequence relationship of glycosylated and unglycosylated *gag* polyproteins of Moloney murine leukemia virus. J. Virol. **35**:41–51.
3. **Enders, G. H., D. Ganem, and H. E. Varmus.** 1987. 5'-Terminal sequences influence the segregation of ground squirrel hepatitis virus RNAs into polyribosomes and viral core particles. J. Virol. **61**:35–41.
4. **Fan, H., H. Chute, E. Chao, and M. Feuerman.** 1983. Construction and characterization of Moloney murine leukemia virus mutants unable to synthesize glycosylated gag polyprotein. Proc. Natl. Acad. Sci. USA **80**:5965–5969.
5. **Ganem, D., and H. Varmus.** 1987. The molecular biology of the hepatitis B viruses. Annu. Rev. Biochem. **56**:651–695.
6. **Gerlich, W. H., and W. S. Robinson.** 1980. Hepatitis B virus contains protein attached to the 5' terminus of its complete DNA strand. Cell **21**:801–810.
7. **Halpern, M. S., J. M. England, D. T. Deery, D. J. Petcu, W. S. Mason, and K. L. Molnar-Kimber.** 1983. Viral nucleic acid synthesis and antigen accumulation in pancreas and kidney of Pekin ducks infected with duck hepatitis B virus. Proc. Natl. Acad. Sci. USA **80**:4865–4869.
8. **Mandart, E., A. Kay, and F. Galibert.** 1984. Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison with woodchuck and human hepatitis B virus sequences. J. Virol. **49**:782–792.
9. **McLachlan, A., D. Milich, A. K. Raney, M. G. Riggs, J. Hughes, J. Sorge, and F. V. Chisari.** 1987. Expression of hepatitis B virus surface and core antigens: influences of pre-S and precore sequences. J. Virol. **61**:683–692.
10. **Melton, D., P. Kreig, M. Rebogliati, T. Maniatis, K. Zinn, and M. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
11. **Moroy, T., J. Etiemble, C. Trepo, P. Tiollais, and M. Buendia.** 1985. Transcription of woodchuck hepatitis in the chronically infected liver. EMBO J. **4**:1507–1514.
12. **Ou, J.-H., O. Laub, and W. J. Rutter.** 1986. Hepatitis B gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. Proc. Natl. Acad. Sci. USA **83**:1578–1582.
13. **Roossinck, M. J., S. Jameel, S. H. Loukin, and A. Siddiqui.** 1986. Expression of hepatitis B viral core region in mammalian cells. Mol. Cell. Biol. **6**:1393–1400.
14. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1983. Deletion mutants of Moloney murine leukemia virus which lack glycosylated *gag* protein are replication competent. J. Virol. **46**:538–546.
15. **Seeger, C., D. Ganem, and H. E. Varmus.** 1986. Biochemical and genetic evidence for the hepatitis B virus replication strategy. Science **232**:477–484.
16. **Seeger, C., D. Ganem, and H. E. Varmus.** 1986. The cloned genome of ground squirrel hepatitis virus is infectious in the animal. Proc. Natl. Acad. Sci. USA **81**:5849–5852.
17. **Sprengel, R., C. Kuhn, C. Manso, and H. Will.** 1984. Cloned duck hepatitis B virus DNA is infectious in Pekin ducks. J. Virol. **52**:932–937.
18. **Sprengel, R., C. Kuhn, H. Will, and H. Schaller.** 1985. Comparative sequence analysis of duck and human hepatitis B virus genomes. J. Med. Virol. **15**:323–333.
19. **Summers, J.** 1981. Three recently described animal virus models for human hepatitis B virus. Hepatology **1**:179–183.
20. **Summers, J., and W. Mason.** 1982. Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. Cell **29**:403–415.
21. **Tiollais, P., C. Pourcel, and A. Dejean.** 1985. The hepatitis B virus. Nature (London) **317**:489–495.
22. **Will, H., R. Cattaneo, H. G. Koch, G. Darai, and H. Schaller.** 1982. Cloned HBV DNA causes hepatitis in chimpanzees. Nature (London) **299**:740–742.
23. **Will, H., R. Cattaneo, E. Pfaff, C. Kuhn, M. Roggendorf, and H. Schaller.** 1984. Expression of hepatitis B virus antigens with a simian virus 40 vector. J. Virol. **50**:335–342.
24. **Will, H., W. Reiser, T. Wiemer, E. Pfaff, M. Büscher, R. Sprengel, R. Cattaneo, and H. Schaller.** 1987. Replication strategy of human hepatitis B virus. J. Virol. **61**:904–911.