

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Hepadnaviral RNA packaging

**Permalink**

<https://escholarship.org/uc/item/51s7z65p>

**Author**

Pollack, Jonathan R.

**Publication Date**

1993

Peer reviewed|Thesis/dissertation

Hepadnaviral RNA packaging

by

Jonathan R. Pollack

**DISSERTATION**

**Submitted in partial satisfaction of the requirements for the degree of**

**DOCTOR OF PHILOSOPHY**

in

Biochemistry

in the

**GRADUATE DIVISION**

of the

**UNIVERSITY OF CALIFORNIA**

**San Francisco**



**To my wife, Anna.**

## ACKNOWLEDGMENTS

While composing this opus, I have had occasion to recall the many individuals to whom I am indebted. First, of course, I wish to thank my graduate advisor, Don Ganem. Throughout my years at the bench, he has offered constant encouragement, valuable ideas and constructive criticism. But more than this, D.G. has been a role model, for he is not only an extraordinarily gifted scientist, but has also expertly balanced the roles of teacher, clinician, and parent. I wish also to thank Dave Morgan and Judy White for their support and assistance as members of my thesis committee.

I thank the many members of the Ganem lab and fourth floor community, past and present. I thank Russell Hirsch for grandfathering the packaging field and adopting me. I am especially grateful to Dan Loeb, who generously spent his time teaching me how to do science. And I thank my tennis partners, Don Macrae, Chris Chang, Takashi Ishikawa, Keiji Ueda, and Ian Taylor for making non-lab hours so enjoyable. I also thank my colleagues in HSE 419, especially Steve Chuck, Chris Chang, Dean Kedes, and Sophie Roy, for their heartfelt, though at times restrained, admiration of my singing.

I thank my parents for their high standards and enthusiasm for scientific inquiry, which have directed me throughout my 26 years of formal education. I owe most, however, to my wife, Anna, and to our growing family, Libby, Mallory, and James, for they have made every moment worthwhile.

Chapter 2 of this thesis is a reproduction of the material as it appears in *Journal of Virology*. Chapter 4 is a manuscript submitted for publication. The coauthor listed in the publication and manuscript directed and supervised the research which forms the basis for these chapters.

## ABSTRACT

### HEPADNAVIRAL RNA PACKAGING

Hepadnaviruses are small enveloped DNA viruses that are hepatotropic and cause liver disease. Like retroviruses, hepadnaviruses replicate by a reverse transcription pathway in which the viral RNA pregenome is reverse transcribed to DNA. This reaction occurs within cytoplasmic subviral core particles; therefore, an important step in replication is the encapsidation of both the RNA pregenome and the viral polymerase.

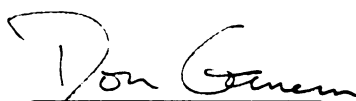
The RNA packaging signal has been previously localized to a short region at the 5' end of the RNA pregenome. RNA packaging is known to require both core and polymerase proteins, but which of these mediates recognition of the packaging signal has remained unknown. In this thesis I characterize the RNA packaging signal and define RNA-protein interactions involved in packaging.

Using nucleases to probe RNA secondary structure, I show that *in vitro* and *in vivo* the packaging signal forms an RNA stem-loop structure with a lower stem, a six nucleotide bulge, and an upper stem with a six nucleotide loop. The RNA structure and sequence requirements for packaging were then determined by transfecting mutant stem-loop constructs into cultured hepatoma cells and assaying for encapsidated RNA. Encapsidation requires the upper and lower stems, the bulge structure (but not its sequence), and specific sequences in the loop.

I next characterized polymerase-RNA interactions involved in packaging. Using a streptavidin-biotin mediated co-precipitation assay, I demonstrate that duck hepadnaviral polymerase produced in reticulocyte lysate binds specifically and with high affinity to the hepadnaviral packaging signal *in vitro*. By mutating the stem-loop region I demonstrate that binding requires recognition of the lower stem and bulge structures,

independently of sequence. Mutations that disrupt binding abolish RNA packaging *in vivo*. However, loop mutants that bind polymerase are also not encapsidated. I therefore conclude that polymerase binding to the packaging signal is necessary but not sufficient to initiate RNA packaging. I further demonstrate that the C terminus of polymerase, required for RNA packaging, is dispensable for RNA binding.

While both hepadnaviruses and retroviruses target their respective RNA genome and polymerase to virions, the mechanisms each evolved to accomplish this are distinct.

A handwritten signature in black ink that reads "Don Ganem". The signature is written in a cursive style with a horizontal line underneath the name.

**Don Ganem, Committee Chairperson**

## TABLE OF CONTENTS

Chapter 1: Introduction.....	1
Chapter 2: An RNA stem-loop structure directs hepatitis B virus genomic RNA encapsidation.....	24
Chapter 3: Use of tat-polymerase fusion constructs to explore the encapsidation function of HBV polymerase protein.....	58
Chapter 4: Duck hepatitis B virus polymerase binds an RNA stem-loop structure to initiate both viral RNA encapsidation and DNA synthesis.....	81
Chapter 5: Conclusion.....	110

## LIST OF TABLES

### Chapter 2:

Table 1. Stem-loop mutations.....	43
-----------------------------------	----

### Chapter 4:

Table 1. DHBV $\epsilon$ stem-loop mutations.....	101
---	-----



## LIST OF FIGURES AND ILLUSTRATIONS

### Chapter 1:

Fig. 1. HBV genome organization.....	15
Fig. 2. The hepadnaviral life cycle.....	16
Fig. 3. Hepadnaviral DNA synthesis.....	17

### Chapter 2:

Fig. 1. Phylogenetic analysis of the $\epsilon$ region.....	44
Fig. 2. The encapsidation assay.....	45
Fig. 3. Mapping the limits of $\epsilon$ .....	46
Fig. 4. Enzymatic probing of $\epsilon$ secondary structure in RNA synthesized <i>in vitro</i> .....	47
Fig. 5. Enzymatic probing of $\epsilon$ secondary structure in RNA from core particles.....	48
Fig. 6. Encapsidation assay of RNAs with lower stem mutations.....	49
Fig. 7. Encapsidation assay of RNAs bearing base changes in the upper stem.....	50
Fig. 8. Enzymatic structure probing of mutated RNAs generated <i>in vitro</i> .....	51
Fig. 9. Encapsidation assay of RNAs with nucleotide changes in the loop and bulge.....	52
Fig. 10. Structure-function relationships in the HBV encapsidation signal.....	53
Fig. 11 (addendum) Enzymatic probing of $\epsilon$ secondary structure at the 5' and 3' ends of RNAs synthesized <i>in vitro</i> .....	54
Fig. 12. (addendum) RNA binding of <i>E. coli</i> -produced HBV polymerase protein.....	55

### Chapter 3:

Fig. 1. Tat-polymerase transactivation of TAR and $\epsilon$ targets.....	73
Fig. 2. Tat-polymerase encapsidation of $\epsilon$ and TAR RNA.....	74

Fig. 3. Transactivation of TAR by tat-polymerase truncations.....	75
Fig. 4. $\epsilon$ RNA encapsidation directed by tat-polymerase truncations.....	76
Fig. 5. $\epsilon$ RNA encapsidation directed by doubly-mutated tat-polymerase.....	77
Fig. 6. $\epsilon$ RNA encapsidation directed by tat-pol.195 and doubly-mutated core donor construct.....	78

**Chapter 4:**

Fig. 1. DHBV polymerase-RNA binding assay.....	102
Fig. 2. RNA binding assay of DHBV $\epsilon$ stem-loop mutants.....	103
Fig. 3. Encapsidation assay of DHBV $\epsilon$ stem-loop mutants.....	104
Fig. 4. DNA priming mediated by DHBV $\epsilon$ stem-loop mutants.....	105
Fig. 5. RNA binding of C-terminally truncated polymerase proteins.....	106
Fig. 6. Polymerase- $\epsilon$ interaction initiates viral RNA encapsidation and DNA synthesis.....	107

**Chapter 5:**

Fig. 1. Comparison of packaging strategies.....	115
---	-----

## **CHAPTER 1**

## INTRODUCTION

### HEPADNA VIRUSES: AN OVERVIEW

The hepadnaviruses (for hepatotropic DNA viruses) are a family of small enveloped DNA animal viruses that share a common structure and liver tropism (28). The human hepatitis B virus is the prototype member and is a major cause of liver disease worldwide. Hepatitis B virus infection is acquired either perinatally (from the infected mother) or through horizontal spread (contact with infected body fluids). Infection can result in the clinical syndrome of acute hepatitis, though the virus is usually cleared rapidly. In 5-10% of infected individuals, however, a persistent infection is established that can lead to chronic liver inflammation, cirrhosis and liver failure (27). Since HBV is not a cytolytic virus it is believed the host immune response contributes to liver damage (12).

Of great clinical consequence, chronic carriers of HBV have a 100-fold increased risk of developing hepatocellular carcinoma compared to age-matched non-carriers (9). The mechanism(s) by which HBV contributes to hepatocellular carcinogenesis are not well understood (29).

A vaccine, composed of recombinantly-produced HBV envelope glycoprotein (74), is now available in some countries to prevent the establishment of HBV infection. However, worldwide, 200 million people are already infected with HBV (71), for which no cure exists. A more detailed understanding of the basic biology of viral replication and infection is essential for the development of future antiviral therapy.

Along with HBV, the hepadnaviral family includes the woodchuck hepatitis virus (WHV; 69), the ground squirrel hepatitis virus (GSHV; 46), the duck hepatitis B virus (DHBV; 47) and the heron hepatitis B virus (HHBV; 67). These viruses share with HBV a

common viral structure, genome organization, and viral replication strategy and provide non-human animal models for the study of hepadnaviral replication and disease.

## **VIRAL STRUCTURE**

The HBV virion (called the Dane particle after its discoverer; 21) is a 42 nanometer (nm) particle, and consists of a lipoprotein envelope surrounding a nucleocapsid that contains the circular DNA genome and viral polymerase. The viral envelope contains three viral glycoproteins L (large), M (middle), and S (small). The capsid, a 27 nm icosahedron with T=3 symmetry, is composed of 160 subunits of viral core protein (30). The hepadnaviral genome (Fig. 1), the smallest of any animal virus, is a 3.2 kilobase (kb), partially double-stranded, relaxed circular DNA molecule (held together by overlapping 5' ends; 68). Polymerase protein is covalently linked to the 5' end of minus strand DNA (4, 32) and an oligoribonucleotide is linked to the 5' end of plus strand DNA (41). This peculiar genomic organization arises from viral replication via a reverse transcription pathway (described below).

## **GENOME ORGANIZATION AND TRANSCRIPTION**

All hepadnaviral genomes contain three overlapping open reading frames (ORFs) templated by minus strand DNA (Fig. 1). ORF PreS1/PreS2/S encodes the L, M and S envelope glycoproteins. ORF PreC/C encodes precore (PreC), a secreted protein of uncertain function, and core (C), the major capsid protein. ORF P encodes the viral polymerase (P protein). A fourth ORF, X, is present only in mammalian hepadnaviral genomes, and appears to function as a viral transcriptional transactivator (20, 66).

HBV transcripts are produced by host RNA polymerase II (28). There are two major classes of transcripts, one subgenomic and one genomic in length (14; Fig. 1). Each class is marked by 5' end heterogeneity, and both classes share a common site of polyadenylation. Subgenomic RNAs (0.7, 2.1, 2.4 kb) encode X protein and envelope

glycoproteins. Genomic RNAs (3.5 kb) are overlength (larger than genome-length) and terminally redundant (owing to first-pass bypassing of polyadenylation signals; 60) and encode PreC, C, and P proteins.

The 5' end heterogeneity displayed by both subgenomic and genomic RNAs allows for the production of more than one protein per reading frame (Fig. 1). For example, the longer genomic RNAs (precore mRNA) encode precore protein, while the shorter genomic RNA begins just 3' of the precore translation start site, and so encodes core protein (which shares the C-terminal extent of PreC ).

## **HEPADNAVIRAL REPLICATION**

The seminal work of Summers and Mason (70), published in 1982, firmly established that hepadnaviruses, although containing DNA genomes, replicated by the reverse transcription of an RNA intermediate. Their conclusions were based on the asymmetric replication of minus and plus DNA strands, the actinomycin D (an inhibitor of DNA-templated polymerization) resistance of minus strand DNA synthesis, and on the detection of replication intermediates containing DNA/RNA hybrids.

By 1990, when I began my experiments in the laboratory, many of the basic elements of hepadnaviral replication were understood (28; Fig. 2 and Fig. 3). Upon entering the cell, the viral genome is delivered to the nucleus where it is transcribed into subgenomic and genomic RNAs. The shortest genomic length RNA serves a dual function: it is the mRNA for core and P protein translation, and it is also the RNA intermediate which templates reverse transcription (14, 22); this RNA is therefore called the RNA pregenome (pgRNA).

The RNA pregenome is encapsidated within cytoplasmic subviral core particles where DNA synthesis occurs. Direct repeats DR1 and DR2, found near the 5' ends of minus and plus DNA strands respectively, are important *cis*-signals for the initiation of DNA synthesis (41, 63, 78). Minus strand DNA synthesis, templated by pgRNA and

primed by P protein (4, 75), is traditionally viewed as progressing from the 3' copy of DR1. A result of P protein priming is the covalent linkage of P protein to minus strand DNA. Plus strand DNA synthesis initiates within DR2, is templated by minus strand DNA and primed by a short capped oligoribonucleotide (41) derived and translocated from the 5' end of pregenomic RNA (a product of RNAase H cleavage). Genome circularization occurs during plus strand synthesis.

Mature core particles containing DNA genomes bud into the endoplasmic reticulum (55) to acquire viral envelope glycoproteins. Mature virions are then exported from the cell via the constitutive vesicular transport pathway.

## **RNA ENCAPSIDATION: AN OVERVIEW**

A requisite step in the replication of hepadnaviruses is the encapsidation of the RNA pregenome and P protein into cytoplasmic subviral core particles. RNA packaging must be extremely selective: not only must the viral pgRNA be discriminated from subgenomic RNAs and precore mRNA (neither of which contain functional *cis*-signals for viral replication) but from the vast pool of cellular mRNAs. How is this selectivity accomplished?

The problem of selective RNA encapsidation has been solved by many different viruses. The mechanisms underlying RNA packaging are best understood in plus-stranded RNA viruses of plants [e.g. tobacco mosaic virus (26) and turnip crinkle virus (65)] and vertebrates [e.g. Sindbis virus (77)]. In these viruses, selective packaging results from direct capsid protein recognition of a specific region of the viral RNA genome (called the packaging signal). Because viral polymerase can be produced during the subsequent round of infection using packaged viral plus strand RNA as the mRNA for its translation, polymerase is not encapsidated within the nucleocapsid.

The encapsidation problem becomes more complex, however, for minus-stranded RNA viruses, double-stranded RNA viruses, and viruses replicating by reverse

transcription. For these viruses, not only must an RNA genome be selectively encapsidated, but an RNA-dependent polymerase must also be packaged in the virion to initiate genome replication on the next infectious cycle.

Why do viruses replicating by reverse transcriptase need to encapsidate a polymerase? Because in all examples studied [including retroviruses (3, 72), hepadnaviruses (70), plant caulimoviruses (57), yeast Ty elements (31) and *Drosophila Copia*-like elements (64)], reverse transcription does not occur freely in the cytoplasm, prior to RNA packaging, but occurs within capsid particles. Encapsidation of the reverse transcription reaction may serve to (1) separate RNA pregenomes destined for templating viral replication from those pregenomes translating viral proteins, (2) increase the efficiency of DNA synthesis (e.g. RNA may be structured in capsids to facilitate template switches during replication), and (3) avoid the possible deleterious effects of indiscriminately reverse transcribing cellular RNAs (25).

The mechanisms underlying the co-encapsidation of genomic RNA and polymerase have been studied extensively in retroviruses, and offer important insights into our understanding of hepadnaviral RNA/ P protein packaging. The basic retroviral genome is organized into three open reading frames: gag [a polypeptide precursor of matrix, capsid, nucleocapsid, and protease], pol (precursor of reverse transcriptase and integrase), and env (envelope glycoproteins). The integrated proviral genome, flanked by the viral long terminal repeats (LTRs), is transcribed into genomic (mRNA for gag, pol) and spliced subgenomic (mRNA for env) RNAs. Genomic RNA is selectively encapsidated and is the template for viral DNA synthesis.

Selective packaging of genomic RNA is mediated by packaging signals on the transcript. In Moloney murine leukemia virus (MoMLV), for example, the packaging signal, psi, encompasses a 350 nucleotide sequence located between U5 (part of the LTR) and the gag coding region of the genome (44, 45). This region is located 3' to the splice



donor site (i.e. within an intron) for subgenomic viral RNAs; therefore subgenomic RNAs do not contain psi and are not packaged. More recent evidence suggests that in MoMLV auxiliary packaging sequences found outside of psi, within U5 and gag, can increase the efficiency of packaging (1, 10, 51).

Most evidence suggests that the retroviral nucleocapsid (NC) portion of the gag precursor polypeptide (which is proteolytically cleaved during capsid maturation) mediates recognition of the packaging signal. NC protein contains an amino acid sequence, the "cys-his box" (39), which has strong similarity to the zinc finger motif (40) of various nucleic acid binding proteins. In MoMLV, mutations introduced in this amino acid sequence prevent encapsidation of viral RNA (33, 49). Purified NC from MoMLV (38) and Rous sarcoma virus (48) has non-specific RNA binding activity, and recently, the NC protein of human immunodeficiency virus type 1 (HIV-1) has been shown to specifically interact with HIV-1 RNA genome encapsidation signals (11).

In retroviruses, gag and pol are found in overlapping translational reading frames and pol is translated as a gag-pol fusion protein by ribosomal frameshifting (15, 36, 43) or stop codon suppression (80). Mature, enzymatically active pol is liberated from the gag-pol fusion protein only following encapsidation. Therefore, polymerase is targeted to the capsid by interactions between the gag domain of the gag-pol fusion and assembling gag subunits.

In retroviruses, the encapsidation of pol and the packaging of viral RNA, while both mediated by gag, are independent of each other. Capsids containing pol are produced in packaging signal-deficient helper cell lines (44), and pol-deficient mutants produce capsids that contain normal amounts of genomic RNA (79). [Interestingly, however, certain pol mutants in Rous sarcoma virus result in capsids lacking the characteristic set of host tRNAs (required to prime first strand DNA synthesis; 56).]

## **HEPADNAVIRAL RNA PACKAGING**

Our understanding of hepadnaviral RNA encapsidation has greatly advanced in the last 3-4 years. The problems faced by hepadnaviruses are similar to those faced by retroviruses: selective packaging of genomic RNA (exclusion of subgenomic RNAs) and the co-encapsidation of polymerase. However, experimental data suggests that basic mechanistic differences may underlie the packaging strategies developed by hepadnaviruses and retroviruses. Much of our current understanding of hepadnaviral packaging has evolved from experiments designed to identify the *cis*-packaging signals and the factors that mediate their recognition.

### **The Packaging Signal**

A *cis*-signal for packaging, the equivalent of the MLV psi sequence, has been localized for both HBV (19, 37) and DHBV (35). The HBV packaging signal, called  $\epsilon$ , has been localized to a 137 nucleotide stretch at the 5' end of pgRNA within the precore reading frame, but not including nearby DR1 sequences (37).  $\epsilon$  sequences are both necessary for encapsidation and, when fused at the 5' end, are sufficient to direct the packaging of a heterologous message (5, 35, 37).

$\epsilon$  sequences are also present at the 5' end of precore mRNA (which differs from pgRNA by a 30 nt 5' extension) but do not function to direct encapsidation of that message (23, 37). Recently it has been demonstrated that mutation of the precore ATG initiation codon results in the encapsidation of precore mRNA (52). Therefore it appears that ribosomes translating through the precore region render  $\epsilon$  non-functional (presumably by disrupting RNA structure).

$\epsilon$  sequences are not found at the 5' ends of subgenomic RNAs, which would appear to suggest a simple explanation why these RNAs are not packaged. However, because  $\epsilon$  sequences are present in the terminally redundant region of pgRNA (R region, Fig. 2) these sequences are also found at the 3' end of pgRNA, and at the 3' ends of all

subgenomic RNAs.  $\epsilon$  sequences, though, do not function to direct encapsidation from this 3' position (35, 37); the basis for this positional effect is not understood. [Psi is located outside of the terminally redundant region of retroviral genomic RNA and so is not present at the 3' end of retroviral subgenomic RNAs.]

$\epsilon$  RNA sequences have the potential to form an RNA stem-loop whose structure is conserved among hepadnaviruses (37). However the existence of this stem-loop structure *in vivo* and its role in packaging have remained unknown. In mammalian hepadnaviruses, the ATG initiation codon for core protein translation is located within  $\epsilon$  sequences, suggesting the possibility of the regulated use of pgRNA for core translation and packaging (37).

Unlike the case for HBV, the DHBV *cis*-packaging requirements appear to be dispersed over a surprisingly large region, encompassing the 5' 1200-1400 nucleotides of the pregenomic message (35). The important packaging signals within this RNA stretch appear to include RNA sequences at the 5' end of pgRNA (which correspond to the HBV  $\epsilon$  sequences) and a second region some 900 nucleotides downstream (Calvert and Summers, personal communication). This downstream region may play an auxiliary role in RNA packaging much as do regions outside of psi in retroviruses. Because this second downstream region is not present in subgenomic RNAs, this suggests one means employed by DHBV to exclude the packaging of these messages.

### **Recognition of the packaging signal**

Since RNA encapsidation is completed prior to enveloping of the nucleocapsid, viral glycoproteins are unlikely to take part in recognition of the packaging signal on pgRNA. Nor is hepadnaviral X protein likely to play a role, given it is not encoded in the avian hepadnaviral genome. Therefore, core protein and the viral polymerase are the two remaining possible viral encoded mediators of  $\epsilon$  recognition. In the past several years, much has been learned of the functions of these proteins.

**Core.** Core protein is a 21 kilodalton (kD) protein in HBV and a 35 kD protein in DHBV. Despite the molecular weight difference, HBV and DHBV core share extensive amino acid sequence homology and predicted tertiary folding (2). Recombinant HBV core expressed in a variety of heterologous systems [*E. coli* (13, 73), yeast (50), insect cells (8), *Xenopus* oocytes (82), and mammalian cells (54)] is able to self-assemble into authentic core particles (27 nm icosahedron). Experiments in the oocyte system reveal core protein dimers to be the initial assembly intermediate; these then rapidly aggregate into capsids (81).

Cores produced in these systems, when studied, appear to package RNA non-specifically; i.e. there is no enrichment for the encapsidation of  $\epsilon$ -containing transcripts (13). Core protein does not contain the "cys-his box" motif believed to participate in packaging-signal recognition in retroviruses and other retro-elements (25, 39). The distinguishing feature of core protein is its arginine-rich C terminus, which consists of four arginine-rich clusters of 3-4 arginines each. The arginine-rich C terminus of core is dispensable for capsid formation (8, 13). However, the non-specific encapsidation of RNA by cores produced in heterologous systems is dependent on the presence of these arginine residues (13). Recent experiments have also shown that deletion of the entire arginine-rich region likewise abolishes specific pgRNA encapsidation in transfected hepatoma cells in culture (53).

**P protein.** Polymerase is a 90 kD protein and has been divided into four domains based on function and amino acid sequence homologies (17, 58). From N to C terminus these four domains are terminal protein (covalently linked to the HBV genome during DNA priming), spacer (tolerates deletions), reverse transcriptase, and RNase H.

P protein is not abundant in virions, and has been historically very difficult to detect (7, 42). Recently, Bartenschlager and Schaller (6) have developed a sensitive assay for its

detection: By introducing a protein kinase A target site into the polymerase coding sequence, P protein isolated from virions is detected after *in vitro* <sup>32</sup>P-kinase labelling. Using this methodology, it has been determined that P protein in virions is indeed a 90 kD species (6), not processed by proteolytic cleavage as is retroviral pol. Knowing the specific activity of the kinase labelling, it has also been estimated that P protein is present in 1-2 copies per virion, roughly in stoichiometric amounts with pgRNA (6).

Hepadnaviral polymerase, even though encoded in the genome in an overlapping translational reading frame with core, is not translated as a core-pol fusion protein (16, 62), as is gag-pol in retroviruses. P protein is instead translated independently of core, likely by "leaky scanning" from the 5' end of pgRNA (24). This fundamental difference between the translation of polymerase in hepadnaviruses and retroviruses has profound consequences for polymerase encapsidation. Specifically, HBV cannot target its polymerase to capsids as does retroviral pol through covalent linkage to the capsid protein.

Suggestions of underlying mechanistic differences between hepadnaviral and retroviral packaging were extended by the recent finding that in both HBV (5) and DHBV (34) polymerase-deficient genomes transfected into cultured hepatoma cells produced core particles lacking pgRNA. Thus, unlike the case for retroviruses, viral RNA packaging is dependent on the expression of polymerase.

During the course of my thesis work, Bartenschlager and Schaller (6), using the *in vitro* P protein kinase-labelling assay, reported the inability to detect HBV P protein in core particles following the transfection of packaging signal-deficient viral genomes. Thus, and also unlike the case for retroviruses, packaging of polymerase is itself dependent on the expression, and co-encapsidation, of  $\epsilon$ -containing RNA.

Much has been learned of the P protein requirements for RNA packaging. P protein can be supplied *in trans* to direct encapsidation of a polymerase-deficient genome; however, there is a marked *cis*-preference for directing the encapsidation of the pgRNA from which it has been translated (5, 34). The efficiency of trans-complementation,

though, can be increased by deleting the packaging signals from the P protein donor mRNA (5). Together these results have been interpreted to mean that the encapsidation function of P protein occurs co-translationally, and/or that P protein is limiting for the packaging reaction and that its local concentration is highest near the pgRNA from which it has been translated.

The function of P protein in packaging is not dependent on its activities in DNA synthesis, since mutant P proteins that are inactive in polymerase and RNase H activities can still function to package RNA (5, 17, 34, 58). Deletion and missense mutations have been found (laboratory-introduced and in naturally-occurring hepadnaviral isolates) in the terminal protein, spacer, reverse transcriptase, and RNase H domains of P protein which abolish its encapsidation function (5, 18, 34, 59). This has been interpreted to mean that the entire polymerase polypeptide is required to mediate its packaging function.

Though much has been learned of the P protein requirements for encapsidation, little is known of its function in the encapsidation reaction. It is not known, for example, which viral protein, core or P protein, mediates the direct recognition of the packaging signal. Because the encapsidation of pgRNA and P protein are co-dependent (i.e. neither is packaged without the other), one attractive model is that P protein binds  $\epsilon$  to initiate encapsidation. Core recognition of a polymerase- $\epsilon$  complex would then ensure the co-encapsidation of both P protein and the viral RNA, solving the mechanistic problem of packaging a polymerase that is not translated as a fusion with the capsid protein. However it is also possible that core (perhaps modified by polymerase) or a cellular protein mediates the direct recognition of  $\epsilon$ .

Subsequent to its function in directing RNA packaging at the 5' end of pgRNA, P protein must polymerize minus-strand DNA synthesis from the 3' copy of DR1, some 3000 nucleotides downstream from  $\epsilon$  on pgRNA. Very recently during the course of my thesis work, though, it has been established that the first step of DNA synthesis occurs at

the 5' end of pgRNA (76; John Tavis and Don Ganem, personal communication). Polymerase acts as a protein primer for the covalent addition of a short (up to 4 nucleotides) DNA sequence whose residues are templated by RNA sequences at the 5' end of pgRNA, within the encapsidation signal (76). This short DNA sequence (linked to P protein) must then dissociate and reanneal to sequences within DR1 at the 3' end of pgRNA where minus strand DNA synthesis continues. Because DNA priming occurs within encapsidation sequences, a possible link between the RNA packaging and DNA priming reactions has been suggested (76).

## CONTENTS OF THE THESIS

When I began my hepadnaviral benchwork, I set out with the goal of characterizing the RNA-protein interactions involved in RNA packaging. I began my work with an extensive characterization of the HBV packaging signal. I chose the HBV packaging signal because it was contained within a much smaller region than the DHBV signal and therefore the protein-RNA interactions might be more straightforward to define. I first determined the RNA secondary structure of HBV  $\epsilon$  sequences *in vitro* and *in vivo* using nucleases specific for base-paired and unpaired regions of RNA. I then determined the structure and sequence requirements for RNA packaging by transfecting a panel of mutant  $\epsilon$  constructs into hepatoma cells in culture and assaying for encapsidated RNA. The results of these experiments are reported in Chapter 2.

I next set out to study the role of P protein in hepadnaviral packaging. Because an HBV polymerase protein I expressed in *E. coli* bound RNA only non-specifically, and because at the time there was no reported preparation of enzymatically active recombinantly-produced P protein, I set out to explore the packaging function of P protein *in vivo*. For this purpose, I fused HBV polymerase to the HIV-1 transactivator protein tat. If P protein interacts with  $\epsilon$  *in vivo*, I reasoned, then a tat-pol fusion might transactivate an

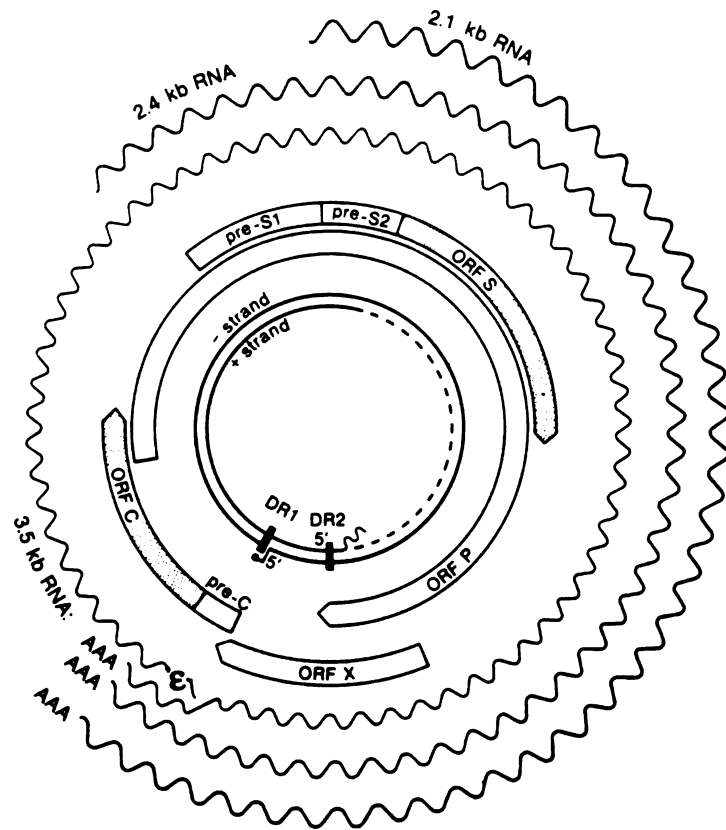
HIV promoter in which TAR (the RNA target of tat) had been replaced by  $\epsilon$ . The results of this and related experiments are reported in Chapter 3.

During the course of my thesis work the expression of enzymatically active DHBV polymerase in rabbit reticulocyte lysate was reported (75). Therefore, I turned my attention to studying P protein- $\epsilon$  interactions in DHBV. Using DHBV polymerase produced in reticulocyte lysate, I developed an assay to detect P protein- $\epsilon$  interactions *in vitro*, based on the streptavidin-biotin mediated co-precipitation of protein/RNA complexes. I then characterized the affinity and the specificity of this interaction. By introducing mutations into the DHBV  $\epsilon$  sequence, I also correlated the RNA structure and sequence requirements of binding with those of RNA packaging *in vivo* and DNA priming *in vitro*. I further began to localize an RNA binding domain of P protein. These results are reported in Chapter 4.

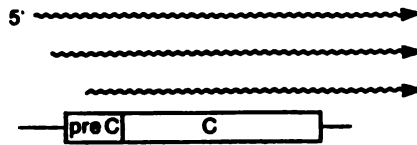
The results of my experiments support specific models for the co-encapsulation of P protein and viral RNA in hepadnaviruses. These models along with other implications of my results are discussed in Chapter 5.



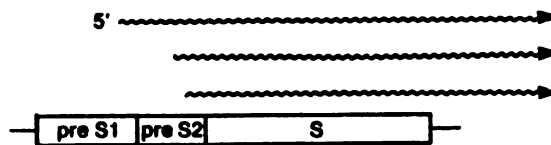
**Figure 1. HBV genome organization.** (Top) The innermost circle represents viral DNA; shown are the direct repeats (DRs), and the genome-linked P protein and capped oligoribonucleotide on minus and plus strands, respectively. The open reading frames encoded by the genome are indicated by the open boxes. The major viral transcripts, genomic (3.5 kb) and subgenomic (2.1 and 2.4 kb), are depicted as wavy lines; the encapsidation signal ( $\epsilon$ ) is shown. (Bottom) Structure of the 5' ends of viral transcripts. The 3.5 kb genomic RNAs initiate either upstream or downstream of the translation start site for preC, thereby encoding either precore or core proteins, respectively. Precore protein contains all the amino acids of core protein plus an N terminal extension. The analogous situation is diagrammed for the 2.1 kb subgenomic RNAs. Reproduced, with permission, from Ganem and Varmus (28).



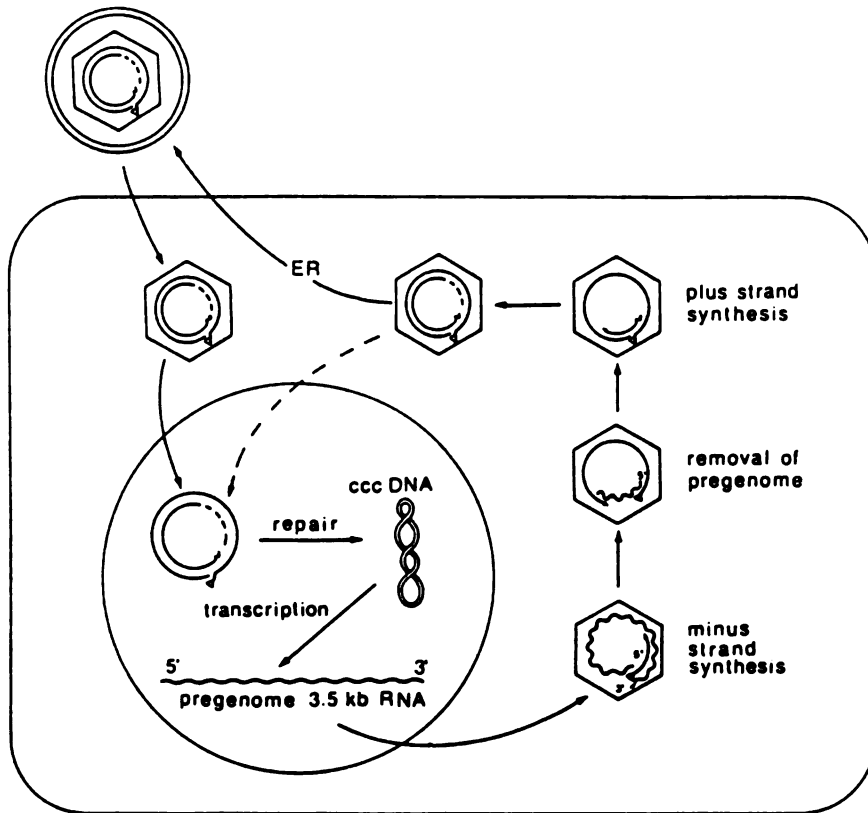
3.5 kb RNA:



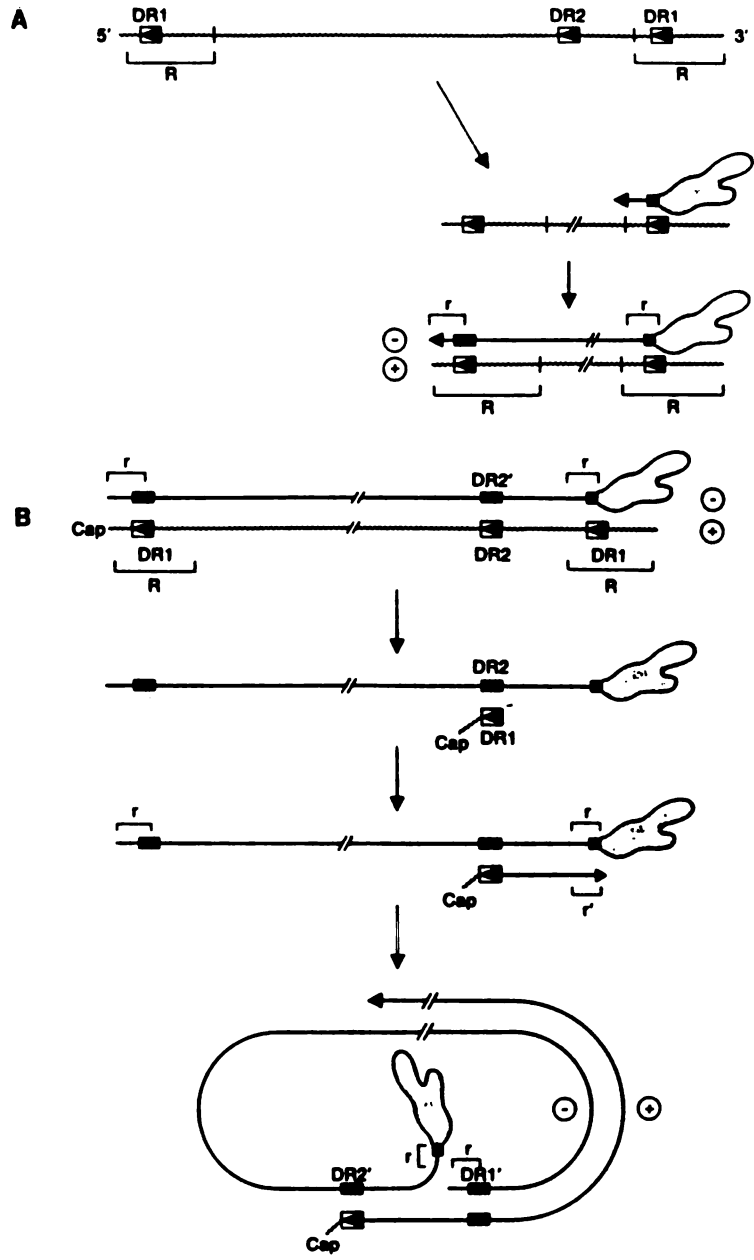
2.1 kb RNA:



**Figure 2. The hepadnaviral life cycle.** Schematic depiction of the sequential events of the hepadnaviral life cycle: hepatocyte entry, transport to the nucleus, genome repair, pgRNA transcription, RNA packaging, minus and plus strand DNA synthesis, envelope acquisition and viral export. For further details, see text. Reproduced, with permission, from Schlicht and Schaller (61).



**Figure 3. Hepadnaviral DNA synthesis.** (A) Minus strand DNA synthesis. P protein serves as a protein primer for the progression of minus strand DNA synthesis from the 3' copy of DR1. Wavy lines denote RNA; straight lines denote DNA. R is the long terminal repeat in pgRNA; r is the short terminal repeat in minus strand DNA. (B) Plus strand DNA synthesis. A short RNA containing DR1 sequence is cleaved from the 5' end of pgRNA and translocated to DR2 to prime plus strand DNA synthesis. Circularization occurs during plus strand synthesis by template transfer between r sequences as indicated. Reproduced, with permission, from Ganem and Varmus (28).



## REFERENCES

1. Adam, M.A. and A.D. Miller. 1988. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virology* 62: 3802-3806.
2. Argos, P. and S.D. Fuller. 1988. A model for the hepatitis B virus core protein: prediction of antigenic sites and relationship to RNA virus capsid proteins. *EMBO J.* 7: 819-824.
3. Baltimore, D. 1970. Viral RNA-dependent DNA polymerase. *Nature* 226:1209-1211.
4. Bartenschlager, R. and H. Schaller. 1988. The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *EMBO J.* 7: 4185-4192.
5. Bartenschlager, R., M. Junker-Niepmann, and H. Schaller. 1990. The P gene product of the hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J. Virol.* 64: 5324-5332.
6. Bartenschlager, R., and H. Schaller. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* 11: 3413-3420.
7. Bavand, M., M. Feitelson, and O. Laub. 1989. The hepatitis B virus-associated reverse transcriptase is encoded by the viral *pol* gene. *J. Virology* 63:1019-1021.
8. Beames, B. and R.E. Lanford. 1993. Carboxy-terminals of the HBV core protein affect capsid formation and the apparent size of the encapsidated HBV RNA. *Virology* 194:597-607.
9. Beasley, R. P., C.C. Lin, L.Y. Hwang, and C.S. Chien. 1981. Hepatocellular carcinoma and hepatitis B virus. *Lancet* 2:1129-1133.
10. Bender, M.A., T.D. Palmer, R.E. Gelinas, and A.D. Miller. 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region. *J. Virology* 61:1639-164.
11. Berkowitz, R.D., J. Luban, and S.P. Goff. 1993. Specific binding of the human immunodeficiency virus type 1 *gag* polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virology* 71:7190-7200.
12. Bianchi, L. and F. Gudat. 1979. Immunopathology of Hepatitis B. *Prog. Liver Dis.* 6:371-392.
13. Birnbaum, F. and M. Nassal. 1990. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virology* 64: 3319-3330.
14. Buscher, M.W., 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.

15. Chamorro, M., N. Parken, and H.E. Varmus. 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *PNAS* 89: 713-717.
16. Chang, L.-J., P. Pryciak, D. Ganem, and H.E. Varmus. 1989. Biosynthesis of the reverse transcriptase of hepatitis B viruses involves *de novo* translational initiation not ribosomal frameshifting. *Nature* 337: 364-368.
17. Chang, L.-J., R. Hirsch, D. Ganem, and H.E. Varmus. 1990. Effects of insertional and point mutations on the functions of the duck hepatitis B virus polymerase. *J. Virology* 64: 5553-5558.
18. Chen, Y., W.S. Robinson, and P.L. Marion. 1992. Naturally occurring point mutation in the C terminus of the polymerase gene prevents duck hepatitis B virus RNA packaging. *J. Virol.* 66: 1282-7.
19. Chiang, P.-W., K.-S. Jeng, C.-P. Hu, and C. Chang. 1992. Characterization of a *cis* element required for packaging and replication of the human hepatitis B virus. *Virology* 186: 701-711.
20. Colgrove, R., G. Simon, and D. Ganem. 1989. Transcriptional activation of homologous and heterologous genes by the hepatitis B virus X gene product in cells permissive for viral replication. *J. Virology* 63: 4019-4026.
21. Dane, D.S., C.H. Cameron, and M. Briggs. 1970. Virus-like particles in serum of patients with Australia antigen associated hepatitis. *Lancet* 1: 695-698.
22. Enders, G. H., D. Ganem, and H.E. Varmus. 1985. Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. *Cell* 42: 297-308.
23. 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.
24. Fouillot, N., S. Tlouzeau, J.M. Rossignol, and O. Jean-Jean. 1993. Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation. *J. Virol.* 67: 4886-4895.
25. Fuetterer, J. and T. Hohn. 1987. Involvement of nucleocapsids in reverse transcription: a general phenomenon? *TIBS* 12: 92-95.
26. Gaddipati, J.P., C.D. Atreya, D. Rochon, and A. Siegel. 1988. Characterization of the TMV encapsidation initiation site on 18s rRNA. *Nuc. Acids Res.* 16: 7303-7313.
27. Ganem, D. 1982. Persistent infection of humans with hepatitis B virus: mechanisms and consequences. *Rev. Infect. Dis.* 4:1026-1047.
28. Ganem, D. and H.E. Varmus. 1987. The molecular biology of the hepatitis B viruses. *Ann. Rev. Biochem.* 56: 651-694.



29. Ganem, D. 1990. Oncogenic viruses: Of marmots and men. *Nature* 347: 320-232.
30. Ganem, D. 1991. Assembly of hepadnaviral virions and subviral particles. *Curr. Top. Microbiol. Immunol.* 168:61-82.
31. Garfinkel, D.J., J.D. Boeke, and G.R. Fink. 1985. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* 42: 507-515.
32. 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: 801-809.
33. Gorelick, R.J., L.E. Henderson, J.P. Hanser, and A. Rein. 1988. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *PNAS* 85: 8420-8424.
34. Hirsch, R., J. Lavine, L. Chang, H. Varmus, and D. Ganem. 1990. Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* 344: 552-555.
35. Hirsch, R., D.D. Loeb, J.R. Pollack, and D. Ganem. 1991. *Cis* - Acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. *J. Virology* 65:3309-3316.
36. Jacks, T. and H. Varmus. 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science* 230: 1237-124.
37. Junker-Niepmann, M., R. Bartenschlager, and H. Schaller. 1990. A short *cis*-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.* 9: 3389-3396.
38. Karpel, R.L., L.E. Henderson, and S. Oroszlan. 1987. Interactions of retroviral structural proteins with single stranded nucleic acids. *J. Biol. Chem.* 262: 4961-4967.
39. Katz, R.A. and J.E. Jentoft. 1989. What is the role of the cys-his motif in retroviral nucleocapsid (NC) proteins? *BioEssays* 11: 176-181.
40. Klug, A. and D. Rhodes. 1987. 'Zinc fingers': a novel protein motif for nucleic acid recognition. *TIBS* 12:464-469.
41. Lien, J.M., C.E. Aldrich, and W.S. Mason. 1986. Evidence that a capped oligonucleotide is the primer for duck hepatitis B virus plus-strand DNA synthesis. *J. Virology* 57: 229-236.
42. Mack, D.H., W. Bloch, N. Nath, and J.J. Sninsky. 1988. Hepatitis B virus particles contain a polypeptide encoded by the largest open reading frame: a putative reverse transcriptase. *J. Virology* 62: 4786-4790.
43. Madhani, H.D., T. Jacks, and H.E. Varmus. 1988. Signals for the expression of the HIV *pol* gene by ribosomal frameshifting. In: Franza, R., B. Cullen, and F. Wong-Stall (eds) *Control of HIV gene expression*. Cold Spring Harbor Lab. pp 119-125.

44. Mann, R., R.C. Mulligan, and D. Baltimore. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* 33: 153-159.
45. 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. Virology* 54: 401-407.
46. Marion, P.L., L.S. Oshiro, D.C. Regnery, G.H. Scullard, and W.S. Robinson. 1980. A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. *PNAS* 77: 2941-2945.
47. Mason, W.S., G. Seal, and J. Summers. 1980. Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *J. Virology* 36: 829-836.
48. Meric, C., J.L. Darlix, and P.F. Spahr. 1984. It is Rous sarcoma virus protein P12 and not P19 that binds tightly to Rous sarcoma virus RNA. *J. Mol Biol* 173: 531-538.
49. Meric, C. and S.P. Goff. 1989. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the cys-his box of the nucleocapsid protein. *J. Virology* 63:1 558-1568.
50. Miyanohara, A., T. Imamura, M. Araki, K. Sugawara, N. Ohtomo, and K. Masubara. 1986. Expression of hepatitis B virus core antigen gene in *Saccharomyces cerevisiae*: synthesis of two polypeptides translated from different initiation codons. *J. Virology* 59: 176-180.
51. Murphy, J.E. and S.P. Goff. 1989. Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: effects on RNA packaging and reverse transcription. *J. Virology* 64: 319-327.
52. Nassal, M., M. Junker-Niepmann, and H. Schaller. 1990. Translational inactivation of RNA function: discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell* 63:1357-1363.
53. Nassal, M. 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J. Virology* 66: 4107-4116.
54. Ou, J.H., O. Laub, and W. J. Rutter. 1986. Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. *PNAS* 83: 1578-1582.
55. Patzer, E.J., G.R. Nakamura, C.C. Simonsen, A.D. Levinson, and R. Brands. 1986. Intracellular assembly and packaging of hepatitis B surface antigen particles occur in the endoplasmic reticulum. *J. Virology* 58: 884-892.
56. Peters, G. and J. Hu. 1980. Reverse transcriptase as the major determinant for selective packaging of tRNA's into avian sarcoma virus particles. *J. Virology* 36:692-700.
57. Pfeiffer, P. and T. Hohn. 1983. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* 33: 781-789.

58. Radziwill, G., W. Tucker, and H. Schaller. Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. 1990. *J. Virology* 64: 613-620.
59. Roychoudhury, A., F. Faruqi, and C. Shih. 1991. Pregenomic RNA encapsidation analysis of eleven missense and nonsense polymerase mutants of human hepatitis B virus. *J. Virology*, 65: 3617-3624.
60. Russnak, R. and D. Ganem. 1990. Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes & Development* 4: 764-776.
61. Schlicht, H.-J. and H. Schaller. 1989. Analysis of hepatitis B virus gene functions in tissue culture and *in vivo*. *Curr. Top. Microbiol. Immunol.* 144: 253-263.
62. Schlicht, H.-J., G. Radziwill, and H. Schaller. 1989. Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core-polymerase fusion proteins. *Cell* 56: 85-92.
63. Seeger, C.D., D. Ganem, and H.E. Varmus. 1986. Biochemical and genetic evidence for the hepatitis B virus replication strategy. *Science* 232: 477-484.
64. Shiba, T. and K. Saigo. 1983. Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. *Nature* 302: 119-124.
65. Sorger, P.K., P.G. Stockley, and S.C. Harrison. 1986. Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly *in vitro*. *J. Mol. Biol.* 191: 639-658.
66. Spandau, D.F. and C.H. Lee. 1988. Trans activation of viral enhancers by the hepatitis B virus X protein. *J. Virology* 62: 427-434.
67. Sprengel, R., E.F. Kaleta, and H. Will. 1988. Isolation and characterization of a hepatitis B virus endemic in herons. *J. Virology* 62: 3832-3839.
68. Summers, J., A. O'Connell, and I. Millman. 1975. Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. *PNAS* 72: 4597-4601.
69. Summers, J.J., J.M. Smolec, and R. Snyder. 1978. A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *PNAS* 75: 4533-4537.
70. Summers, J. and W. S. Mason. 1982. Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* 29:403-415.
71. Szmuness, W. 1978. Hepatocellular carcinoma and the hepatitis B virus: evidence for a causal association. *Prog. Med. Virol.* 24: 40-69.
72. Temin, H. and S. Mizutani. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226: 1211-1213.

73. Uy, A., V. Bruss, W.H. Gerlich, H.G. Kochel, and R. Thomssen. 1986. Precore sequence of hepatitis B virus inducing e antigen and membrane association of the viral core protein. *Virology* 155: 89-96.
74. Valenzuela, P., A. Medina, W.J. Rutter, G. Ammerer, and B.D. Hall. 1982. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* 298: 347-350.
75. Wang, G.-H., and C. Seeger. 1992. The reverse transcriptase of hepatitis B virus acts as a protein primer for viral DNA synthesis. *Cell* 71: 663-670.
76. Wang, G.-W. and C. Seeger. 1993. Novel mechanism for reverse transcription in hepatitis B viruses. *J. Virology* 67:6507-6512.
77. Weiss, B., H. Nitschko, I. Ghattas, R. Wright, and S. Schlesinger. 1989. Evidence for specificity in the encapsidation of sandbis virus RNAs. *J. Virology* 63: 5310-5318.
78. Will, H., W. Reiser, T. Weimer, E. Pfaff, M. Buscher, R. Sprengel, R. Cattaneo, and H. Schaller. 1987. Replication strategy of human hepatitis B virus. *J. Virology* 61: 904-911.
79. Yoshimura, F.K. and J.M. Yamamura. 1981. Four Moloney murine leukemia virus-infected rat cell clones producing replication-defective particles: protein and nucleic acid analyses. *J. Virology* 38: 895-905.
80. Yoshinaka, Y., I. Katoh, T.D. Copeland, and S. Oroszlan. 1985. Murine leukemia virus protease if encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *PNAS* 82: 1618-1622.
81. Zhou, S. and D.N. Standring. 1992. Hepatitis B virus capsid particles are assembled from core-protein dimer precursors. *PNAS* 89: 10046-10050.
82. Zhou, S., S.Q. Yang, and D.N. Standring. 1992. Characterization of hepatitis B virus capsid particle assembly in *Xenopus* oocytes. *J. Virology* 66: 3086-3092.

## **CHAPTER 2**

**AN RNA STEM-LOOP STRUCTURE DIRECTS HEPATITIS B VIRUS  
GENOMIC RNA ENCAPSIDATION**

**Jonathan R. Pollack<sup>1</sup> and Don Ganem<sup>2,3</sup>**

**Department of Biochemistry and Biophysics<sup>1</sup>, Howard Hughes Medical Institute<sup>2</sup>, and  
Departments of Microbiology and Immunology and Medicine<sup>3</sup>, University of California  
Medical Center, San Francisco, California 94143-0502.**

## **ABSTRACT**

Selective encapsidation of hepatitis B virus (HBV) genomic RNA within cytoplasmic core particles requires recognition of the *cis*-encapsidation signal, (termed  $\epsilon$ ) located at the 5' end of genomic RNA. By transfecting plasmids expressing chimeric RNAs bearing HBV sequences fused to LacZ, we have mapped the minimal region of  $\epsilon$  to the 5' 94 nts of genomic RNA. Enzymatic probing of the RNA secondary structure in this region (using either *in vitro* transcripts or RNA extracted from HBV core particles) reveals a stem-loop structure containing a lower stem, a 6 nt bulge, an upper stem with a single unpaired U residue and a 6 nt loop. The functional role of this structure in encapsidation was explored by examining the effects of mutations in  $\epsilon$  on encapsidation of RNA *in vivo*. These studies reveal that (i) in the lower stem, base-pairing but not specific primary sequence is required for function; (ii) there is no requirement for base-pairing in the lower portion of the upper stem, but base-pairing elsewhere in this stem contributes to packaging efficiency; (iii) the presence of the 6 nt bulge, but not its primary sequence, is important for function; and (iv) specific nucleotide sequences in the loop and in regions of the upper stem are critical for RNA encapsidation.

## INTRODUCTION

Hepadnaviruses are small hepatotropic viruses that cause acute and chronic hepatitis and strongly predispose to the development of hepatocellular carcinoma (11). These viruses include the prototype human hepatitis B virus (HBV) and several related viruses of mammals and birds. All hepadnaviruses contain 3-3.3 kb. DNA genomes whose replication proceeds by reverse transcription of an RNA intermediate (26). Reverse transcription occurs principally within subviral particles (or cores) composed of the viral nucleocapsid protein (C), the polymerase (P) and the RNA template for the reaction. The latter is also called pregenomic RNA (pgRNA), in view of its role as the precursor to the DNA form of the genome. Thus, these DNA viruses encapsidate their genome as RNA, and this encapsidation reaction is an essential early step in the pathway of genomic replication.

The RNA packaging reaction displays remarkable selectivity for pgRNA. HBV-infected cells produce a variety of viral transcripts of both genomic and subgenomic length. The genomic RNAs include the true pgRNA as well as minor additional transcripts with start sites up to 31 nts 5' to the pgRNA cap site (9, 28). These RNAs, otherwise identical in structure to pgRNA, are not encapsidated (10); they function only as mRNAs for a secreted variant of the C protein known as preC. The subgenomic mRNAs (of 2.4, 2.1 and possibly 0.7 kb) are 3' coterminal with pgRNA but also are not packaged into cores (10).

Selective encapsidation of pg RNA is now known to be dependent on both C and P gene product(s) (2,14,22,23) and on a *cis*-acting encapsidation signal, termed  $\epsilon$ , on pg RNA. This site has been localized to an 85-137 nt. stretch near the 5' end of the HBV pg RNA transcript (2,16), while a much larger region is required to direct efficient pgRNA encapsidation in the avian hepadnaviruses (15). The  $\epsilon$  sequences are also present on preC



mRNAs, but ribosomes translating the preC region (which overlaps with  $\epsilon$ ) apparently preclude their functioning to direct encapsidation of these transcripts (21). Because of the terminal redundancy of the pg RNA transcript, and the coterminal nature of all viral transcripts, the  $\epsilon$  sequences are also present at the 3' ends of pg RNA and subgenomic RNAs, but do not direct encapsidation from this position (15); the basis for this positional effect is not understood.

While computer-based secondary structure prediction and phylogenetic analyses have suggested potential secondary structures in the  $\epsilon$  region (16), the existence of such structures has not been demonstrated and their potential roles in encapsidation are unknown. In this paper, we further define the minimal  $\epsilon$  element, determine the secondary structure of this RNA region, and explore the functional role of this RNA structure in encapsidation *in vivo*.

## **MATERIALS AND METHODS**

**Materials.** Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions. RNase T<sub>1</sub> and RNase T<sub>2</sub> were purchased from BRL, RNase CV<sub>1</sub> was purchased from Pharmacia, and AMV reverse transcriptase was purchased from Seikagaku America, Inc. Radionucleotides were purchased from Amersham Corp.

**Plasmids.** All HBV nucleotide positions are numbered from the unique EcoRI site of HBV adw2 (27). In this numbering scheme, nt 1815 is the transcription initiation site of pg RNA.

Helper plasmid pCMV-CP was constructed by inserting HBV sequences from the StyI site (nt 1880) to the TaqI site (nt 2012) into the polylinker of pCDNA1 (Invitrogen),

downstream of the CMV-IE promoter. pCMV-CP contains the env(-) termination mutation in the S gene (6) and therefore does not express HBV envelope proteins.

pE-BS was constructed by cloning the 172 nt polymerase chain reaction (PCR)-amplified fragment containing nts 1815 to 1986 of HBV (using linearized HBV monomer as a template) into the HindIII and BamHI sites of pBS(-) (Stratagene). pLacZ was constructed from pON249 (12) by removing a 364 nt PvuII fragment within the LacZ gene so the resultant CMV promoter-driven LacZ transcript approximates the length of HBV pg RNA. The HindIII-KpnI (KpnI site in polylinker) HBV fragment of pE-BS was inserted to replace the HindIII-KpnI fragment at the 5' end of LacZ in pLacZ to generate pE-LacZ. In pE-LacZ, transcription from the CMV-IE promoter initiates at the first nt of the HBV sequence.

The other HBV-LacZ fusions, used in mapping the limits of  $\epsilon$  (Fig. 3), were created by cloning the appropriate PCR-amplified fragment into the HindIII-KpnI sites of pLacZ. pb-LacZ, pc-LacZ, pd-LacZ, pe-LacZ, and pf-LacZ contain HBV nts 1842-1986, 1842-1927, 1829-1986, 1815-1927, and 1815-1908, respectively. pDHBV-LacZ was created by cloning the PCR-amplified fragment containing the 5' 172 nts of DHBV3 (25) pg RNA into the HindIII-KpnI sites of pLacZ.

Small deletions and all nucleotide changes in the stem-loop structure sequences, listed in Table 1, were constructed by site-directed mutagenesis by standard methods (18), by first introducing the mutation into pE-BS. After confirmation by DNA sequencing, the HindIII-KpnI fragment containing the mutated stem-loop sequence was then cloned into the HindIII-KpnI sites of pLacZ.

**Computer Prediction of Secondary Structure.** RNA secondary structure prediction for wild-type and mutant constructs was performed using the software of Martinez (20) and Abrahams et. al. (1).

***In Vitro* Structure Determination.** RNA was transcribed *in vitro* from SmaI-linearized pE-BS template using T3 RNA polymerase. The resultant RNA contains the 5' 172 nts of HBV genome plus 14 nts of polylinker upstream and 2 nts of polylinker downstream. The RNA was gel purified by electrophoresis on a 4% w/v (3%NuSieve, 1%SeaPlaque [FMC BioProducts]) agarose gel in Tris-Acetate buffer and recovered by phenol extraction. For structure mapping of mutated RNAs, the relevant mutation introduced into pE-BS was used for template.

RNA renaturation and enzymatic digestion with RNase T<sub>1</sub>, RNase T<sub>2</sub> and RNase CV<sub>1</sub> were performed exactly as per Skinner et. al. (24). The RNA from one fifth of each digestion reaction was used for the primer extension reaction, performed exactly as described by Black and Pinto (5) using a [ $\gamma$ -<sup>32</sup>P]ATP labeled oligonucleotide primer (5'-GAAGTCAGAAGGCAAAAACG-3') complementary to nts at the 3' end of the RNA. Primer extension products were then separated on a 6% denaturing polyacrylamide gel and detected by autoradiography.

For the enzymatic structure determination of RNA encapsidated within core particles, RNA extracted from core particles purified from 1/2 of a 100 mm plate was used for enzymatic digestion as described above.

**Cell Culture and Transfections.** HepG2 human hepatoma cells were grown in HME16 medium supplemented with 10% fetal calf serum, 0.14% sodium bicarbonate and 2 mM glutamine, and passaged every 2-3 days at a 1:3 dilution. DNA transfections were performed by the calcium phosphate coprecipitation method exactly as described previously (13).

**RNA preparation and RNase protection assay.** PolyA<sup>+</sup> total cellular RNA was purified 72 hrs post-transfection as previously described (13,19). RNA within cytoplasmic core particles was isolated 72 hrs post-transfection as previously described (13,19) by

polyethylene glycol precipitation of core particles followed by proteinase K digestion, phenol extraction and precipitation of nucleic acid. RNase protection analysis was performed as previously described (14) on polyA+ RNA or core RNA prepared from equivalent numbers (1/2 of a 100 mm plate) of transfected HepG2 cells. Synthesis of [ $\alpha$ - $^{32}$ P]CTP-labeled RNA probes was carried out as previously described (14); Probe lacZ-P108 (15) is a 475 nt riboprobe with about 50 nonhybridizing polylinker nts and 425 nts of LacZ complementary to the LacZ fragment contained in pLacZ. Intensity of RNase-protected bands was quantitated by phosphorimager (Molecular Dynamics). At least two independent transfections and RNase protection assays were performed for each mutant construct tested.

## RESULTS

**Computer and phylogenetic analysis of the  $\epsilon$  region.** Computer-based secondary structure prediction suggests several potential secondary structures within the region of the HBV encapsidation signal,  $\epsilon$ . One plausible structure, first pointed out by Junker-Niepmann et al. (16) and shown in Fig. 1, is supported by phylogenetic comparisons. As shown, despite significantly divergent primary sequences, grossly similar secondary structures can be drawn for both the mammalian and avian hepadnaviruses. Note, however, that the upper stem is less well conserved within the avian grouping. In the heron hepatitis B virus (HHBV), as indicated by the nucleotide substitutions in Fig. 1, and in two infectious duck hepatitis B virus (DHBV) isolates (DHBV-S5 and DHBV-S31; not shown), nucleotide changes are found which in each case disrupt three potential base pairs within the upper stem. These findings emphasize the need for direct experimental characterization of the RNA structure in this region and its relation to RNA packaging *in vivo*.

**Encapsidation of heterologous RNA directed by HBV  $\epsilon$  region.** To study the function of the packaging signal *in vivo*, we employed an assay based on the ability of  $\epsilon$  sequences to mediate the encapsidation of heterologous RNAs. HepG2 hepatoma cells were transfected with plasmid pE-LacZ (Fig. 2a) which expresses a chimeric RNA of genomic length bearing the 5' 172 nts of HBV pg RNA (which should contain the entire  $\epsilon$  region [16]) fused to LacZ. This RNA is encapsidated by C and P proteins supplied *in trans* from helper plasmid pCMV-CP (Fig. 2a), which does not contain  $\epsilon$  and which donates P protein efficiently (2). Following transfection, we harvested from equivalent numbers of transfected cells either the total cellular polyA<sup>+</sup> RNA or the RNA contained within purified cytoplasmic core particles. The RNA present in equal portions of each preparation was quantified by RNase protection using a riboprobe derived from LacZ

sequences; the ratio of encapsidated to total polyA+ RNA is a measure of the packaging efficiency.

Fig. 2b, lanes 4 and 5 show the results from a cotransfection of the helper plasmid and plasmid pLacZ, which expresses only LacZ sequences. The LacZ RNA accumulates in the cytoplasm, as we find a protected band of the expected size (425 nts) in the sample containing polyA+ RNA (lane 5). However, this RNA is not encapsidated, as no protected band is found in the sample of RNA extracted from core particles (lane 4). By contrast, transfection of the chimeric plasmid pE-LacZ bearing the  $\epsilon$  region fused to LacZ results in efficient RNA encapsidation within cores (lanes 6, 7). The efficiency of encapsidation (ca. 50% of the total RNA) is similar to that of wild-type HBV pgRNA, indicating that all the relevant *cis*-acting information is contained within these 172 nts of HBV.

**Mapping the Limits of  $\epsilon$ .** To better define the limits of the  $\epsilon$  region, and in particular to determine whether HBV sequences containing only the proposed stem-loop structure are sufficient to direct encapsidation, we generated the constructs shown in Fig. 3a. These constructs were cotransfected with helper plasmid pCMV-CP, and levels of encapsidated RNA determined by RNase protection (Fig. 3b). As before, the 5' 172 nt region directs efficient packaging of LacZ RNA sequences (pE-LacZ [construct a]; lanes 3, 4), while the RNA with LacZ alone is not encapsidated (lanes 1, 2). Among these constructs, construct f (lanes 13, 14) represents the smallest HBV region sufficient to direct RNA encapsidation with nearly wild-type efficiency. Thus,  $\epsilon$  is contained within the 5' 94 nts of pg RNA found in construct f; this includes the proposed stem-loop structure and the 30 nts upstream of it. Constructs b, c and d (lanes 5-10) harbor deletions of varying extent upstream of the stem-loop and are packaged 5-10 fold less efficiently, indicating that sequences in this region, while not absolutely essential for encapsidation, contribute measurably to its efficiency.

Since the proposed stem-loop structure is conserved phylogenetically, we sought to determine whether a similar structure with a different primary sequence could still be packaged by HBV C and P proteins. We therefore tested construct g (pDHBV-LacZ), which contains the putative DHBV stem-loop structure fused to LacZ RNA. As shown in Fig. 3b (lanes 15, 16), this RNA is not encapsidated by HBV C and P proteins.

**Probing the structure of  $\epsilon$  in RNA synthesized *in vitro*.** To obtain direct biochemical evidence for secondary structure within  $\epsilon$ , we probed the structure of this region in RNA transcripts generated *in vitro*. A 188 nt RNA containing the  $\epsilon$  region was transcribed *in vitro*, self-annealed, and then subjected to cleavage by RNases specific for either single-stranded or base-paired regions. The positions of cleavage sites were then determined by primer extension, using an end-labelled primer homologous to sequences at the 3' end of the RNA. Fig. 4b (lanes 1-4) shows typical results obtained by cleavage with RNase T<sub>2</sub>, an RNase specific for single-stranded regions (8); products of primer extension on these cleaved RNAs are run next to the corresponding sequencing ladder generated from the same primer (lanes 5-8). Increasing the concentration of the enzyme yields two sets of prominent bands (lanes 1-3) not found in the absence of enzyme (lane 4), representing positions of single-stranded nucleotides. A single structure is accommodated by this information; this structure is identical to that predicted by phylogenetic analysis and is summarized in Fig. 4a (the RNase T<sub>2</sub> cleavage sites are here indicated by arrows). The two sets of prominent bands correspond to cleavages in the loop and bulge regions. Digestion with RNase T<sub>1</sub>, which specifically cleaves single-stranded RNA after G residues (8), yields cleavages only after G residues present in the loop and bulge (data not shown), in close agreement with the RNase T<sub>2</sub> results.

Fig. 4b also shows the results of similar experiments employing RNase CV<sub>1</sub> (lanes 9-12), which is specific for base-paired or stacked regions of RNA (8). Two prominent

sets of bands appear, corresponding to cleavages in the upper and lower stems of this structure (depicted as arrowheads in Fig. 4a).

The *in vitro* transcript used in these experiments contains 14 nts of polylinker sequences upstream of the HBV sequences. To exclude the possibility that these extraneous sequences altered the folding of the RNA, 12 of these nts were deleted from the DNA template by oligonucleotide mutagenesis (leaving the 5' GG required for efficient transcription by T3 RNA polymerase); the RNase cleavage patterns on this transcript are identical to those of its parent (data not shown). Thus, the RNA corresponding to the  $\epsilon$  region synthesized *in vitro* adopts the stable secondary structure depicted in Fig. 4a, with a lower stem, a 6 nt bulge, an upper stem with a single unpaired U residue, and a 6 nt loop.

**Structure of  $\epsilon$  in RNA encapsidated within core particles.** We next sought to determine whether the secondary structure observed in transcripts generated *in vitro* actually forms *in vivo*.  $\epsilon$ -LacZ chimeric RNA was therefore extracted from cytoplasmic core particles purified from HepG2 cells cotransfected with pE-LacZ and helper plasmid pCMV-CP. This deproteinized RNA was directly subjected to cleavage by RNase T2 (Fig. 5, lane 6) and RNase CV1 (Fig. 5, lane 7), without prior denaturation and refolding. The characteristic patterns of cleavage sites for the loop and bulge (lane 6) and for the upper stem and lower stem (lane 7) are evident. While we cannot rule out the existence of higher order structures resulting from RNA-protein interactions *in vivo*, our data strongly suggest that the stem-loop secondary structure, as depicted in Fig. 4a, exists *in vivo* in RNAs known to have been packaged.

**Mutational analysis of stem-loop structure and function.** To obtain genetic evidence for the existence of the stem-loop structure *in vivo*, and to explore its functional role in RNA packaging, we next introduced numerous mutations into the  $\epsilon$  region of chimeric  $\epsilon$ -LacZ constructs and assayed encapsidation *in vivo*, using the transfection assay



described in Fig. 2. The exact nucleotide sequences of these mutations are detailed in Table 1; the mutational lesions are also schematically depicted above the results of their corresponding packaging assays in Figs. 6, 7 and 9.

**(i) Stem mutations.** First we examined the functional effects of gross deletions within the stem-loop structure (Fig. 6). Complete deletion of either the upper stem (p $\Delta$ Upper; lanes 11, 12,) or the lower stem (p $\Delta$ Lower; lanes 13, 14) region abolishes RNA packaging. Thus the encapsidation function is not localized to one subdomain of the structure.

We next introduced less radical mutations into the lower stem to determine the importance of base pairing there for RNA packaging (Fig. 6). When we change four consecutive nts in either the left (pLowerL; lanes 5, 6) or right (pLowerR; lanes 7, 8) side of the stem to disrupt base-pairing in this region, in each case RNA is no longer packaged. In the doubly-mutated construct (pLowerL/R; lanes 9, 10), which contains both the changes in the left and right sides of the stem to restore base-pairing in this region, RNA is now encapsidated, albeit at 3-4 fold reduced levels compared to the wild-type construct (pE-LacZ; lanes 3, 4). This confirms that the lower stem of the structure forms *in vivo* and indicates that base-pairing in this region is a structural requirement for encapsidation.

We next introduced mutations to test the role of base-pairing in the upper stem for RNA packaging (Fig. 7a). When we alter 3 consecutive residues in the lower left side of this stem to disrupt base pairing, RNA is still encapsidated (pUpper1L; lanes 5, 6) at levels only 2-3 fold lower than wild-type (lanes 3, 4); this suggests that base pairing in this region of the upper stem is not essential for encapsidation. When we alter 3 consecutive bases in the lower right side of this stem to disrupt base pairing, RNA is no longer encapsidated (pUpper1R; lanes 7, 8); given the lack of importance of base pairing here, this suggests that these specific nucleotides are critical for encapsidation. The doubly-mutated construct, which contains both sets of changes to restore base pairing, does not direct RNA encapsidation (pUpper1L/R; lanes 9, 10); this is consistent with the fact that changing the

bases in the right side is not tolerated. In order to prove that in the doubly-mutated construct base pairing is indeed restored, we enzymatically probed the structure of these three mutated RNAs synthesized *in vitro* (Fig. 8). In each case the structure behaves as predicted for the introduced changes: mutating the left side (lane 12) or right side (lane 14) of the stem disrupts the structure locally, while in the doubly-mutated construct (lane 16) the wild-type structure is restored.

A similar set of mutations was introduced in the upper region of the upper stem (Fig. 7b). When four residues are altered in the left side (pUpper2L, lanes 5, 6) or in the right side (pUpper2R, lanes 7, 8) to disrupt base pairing, in each case RNA is encapsidated at 4-10 fold reduced levels compared to wild-type (lanes 3, 4). In the doubly-mutated construct (pUpper2L/R; lanes 9, 10), which contains both sets of mutations and should restore base pairing, RNA is now encapsidated at nearly wild-type levels. This reaffirms the existence of the upper stem *in vivo* and suggests that while base pairing in the lower portion of this stem is dispensable for packaging (Fig. 7a), base pairing in its upper region, while not absolutely essential, increases packaging efficiency.

Interestingly, when we alter the uppermost *five* consecutive residues of the upper stem (one more than altered in the pUpper2 constructs of Fig. 7b), RNA is no longer packaged in either of the singly-mutated constructs, and restoration of base pairing in the corresponding doubly-mutated construct does not restore encapsidation (data not shown). Given that in the doubly-mutated pUpper2L/R construct of Fig. 7b *four* consecutive base changes that maintain base pairing are tolerated, this finding suggests that the specific nucleotides of the uppermost base pair are a primary sequence requirement for encapsidation; i.e. they cannot be altered.

**(ii) Loop and bulge mutations.** In the constructs of Fig. 9a, mutations were introduced to determine the functional role of specific sequences in the loop and bulge regions of the  $\epsilon$  structure. When the first four nucleotides in the loop are changed, RNA is no longer encapsidated (pLoop1-4; lanes 5, 6). In contrast, when the same four base

changes are introduced in the bulge, RNA is encapsidated (pBulge1-4; lanes 7, 8) at an efficiency near that of the wild-type construct (lanes 3, 4). Changing the last two bases in the bulge also does not reduce encapsidation efficiency (data not shown). However, when the bulge is deleted entirely, RNA is not packaged (p $\Delta$ bulge; lanes 9, 10). To rule out the possibility that the above mutations alter the overall structure of the stem-loop, we have again examined the structures of these RNAs synthesized *in vitro*, using nuclease accessibility experiments (Fig. 8). In each case the structure behaves as predicted for the introduced changes: base changes in the loop (Fig. 8, lane 8) and bulge (Fig. 8, lane 10) preserve the overall structure, and deletion of the bulge eliminates the bulge without other gross effects on the stems or loop (data not shown). Thus, it appears that primary sequence in the loop is important for encapsidation; by contrast, the bulge plays principally a structural role, with its primary sequence being relatively unimportant.

Since specific primary sequences in the loop appear to be important for encapsidation, we next mutated each of the six nucleotides in the loop individually, making primarily purine to purine and pyrimidine to pyrimidine changes (Fig. 9b). Alterations at either of the first two positions of the loop has no effect on encapsidation efficiency (lanes 5, 6 and 7, 8); however, changing the base at the third position abolishes packaging (lanes 9, 10), and alterations at the fourth or fifth positions diminish levels of encapsidated RNAs over 20-fold (lanes 11, 12 and 13, 14). Altering the nucleotide at the sixth position (lanes 15, 16) also reduces encapsidation efficiency, but less dramatically (ca. 3-fold). Thus, specific nucleotides in the loop appear to play a crucial role in encapsidation.

Finally, we have examined the role of the single unpaired U residue in the upper stem in encapsidation (Fig. 9a). Deletion of this residue (p $\Delta$ U; lanes 11, 12) reduces levels of packaged RNAs at least 10-fold compared to the wild-type construct (lanes 3, 4); this demonstrates the importance of the unpaired U for encapsidation. The integrity of the overall stem-loop structure in this mutated RNA was confirmed by enzymatic probing of the RNA synthesized *in vitro* (data not shown). We have not changed the unpaired U to

other residues (to determine whether it plays primarily a structural role or is recognized specifically) because other bases at this position are predicted by computer secondary structure analysis to alter the base pairing partners within the upper stem.

## DISCUSSION

In this study we demonstrate that the 5' 94 nucleotides of HBV pg RNA that defines the  $\epsilon$  region form a stable stem-loop structure *in vitro* and *in vivo*, and that this structure is functionally important for encapsidation. Mutational analysis leads to a picture of the functional encapsidation signal that includes (i) structural requirements for base pairing in the lower stem and the upper regions of the upper stem, and for the presence of the bulge; these are relatively independent of the primary sequence in these regions, (ii) a requirement for a single unpaired U residue in the upper stem, and (iii) a requirement for specific nucleotide sequences in the loop and in regions of the upper stem. These key features are summarized schematically in Fig. 10.

While enzymatic structure probing demonstrates the existence of the upper stem both *in vitro* and *in vivo*, only a single set of mutations in this stem (pUpper2 constructs; Fig. 7b) supports a functional role for base pairing in the upper stem in encapsidation, and this role was relatively modest; disrupting base pairing reduced but did not eliminate packaged RNA. Mutations in the lower region of the upper stem (pUpper1 constructs; Fig. 7a) suggest that primary sequence and not base pairing is essential there. Thus, while base pairing in some regions of the upper stem appears to increase packaging efficiency, perhaps by presenting essential bases in the loop in a favorable orientation, disruptions in base pairing in the upper stem overall appear better tolerated than in the lower stem. This is entirely consistent with the phylogenetic analysis, in which apparent disruptions of base-pairing in the upper stem seem to be tolerated in several infectious avian hepadnaviral isolates (cf. Fig. 1).

Our results in mapping the 5' and 3' extents of the HBV  $\epsilon$  region differ in two minor respects from those reported by other investigators (2,16). First, those studies reported that 29 nts 3' of the stem-loop structure are essential for encapsidation while we find no such requirement. Because the foreign sequences which lie 3' to the stem-loop structure in their chimeric RNAs differ from ours, it is possible that the differences in these flanking sequences are responsible for this discrepancy. Second, in those studies sequences upstream of the stem-loop structure were not found to increase packaging efficiency, while we find that their inclusion increases encapsidation efficiency about 5-fold. In this context we note that Chiang et. al. (7) found that deletion of these upstream sequences from wild-type pg RNA also resulted in a clear reduction of packaging efficiency; these results agree well with our findings and support the inference that sequences 5' to the stem-loop contribute to RNA packaging. Whether this effect is mediated by specific sequences upstream of the stem-loop structure or by a spacing requirement between the 5' end of the transcript and the stem-loop structure remains to be determined.

We also observe a difference in the encapsidation efficiency of HBV pg RNA compared to that reported by others (2), who found that HBV encapsidates only ca. 10% of its pg RNA. This efficiency is considerably less than that previously observed (14) for the avian hepadnavirus DHBV (ca. 50%) However, in our hands HBV and DHBV encapsidate their pgRNAs with comparable efficiency (ca. 30-50%). We are uncertain of the reason for these differing estimates of HBV packaging efficiency, but note that we employ a different HBV subtype from that used in the earlier experiments.

Because  $\epsilon$  resides in the terminally redundant portion of pg RNA, and because of the coterminal nature of all viral RNA transcripts,  $\epsilon$  sequences are found at the 3' ends of pg RNA and all subgenomic RNAs. However, at the 3' position these sequences do not function to direct encapsidation (15). One possible explanation for this positional discrimination is that sequences upstream of the 3' copy of  $\epsilon$  interfere with the formation of

the stem-loop structure (perhaps by favoring the formation of an alternate structure); these inhibitory sequences are not present upstream of the 5' copy of  $\epsilon$ . To test this idea, we generated an *in vitro* transcript that mimics the major subgenomic mRNA, containing a single  $\epsilon$  region 3' to 1850 nts of HBV sequence. Enzymatic probing of this RNA shows that the characteristic  $\epsilon$  stem-loop structure forms stably in such transcripts [Fig. 11 (addendum)]. However, we cannot exclude the possibility that this stem-loop structure forms in shorter RNAs derived from the full-length RNA during the cleavage reaction.

These studies of  $\epsilon$  structure and function set the stage for a more detailed biochemical dissection of encapsidation. The central question now is: what is the molecular basis of  $\epsilon$  recognition? Clearly, C and P proteins are required for encapsidation *in vivo*, as judged by mutational analyses (2,14,22,23). C protein is known to have RNA-binding properties, but this activity is relatively sequence-nonspecific (4). It has been shown recently that encapsidation of  $\epsilon$ -containing RNAs and encapsidation of P protein are tightly coupled; neither is encapsidated without the other (3). A simple and attractive model is that P protein directly recognizes the encapsidation signal and targets the RNA for encapsidation into assembling cores, perhaps through noncovalent C-P interactions. In support of this, a recent report suggests that recombinant P protein made in bacteria can specifically recognize RNAs containing  $\epsilon$  sequences (17). However, this recognition operated on  $\epsilon$  elements at the 3' end of *in vitro* transcripts, a position in which they do not function in encapsidation *in vivo*; the bound RNAs also contained sequences involved in the initiation of reverse transcription, another function likely to involve P protein binding. We have recently prepared recombinant P proteins in *E. coli* and thus far we have observed only sequence-nonspecific RNA binding [Fig. 12 (addendum)]; clearly, more extensive studies of this issue are required.

Of course, this simple model is only one of several possibilities. The genetic data is equally compatible with a complex of C and P proteins being involved in the direct binding to  $\epsilon$ , or even with more improbable models. For example, P protein might function to

modify C protein to convert it to a sequence-specific RNA recognition protein.

Distinguishing among these possibilities will require development of biochemical assays for  $\epsilon$  recognition *in vitro*, and attempts to accomplish this are currently underway.

## **ACKNOWLEDGMENTS**

We thank Dan Loeb and Sophie Roy for many helpful discussions throughout the course of this work.

This work was supported by grants from the National Institutes of Health.

**TABLE 1 Stem loop mutations.** The primary sequence of the wild-type (WT) stem-loop region is shown at top, annotated above with arrows indicating the sequence components of the upper stem, lower stem, loop and bulge. Subsequent rows show the mutant designations along with their nucleotide changes. (.) represents no change of nucleotide; (-) represents deletion of a nucleotide. The encapsidation efficiency of constructs compared to pE-LacZ (WT) is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable encapsidated RNA.

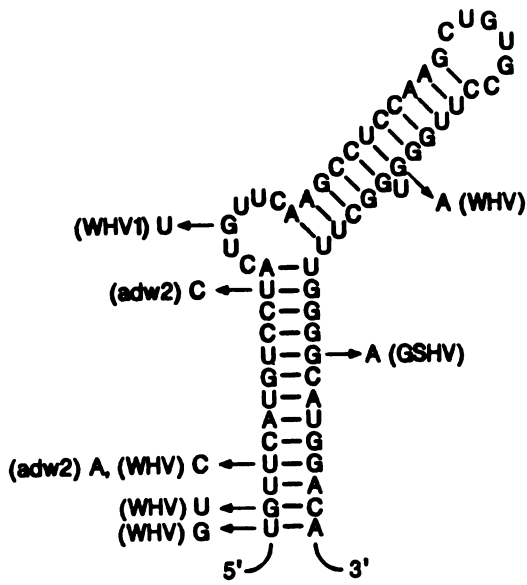


TABLE 1. Stem-loop mutations

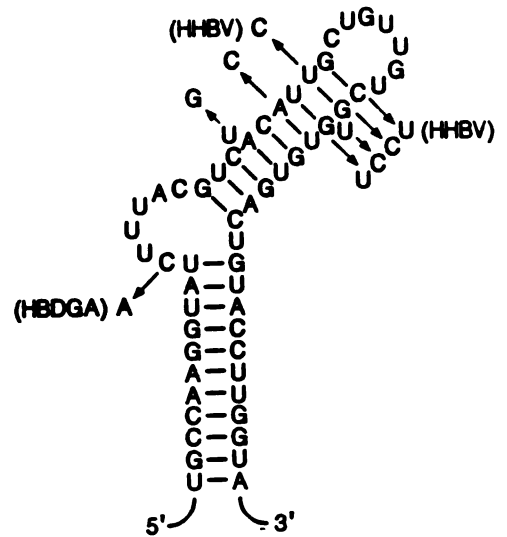
Mutation	Sequence						Encapsidation efficiency
	Lower stem	Bulge	Upper stem	Loop	Upper stem	Lower stem	
WT	UGUACAUGUCCACUUGUCAAGCCUCCAAGCUGUGCCUUGGUUGGCUUUGGGCAUGGACA						++
Δupper	.....						-
Δlower	.....						-
LowerL	.....CAGG.....						-
LowerR	.....CCTG.....						-
lowerL/R	.....CAGG.....CCTG.....						+
Upper1L	.....UCG.....						+
Upper1R	.....CGA.....						-
Upper1L/R	.....UCG.....CGA.....						-
Upper2L	.....GGUU.....						+
Upper2R	.....AACC.....						+
Upper2L/R	.....GGUU.....AACC.....						++
Loop1-4	.....UCUG.....						-
Bulge1-4	.....UCUG.....						++
Δbulge	.....						-
ΔU	.....						+/-
Loop1	.....U.....						++
Loop2	.....A.....						++
Loop3	.....A.....						-
Loop4	.....C.....						+/-
Loop5	.....A.....						+/-
Loop6	.....U.....						+

**FIG. 1 Phylogenetic analysis of the  $\epsilon$  region. RNA secondary structure proposed for the mammalian and avian hepadnaviruses based on computer prediction and phylogenetic analysis. HBV subtype ayw and DHBV16 sequences are shown, with nucleotide changes found in selected hepadnaviral isolates indicated by arrows and the corresponding hepadnaviral isolate indicated in parentheses. HBDGA is a DHBV isolate. All HBV, DHBV, woodchuck (WHV) and ground squirrel (GSHV) hepadnaviral sequences were obtained from GenBank.**

**Mammalian  
Hepadnaviruses  
(HBV Subtype ayw)**



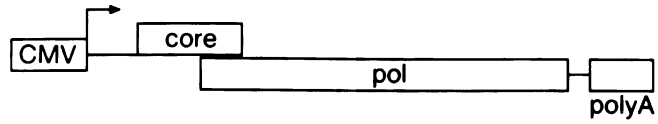
**Avian  
Hepadnaviruses  
(DHBV16)**



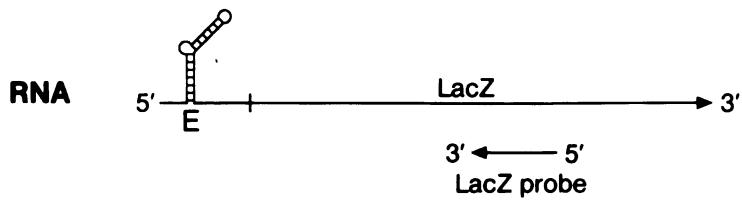
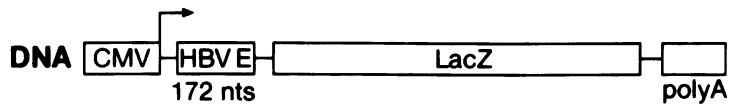
**FIG. 2 The encapsidation assay.** (a) Schematic representation of the helper plasmid pCMV-CP (Donor), which uses the CMV-IE promotor to drive expression of HBV C and P proteins, and plasmid pE-LacZ (Recipient), from which is transcribed a chimeric RNA bearing  $\epsilon$  (HBV E) fused to LacZ. (b) RNase protection analysis of total polyA<sup>+</sup> and encapsidated RNA. From equivalent numbers of transfected cells, we prepared either total polyA<sup>+</sup> RNA or RNA extracted from purified cytoplasmic core particles. RNA from equal portions of each preparation was quantified by RNase protection using a riboprobe schematically depicted by arrow. When annealed to complementary LacZ sequences and digested with RNase, this probe generates a 425 nt protected fragment. Lane 1: size standards; lane 2: undigested probe; lane 3: probe digested in the absence of added RNA sample; lanes 4, 6: core RNA (C) from cells cotransfected with helper plasmid pCMV-CP along with pLacZ and pE-LacZ, respectively; lanes 5, 7: total polyA<sup>+</sup> RNA (A) from cells cotransfected with helper plasmid pCMV-CP along with pLacZ and pE-LacZ, respectively. Lanes 4-7 are a reproduction of a section of the autoradiograph (lanes 1-4) of Fig. 3.

**A.**

**Donor**



**Recipient**

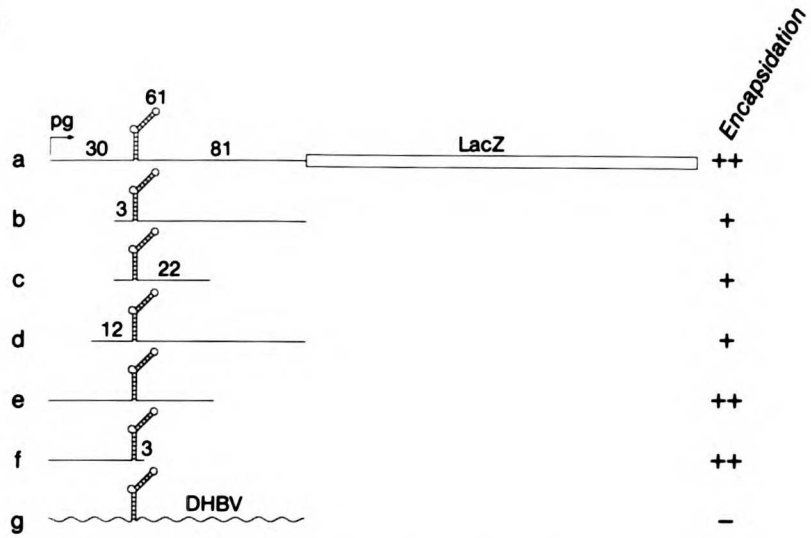


**B.**

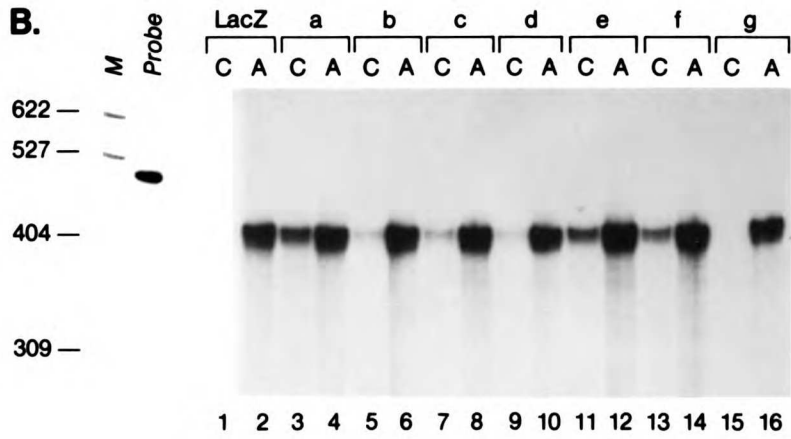


**FIG. 3 Mapping the limits of  $\epsilon$ .** (a) Schematic representations of the chimeric plasmid series containing varying extents of the 5' region of HBV pg RNA fused to LacZ. "Pg" indicates transcription initiation from the authentic pg RNA start site. The number of nts contained in each construct is indicated by the numbers above the constructs. pE-LacZ appears here as construct a. pDHBV-LacZ [g] contains the 5' sequences of DHBV pg RNA. The encapsidation efficiency of constructs compared to pE-LacZ (WT) is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable encapsidated RNA. (b) RNase protection analysis of total polyA+ and encapsidated RNA. Methods are identical to those described in the legend of Fig 2b. M: size standards; Probe: undigested probe; lanes 1, 3, 5, 7, 9, 11, 13, 15: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ [a], pb-LacZ [b], pc-LacZ [c], pd-LacZ [d], pe-LacZ [e], pf-LacZ [f], and pDHBV-LacZ [g], respectively; lanes 2, 4, 6, 8, 10, 12, 14, 16: polyA+ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pb-LacZ, pc-LacZ, pd-LacZ, pe-LacZ, pf-LacZ, and pDHBV-LacZ, respectively.

**A.**



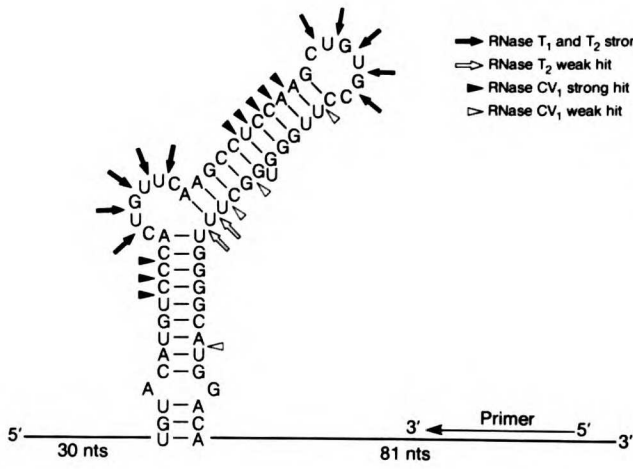
**B.**



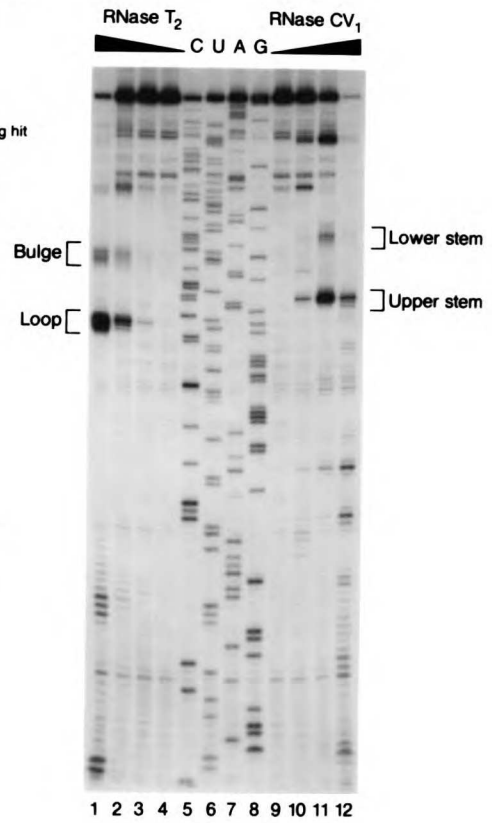
**FIG. 4 Enzymatic probing of  $\epsilon$  secondary structure in RNA synthesized *in vitro*.** (a) Schematic representation of the RNase cleavage sites drawn onto the RNA secondary structure of the  $\epsilon$  region as derived from enzymatic probing. (b) A 188 nt RNA spanning the  $\epsilon$  region was transcribed *in vitro*, renatured, and cleaved by RNases specific for single-stranded (RNase T<sub>1</sub> and RNase T<sub>2</sub>) or base-paired (RNase CV<sub>1</sub>) regions. The position of cleavage sites was determined by primer extension from an end-labelled primer homologous to sequences at the 3' end of the RNA. Lanes 1, 2, 3 and 4: RNase T<sub>2</sub> digestion with 60, 6.0, 0.6, and 0 U/ml, respectively; lanes 5-8: sequencing ladder generated from the same primer; lanes 9, 10, 11, and 12: RNase CV<sub>1</sub> digestion with 0, 0.25, 2.5 and 25 U/ml, respectively. Positions of nt cleavage corresponding to the upper and lower stems, the loop and bulge are indicated alongside the lanes.



**A.**

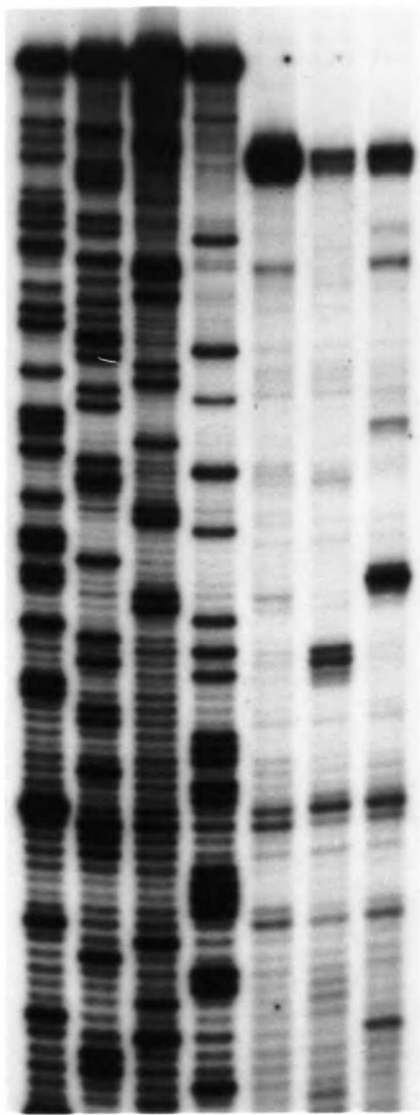


**B.**



**FIG 5. Enzymatic probing of  $\epsilon$  secondary structure in RNA from core particles.** E-LacZ chimeric RNA was extracted from core particles purified from HepG2 cells cotransfected with pE-LacZ and helper plasmid pCMV-CP. RNA was directly subjected to cleavage by RNases and the position of cleavage sites determined by primer extension from an end-labelled primer homologous to sequences at the 3' end of the RNA. Lanes 1-4 : sequencing ladder generated from the end-labelled primer; lanes 5, 6, 7: core RNA digested with no enzyme, 600 U/ml RNase T<sub>2</sub>, and 25 U/ml RNase CV<sub>1</sub>, respectively. Positions of nt cleavage corresponding to the upper and lower stems, the loop and bulge are indicated alongside the lanes.

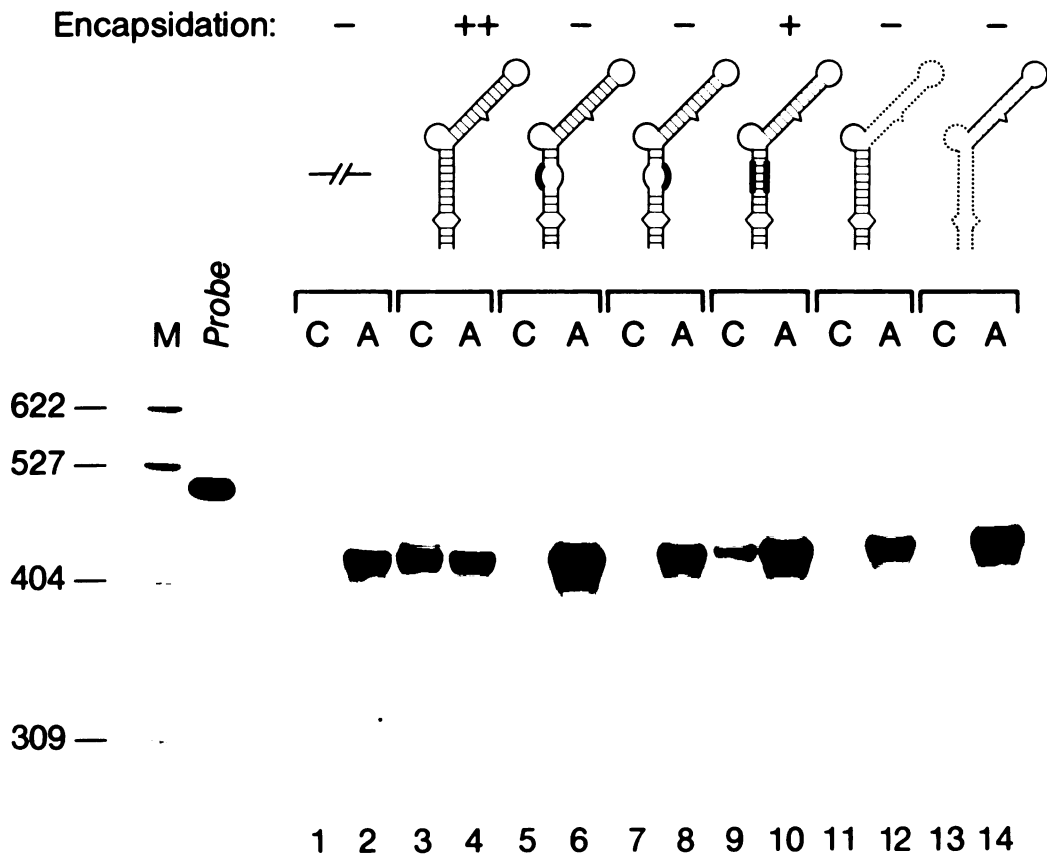
C U A G — T<sub>2</sub> CV<sub>1</sub>



← Lower stem  
← Bulge  
← Upper stem  
← Loop

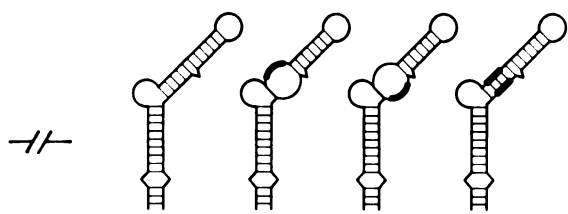
1 2 3 4 5 6 7

**FIG. 6 Encapsidation assay of RNAs with lower stem mutations.** Methods are exactly as described in the legend of Fig 2b. M: size standards; Probe: undigested probe; lanes 1, 3, 5, 7, 9, 11, and 13: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLowerL, pLowerR, pLowerL/R, pUpper, and pΔlower, respectively; lanes 2, 4, 6, 8, 10,12 and 14: polyA+ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLowerL, pLowerR, pLowerL/R, pUpper and pΔlower, respectively. The encapsidation efficiency of constructs compared to pE-LacZ (WT) is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable encapsidated RNA.



**FIG. 7 Encapsidation assay of RNAs bearing base changes in the upper stem.** Methods are identical to those described in the legend of Fig 2b. (a) M: size standards; Probe: undigested probe; lanes 1, 3, 5, 7, 9: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper1L, pUpper1R, and pUpper1L/R, respectively; lanes 2, 4, 6, 8, 10: polyA+ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper1L, pUpper1R, pUpper1L/R, respectively. (b) M: size standards; Probe: undigested probe; lanes 1, 3, 5, 7, 9: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper2L, pUpper2R, and pUpper2L/R, respectively; lanes 2, 4, 6, 8, 10: polyA+ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pUpper2L, pUpper2R, and pUpper2L/R, respectively. The encapsidation efficiency of constructs compared to pE-LacZ (WT) is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable encapsidated RNA.

Encapsidation: -    ++    +    -    -



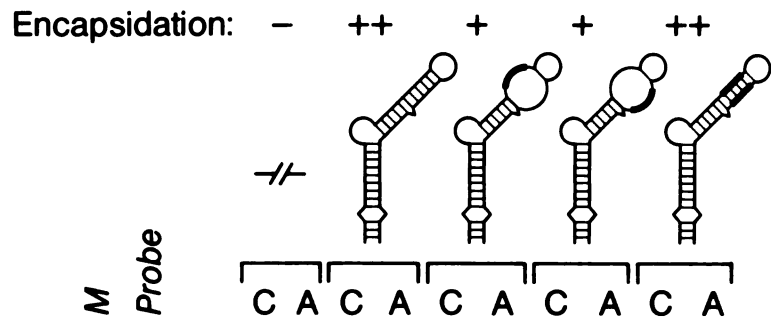
M  
Probe

C A C A C A C A C A

622 —  
527 —  
404 —  
309 —

1 2 3 4 5 6 7 8 9 10

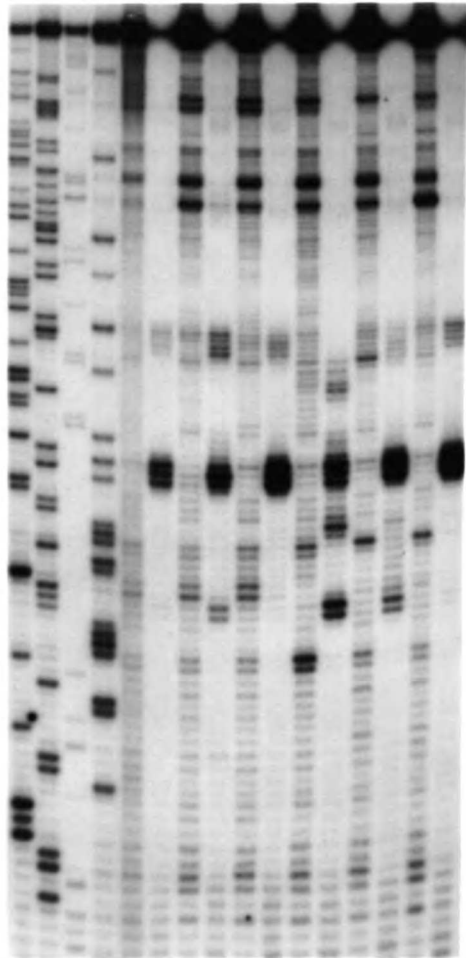






**FIG. 8 Enzymatic structure probing of mutated RNAs generated *in vitro*.** RNAs containing the indicated stem loop mutations were synthesized *in vitro*, renatured, and cleaved by RNase T<sub>2</sub>, specific for single-stranded regions. The position of cleavage sites was determined by primer extension from an end-labelled primer homologous to sequences at the 3' end of the RNA. Lanes 1-4 : sequencing ladder generated from the end-labelled primer; lanes 6, 8, 10, 12, 14, 16: 60 U/ml RNase T<sub>2</sub> digestion of WT, Loop1-4, Bulge1-4, Upper1L, Upper1R, and Upper1L/R RNAs, respectively; lanes 5, 7, 9, 11, 13, 15: no enzyme digestion control of WT, Loop1-4, Bulge1-4, Upper1L, Upper1R, and Upper1L/R RNAs, respectively. Positions of nt cleavage corresponding to the loop and bulge are indicated alongside the lanes.

WT    Loop1-4    Bulge1-4    Upper1L    Upper1R    Upper1L/R  
 C U A G    - + - + - + - + - +    RNase T<sub>2</sub>



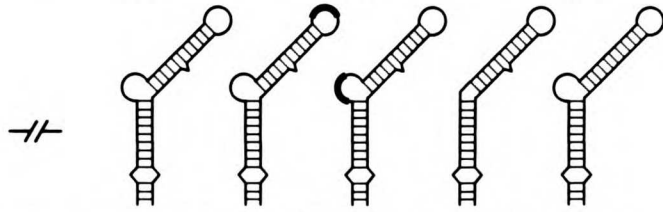
] Bulge  
 ] Loop

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

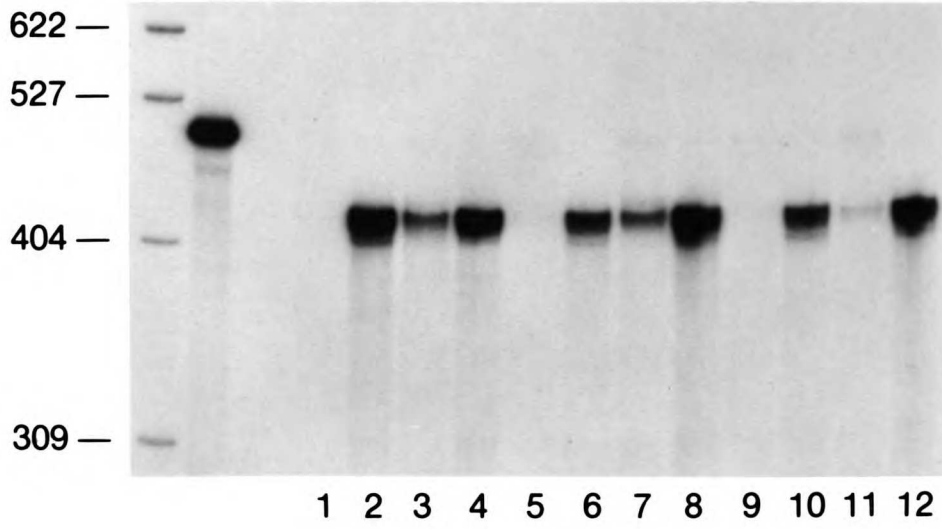
**FIG. 9 Encapsidation assay of RNAs with nucleotide changes in the loop and bulge.** Methods are identical to those described in the legend of Fig 2b. (a) M: size standards; Probe: undigested probe; lanes 1, 3, 5, 7, 9, 11: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1-4, pBulge1-4, p $\Delta$ bulge, and p $\Delta$ U, respectively; lanes 2, 4, 6, 8, 10 and 12: polyA+ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1-4, pBulge1-4, p $\Delta$ bulge, and p $\Delta$ U, respectively. (b) M: size standards; Probe: undigested probe; lanes 1, 3, 5, 7, 9, 11, 13 and 15: core RNA (C) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1, pLoop2, pLoop3, pLoop4, pLoop5, and pLoop6, respectively; lanes 2, 4, 6, 8, 10, 12, 14 and 16: polyA+ RNA (A) from cells cotransfected with pCMV-CP along with pLacZ, pE-LacZ, pLoop1, pLoop2, pLoop3, pLoop4, pLoop5, and pLoop6, respectively. The encapsidation efficiency of constructs compared to pE-LacZ (WT) is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable encapsidated RNA.

Encapsidation:

-    ++    -    ++    -    +/-

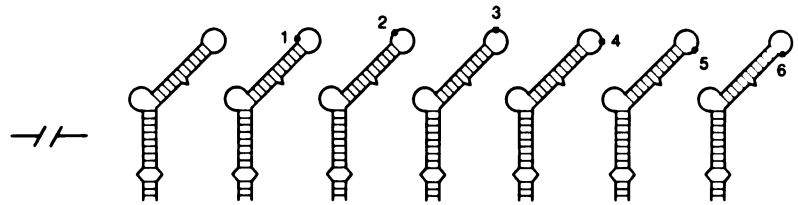


C A C A C A C A C A C A



1 2 3 4 5 6 7 8 9 10 11 12

Encapsidation: -    ++    ++    ++    -    +/-    +/-    +



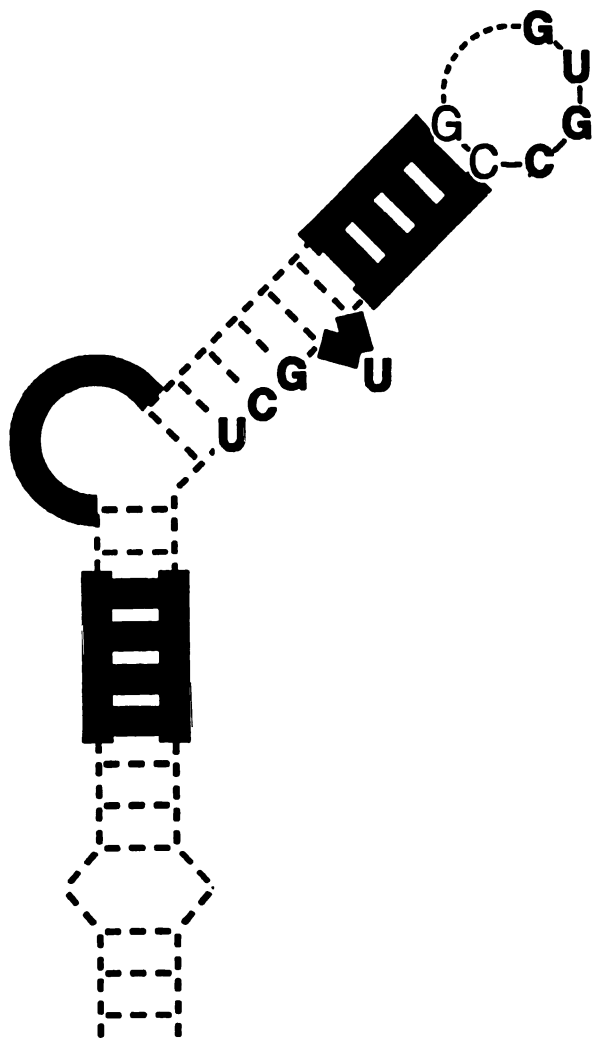
C A C A C A C A C A C A C A C A

M  
Probe  
622 —  
527 —  
404 —  
309 —

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



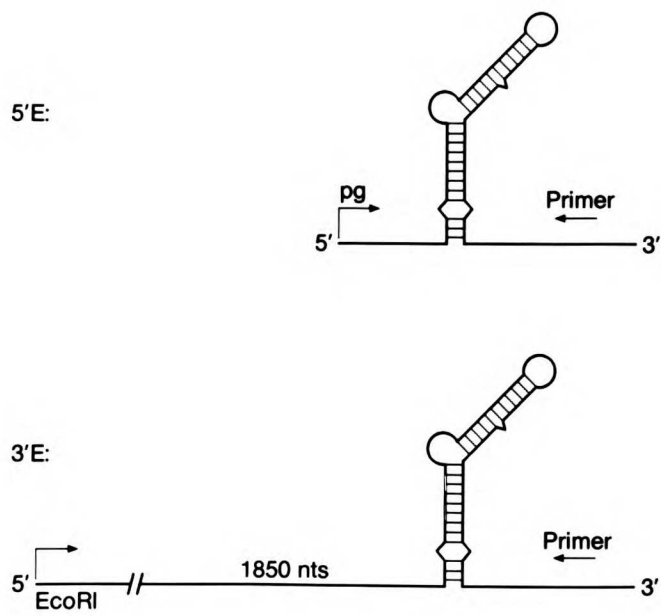
**FIG. 10 Structure-function relationships in the HBV encapsidation signal.** The secondary structure of the  $\epsilon$  region is shown with the functional requirements for encapsidation represented as follows: thick solid lines indicate important structural requirements, while specific sequence requirements are indicated by bold-faced nucleotides. Additional likely sequence requirements (see text) are indicated by plain-faced nucleotides. Dashed lines denote regions where no important functions in encapsidation have yet been assigned. We note that base pairing in the lower portion of the lower stem is not conserved phylogenetically (cf. Fig. 1), but its potential importance in encapsidation has not been addressed experimentally.



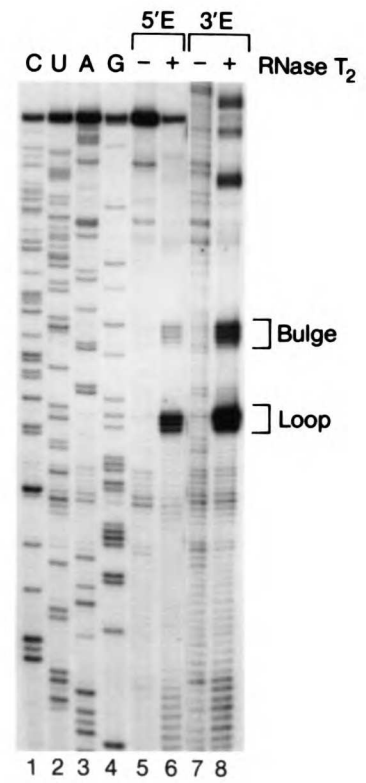
**FIG. 11 (addendum) Enzymatic probing of  $\epsilon$  secondary structure at the 5' and 3' ends of RNAs synthesized *in vitro*. (a) Schematic representation of the HBV RNAs generated *in vitro* which contain  $\epsilon$  sequences at their 5' (5'E) or 3' (3'E) ends. (b) RNAs containing 5' or 3'  $\epsilon$  sequences were transcribed *in vitro*, renatured, and cleaved by RNase T<sub>2</sub>, specific for single-stranded regions. The position of cleavage sites was determined by primer extension from an end-labelled primer homologous to sequences at the 3' end of the RNA. Lanes 1-4: sequencing ladder generated from the labelled primer; lanes 5, 6 and lanes 7, 8: 5' and 3'  $\epsilon$  RNA substrates, respectively, digested with either 0 (lanes 5, 7) or 6.0 U/ml (lanes 6, 8) RNase T<sub>2</sub>. Positions of nt cleavage corresponding to the loop and bulge are indicated alongside the lanes.**



**A.**

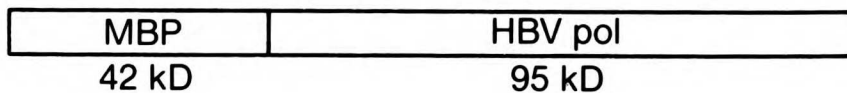


**B.**

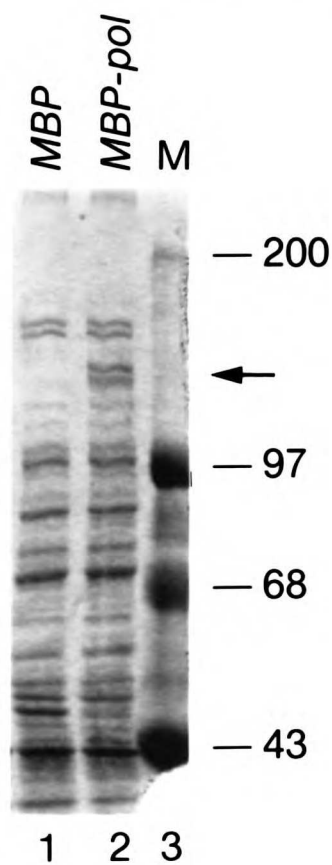


**FIG. 12 (addendum) RNA-binding of *E.coli*-produced HBV polymerase protein.** (a) Schematic representation of the maltose binding protein (MBP) - HBV polymerase fusion protein, constructed in the pMAL-c2 vector (New England Biolabs). (b) Coomassie-stained SDS-polyacrylamide gel of *E. coli* crude lysates produced by freeze/thaw lysis and sonication. MBP-pol expression was induced with 0.01 mM IPTG (8°C, 16 hrs). Lane 1: lysate from *E. coli* harboring pMAL-c2 (parent vector); lane 2: lysate from *E. coli* harboring pMBP-pol; lane 3: MW standards. MBP-pol is a 130 kD species (doublet) indicated by the arrow. (c) Northwestern RNA-protein binding assay. MBP-pol fusion protein was purified from a 2 liter *E. coli* prep using an amylose resin column according to the supplier (New England Biolabs). 1 ug MBP-pol fusion protein was subjected to SDS-PAGE, electroblotted onto a nitrocellulose (NC) membrane, and probed with <sup>32</sup>P-labelled RNAs exactly as described by Kochel et al. (17). Lanes 1, 2 and lanes 3, 4: MBP-pol NC membrane probed with either an ε or non-ε (pBS-; Statagene) containing <sup>32</sup>P-labelled RNA, respectively, in the presence of either a 200-fold (lanes 1, 3) or 2000-fold (lanes 2, 4) weight excess of *E. coli* rRNA.

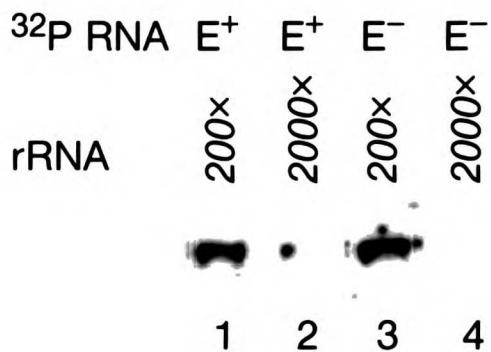
### A. MBP-pol construct



### B. Coomassie



### C. Northwestern



## REFERENCES

1. Abrahams, J.P., M. van den Berg, E. van Batenburg, and C. Pleij. 1990. Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. *Nucl. Acids Res.* 18: 3035-3044.
2. Bartenschlager, R., M. Junker-Niepmann, and H. Schaller. 1990. The P gene product of the hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J. Virol.* 64:5324-5332.
3. Bartenschlager, R. and H. Schaller. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* 11: 3413-3420.
4. Birnbaum, F. and M. Nassal. 1990. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virol.* 64:3319-3330.
5. Black, D.L. and A.L. Pinto. 1989. U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol. Cell. Biol.* 9: 3350-3359.
6. Bruss, V. and D. Ganem. 1991. The role of envelope proteins in hepatitis B virus assembly. *Proc. Natl. Acad. Sci. USA* 88: 1059-1063.
7. Chiang, P.-W., K.-S. Jeng, C.-P. Hu, and C. Chang. 1992. Characterization of a *cis* element required for packaging and replication of the human hepatitis B virus. *Virology* 186: 701-711.
8. Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J.-P. Ebel, and B. Ehresmann. 1987. Probing the structure of RNAs in solution. *Nuc. Acids Res.* 15: 9109-9128.
9. Enders, G.H., D. Ganem, and H.E. Varmus. 1985. Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. *Cell* 42: 297-308.
10. 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.
11. Ganem, D. and H.E. Varmus. 1987. The molecular biology of the hepatitis B viruses. *Ann. Rev. Biochem.* 56: 651-694.
12. Geballe, A.-P., R.R. Spaete, and E.S. Mocarski. 1986. A *cis*-acting element within the 5' leader of a cytomegalovirus b transcript determines kinetic class. *Cell* 46: 865-872.
13. Hirsch, R., R. Colgrove, and D. Ganem. 1988. Replication of duck hepatitis B virus in two differentiated human hepatoma cell lines after transfection with cloned viral DNA. *Virology* 167:136-142.

14. Hirsch, R., J. Lavine, L. Chang, H. Varmus, and D. Ganem. 1990. Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* 344: 552-555.
15. Hirsch, R., D.D. Loeb, J.R. Pollack, and D. Ganem. 1991. *Cis* - Acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. *J. Virol.* 65: 3309-3316
16. Junker-Niepmann, M., R. Bartenschlager, and H. Schaller. 1990. A short *cis*-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.* 9:3389-3396.
17. Kochel, H.G., M. Kann, and R. Thomssen. 1991. Identification of a binding site in the hepatitis B virus RNA pregenome for the viral pol gene product. *Virology* 182: 94-101.
18. Kunkel, T.A., J.D. Roberts, and R.A. Zabor. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meth. Enzymol.* 154:376-382.
19. Lavine, J., R. Hirsch, and D. Ganem. 1989. A system for studying the selective encapsidation of hepadnaviral RNA. *J. Virol.* 63: 4257-4263.
20. Martinez, H. M. 1988. An RNA secondary structure workbench. *Nucl. Acids Res.* 16: 1789-1798.
21. Nassal, M., M. Junker-Niepmann, and H. Schaller. 1990. Translational inactivation of RNA function: discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell* 63: 1357-1363.
22. Nassal, M. 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J. Virol.* 66: 4107-4116.
23. Roychoudhury A., F. Faruqi, and C. Shih. 1991. Pregenomic RNA encapsidation analysis of eleven missense and nonsense polymerase mutants of human hepatitis B virus. *J. Virol.* 65: 3617-3624.
24. Skinner, M.A., V.R. Racaniello, G. Dunn, J. Cooper, P.D. Minor and J.W. Almond. 1989. New model for the secondary structure of the 5' non-coding RNA of poliovirus is supported by biochemical and genetic data that also show that RNA secondary structure is important in neurovirulence. *J. Mol. Biol.* 207: 379-392.
25. 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.
26. Summers, J. and W.S. Mason. 1982. Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* 29:403-415.
27. Valenzuela, P., M. Quiroga, J. Zaldivar, P. Gray, and W.J. Rutter. 1980. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. *ICN/UCLA Symp. Mol. Cell. Biol.* 18: 57-70.
28. Will, H., W. Reiser, T. Weimer, E. Pfaff, M. Buscher, R. Sprengel, R. Cattaneo, and H. Schaller. 1987. Replication strategy of human hepatitis B virus. *J. Virol.* 61: 904-911.

## **CHAPTER 3**

**USE OF TAT-POLYMERASE FUSION CONSTRUCTS TO EXPLORE THE  
ENCAPSIDATION FUNCTION OF HBV POLYMERASE PROTEIN**

## **ABSTRACT**

Selective encapsidation of HBV pregenomic RNA is dependent on the presence of the  $\epsilon$  (encapsidation signal) RNA stem-loop structure, and on both core and polymerase proteins. However, it is not known whether either or both of these proteins participates in the direct recognition of the  $\epsilon$  RNA stem-loop.

In order to study the role of polymerase in the packaging reaction, we have fused HBV polymerase to the HIV-1 transactivator protein *tat*, a protein that functions by binding the RNA target TAR to augment expression from the HIV-1 LTR promoter. The *tat*-polymerase fusion transactivates the HIV-1 LTR promoter (via the *tat* portion of the fusion) and is competent to direct encapsidation of  $\epsilon$ -containing RNA (via the polymerase portion). However, the *tat*-polymerase fusion cannot transactivate an HIV-1 LTR promoter containing an  $\epsilon$  RNA sequence (in place of TAR), and is unable to direct encapsidation of TAR-containing RNA. Further, C-terminal truncations of the *tat*-polymerase fusion are unable to direct encapsidation of  $\epsilon$ -containing RNA; therefore, we are as yet unable to localize a discrete packaging domain of HBV polymerase.



## **INTRODUCTION**

Selective encapsidation of the human hepatitis B virus (HBV) pg RNA is dependent on both core and polymerase gene product(s) (1, 7, 11, 13) and on a *cis*-acting encapsidation signal, termed  $\epsilon$ , on pg RNA (1, 9).  $\epsilon$  consists of an RNA stem-loop whose structure and (in portions) sequence direct encapsidation (12). Encapsidation of  $\epsilon$ -containing RNA and polymerase is tightly coupled; neither is encapsidated without the other (1, 2, 7). This has led to a model in which encapsidation is initiated by polymerase recognition of  $\epsilon$ . However, such recognition has yet to be demonstrated either *in vivo* or *in vitro*. *In vitro* RNA binding studies have been hindered by the inability to express soluble, enzymatically active HBV polymerase protein in a variety of recombinant vectors. It is also possible that core protein and/or cellular proteins are involved in the recognition of  $\epsilon$ .

Mutations have been reported in all of the subdomains of polymerase (terminal protein, spacer, reverse transcriptase, and RNaseH) which abolish its packaging function (1, 7, 13). This has been interpreted to mean that all of polymerase is required to mediate encapsidation. Importantly, however, protein stability data on these mutants has not been available, due principally to the low abundance of polymerase in cells. It is equally plausible then that these mutations merely destabilize the polymerase polypeptide.

In order to study the role of polymerase in the packaging reaction, and to localize a packaging domain of polymerase, we have fused HBV polymerase to the human immunodeficiency virus type 1 (HIV-1) transactivator protein *tat*, a protein that functions in HIV-1 replication by binding the RNA target TAR to augment expression of HIV-1 transcripts. In this study we show that the *tat*-polymerase fusion is able to transactivate the HIV promoter (via the *tat* portion of the fusion) and is competent to direct encapsidation of  $\epsilon$ -containing RNA (via the polymerase portion of the fusion). However, we find that the *tat*-polymerase fusion cannot transactivate an HIV-1 LTR promoter containing an  $\epsilon$  RNA

sequence (in place of TAR), and is unable to encapsidate TAR-containing RNA. These results suggest specific models for polymerase/core recognition of  $\epsilon$ . We also show that while C-terminal truncations of the tat-polymerase fusion are stably expressed (as determined by transactivation of the HIV-1 LTR promoter), they are unable to direct encapsidation of  $\epsilon$ -containing RNA; therefore, we are as yet unable to localize a discrete packaging domain of HBV polymerase.

## **MATERIALS AND METHODS**

**Materials.** Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions. Radionucleotides were purchased from Amersham Corp.

**Plasmids.** All HBV nucleotide positions are numbered from the unique EcoRI site of HBV adw2 (16). In this numbering scheme, nt 1815 is the transcription initiation site of pg RNA.

Plasmid pCMV-CP (12), which expresses core and polymerase, was constructed by inserting overlenght HBV sequences from nts 1880-2012 into the polylinker of pCDNA1 (Invitrogen), downstream of the CMV-IE promoter. Plasmid pCMV-C (also called pCMV-CP.626), which expresses core protein, contains an amber (UAG) nonsense mutation at amino acid position 626 of polymerase created by cloning oligonucleotide CTAGTCTAGACTAG (Pharmacia) into the HpaI site (Russell Hirsch, unpublished data). pCMV-P, which expresses polymerase, contains a TaqI (nt 2012) HBV monomer cloned into pCDNA1.

pSVtat-pol, which encodes the tat-polymerase fusion protein, was created by cloning polymerase sequence (nts 2305-1982) in frame into the HindIII site (amino acid position 67) of pSVTAT (15; gift of B.M. Peterlin). pSVTAT(1/67) encodes amino acids 1-67 of tat (15; gift of B.M. Peterlin). Ochre (TAA) nonsense mutations were introduced at amino acid positions 195, 533 652 and 788 of polymerase by digestion with NdeI (within hemagglutinin epitope tag), SpeI, EcoRV, and NcoI, respectively, followed by blunting and insertion of the self-annealing oligonucleotide CGTTAATTAATTAACG. Ochre mutations were confirmed by restriction enzyme digestion and DNA sequencing. The amber (UAG) nonsense mutation at amino acid position 626 of polymerase was

created by swapping in the appropriate restriction fragment from pCMV-CP.626. Doubly mutated polymerases were engineered by swapping in appropriate restriction fragments, and confirmed by restriction enzyme digestion.

HIV-TAR-CAT [HIVSCAT, (15)] is a gift of B.M. Peterlin. HIV- $\epsilon$ -CAT was constructed by replacing the SphI-HindIII fragment of HIV-TAR-CAT with the polymerase chain reaction (PCR)-amplified fragment containing HBV nts 1845-1986 (along with sequences between the SphI site and the start site of HIV-1 transcription), so as to replace TAR with  $\epsilon$  sequences.

pE-LacZ, which expresses the chimeric  $\epsilon$ -LacZ RNA, has been described elsewhere (12). pTAR-LacZ was constructed by cloning the PCR-amplified fragment containing HIV-1 nts 1-85 (+1 is the transcription initiation site of HIV-1 LTR promoter) from pHIV-TAR-CAT into the HindIII-KpnI fragment of pE-LacZ so as to replace  $\epsilon$  sequences with HIV-1 TAR. In pTAR(30)-LacZ, TAR sequences are preceded by the 5'-terminal 30 nts of HBV pgRNA (introduced in the PCR primer).

**Cell Culture and Transfections.** HepG2 human hepatoma cells and HeLa cells were grown in HME16 medium supplemented with 10% fetal calf serum, 0.14% sodium bicarbonate and 2 mM glutamine, and passaged every 2-3 days at a 1:3 dilution. DNA transfections were performed by the calcium phosphate coprecipitation method exactly as described previously (6).

**Assays of CAT activity.** Cell extracts for CAT activity were prepared by freeze-thaw lysis (14). CAT activity was determined in extracts using a phase separation procedure. In brief, 5  $\mu$ l cell extract was incubated in a 100  $\mu$ l volume containing 250 mM Tris-HCl pH 8.0, 0.5 mM n-butyryl CoA (Sigma) and 0.2  $\mu$ Ci  $^{14}$ C-chloramphenicol (58.1 mCi/mmol) for 15-60 minutes at 37°C. Phase extraction was performed with an equal volume of 2:1 hexane/xylene mix, and levels of n-butyrylated chloramphenicol determined by liquid

scintillation counting. CAT activity is reported as fold-transactivation relative to CAT activity resulting from transfection of target CAT reporter alone.

All results are reported as corrected for transfection efficiency using the HGH Allegro kit (Nichols Institute) used according to manufacturer's instruction.

**RNA preparation and RNase protection assay.** PolyA<sup>+</sup> total cellular RNA was purified 72 hrs post-transfection as previously described (6). RNA within cytoplasmic core particles as previously described (10) by polyethylene glycol precipitation of core particles followed by proteinase K digestion, phenol extraction and precipitation of nucleic acid. RNase protection analysis was performed as previously described (7) on polyA<sup>+</sup> RNA or core RNA prepared from equivalent numbers of transfected HepG2 cells. Synthesis of [ $\alpha$ -<sup>32</sup>P]CTP-labeled RNA probes was carried out as previously described (7); Probe lacZ-P108 (8) is a 475 nt riboprobe with about 50 nonhybridizing polylinker nts and 425 nts of LacZ complementary to the LacZ fragment contained in pLacZ.

## **RESULTS AND DISCUSSION**

**Experimental strategy.** To explore the encapsidation function of HBV polymerase *in vivo* we have fused polymerase to the HIV-1 transactivator protein tat. HIV-1 tat protein is known to function in HIV-1 replication by binding to an RNA hairpin, TAR, in the nascent HIV-1 RNA transcript and, from this position, up-regulating expression from the HIV-1 LTR promoter (4, 5). Tat has two functionally separate domains (3), an arginine-rich RNA binding domain (which binds TAR) and an activation domain (which transactivates transcription of the HIV-1 LTR promoter). Selby and Peterlin (15) in fact demonstrated that a heterologous RNA binding domain could replace that of tat without disturbing its transactivation function. Specifically, these authors showed that a tat-R17 bacteriophage coat protein fusion transactivated an HIV-1 LTR promoter in which TAR was replaced by the R17 coat protein operator (the RNA binding sequence of the R17 coat protein). In this case, the R17 coat protein functioned to bring the activation domain of tat in proximity of the HIV-1 LTR promoter to effect transactivation.

Based on this strategy, we sought to replace the RNA binding domain of tat with the putative RNA binding polymerase protein. If polymerase binds  $\epsilon$  *in vivo*, we might then expect a tat-polymerase fusion to transactivate an HIV-1 LTR promoter in which TAR has been replaced by  $\epsilon$  sequences; in this case the polymerase portion of the fusion would mediate RNA binding (to  $\epsilon$ ) while the tat portion would function in transactivation. Conversely, we might also expect a tat-polymerase fusion to direct encapsidation of TAR-containing RNA; in this case the tat portion of the fusion would mediate RNA binding (to TAR) while the polymerase portion would direct encapsidation.

Since the expression of tat-polymerase fusion proteins can be monitored by their ability to transactivate the wild-type HIV-1 LTR promoter (via the tat portion interacting with TAR), we can also truncate the polymerase portion of the fusion in order to localize a

subregion sufficient to direct RNA encapsidation. Expression data allows us to interpret the effects of truncations that abolish encapsidation function.

**Tat-polymerase fusion transactivates TAR but not  $\epsilon$  target.** A tat-polymerase fusion (pSVtat-pol; Fig. 1a) was constructed by cloning the full-length HBV polymerase gene downstream and in frame with HIV-1 tat (amino acids 1-67). Amino acids 1-67 of tat were used because of a convenient restriction site and because this fragment is known to effect near-wildtype levels of transactivation (15). We expect the tat-polymerase fusion to bind TAR (via its tat domain) and effect transactivation of the HIV-1 LTR promoter (also via its tat domain). To verify this prediction, the tat-polymerase fusion construct was cotransfected along with the reporter target HIV-TAR-CAT (Fig. 1a), which contains the chloramphenicol acetyltransferase (CAT) gene downstream of the wildtype HIV-1 LTR promoter (which contains TAR), into HeLa cells. Tat-mediated transactivation was assayed by measuring CAT enzymatic activity 48 hours post-transfection.

Fig. 1b shows the results of a typical experiment, reported as fold-transactivation compared with transfection of the target construct (in this case HIV-TAR-CAT) alone. The tat-polymerase fusion (row 2) transactivates a TAR target as well as (or in this case better than) tat (aa 1-67, hereafter referred to as tat, row 1). Therefore we conclude that the tat portion of the fusion is active for both TAR binding and transactivation.

We next asked whether the tat-polymerase fusion would function to transactivate an HIV-1 LTR promoter in which TAR sequences have been replaced by  $\epsilon$ . For this purpose we constructed HIV- $\epsilon$ -CAT (Fig. 1a), which contains the  $\epsilon$  stem-loop structure immediately adjacent to the start site of HIV-1 LTR transcription. As expected, cotransfection of tat, which should not bind  $\epsilon$ , along with HIV- $\epsilon$ -CAT (row 3) results in no transactivation. Cotransfection of the tat-polymerase fusion with the HIV- $\epsilon$ -CAT target (row 4) also results in no detectable transactivation (the difference reported is small and not reproducible). The tat-polymerase fusion is also unable to transactivate the HIV- $\epsilon$ -CAT

target in HepG2 hepatoma cells, and no activation of  $\epsilon$  is observed when tat-polymerase is co-expressed with HBV core protein (data not shown). Additionally, the fusion is unable to transactivate an HIV- $\epsilon$ -CAT target in which HIV-1 transcription initiates at the authentic start site of HBV pg RNA, 30 nt 5' of the stem-loop structure (data not shown).

There are many possible reasons why the tat-polymerase fusion might fail to activate an HIV-1 LTR promoter containing  $\epsilon$  in place of TAR. It is possible that polymerase- $\epsilon$  interaction is of insufficient affinity to produce tat-mediated transactivation. Also, the polymerase- $\epsilon$  interaction required to effect transactivation must take place in the nucleus, while the polymerase- $\epsilon$  interactions in packaging are normally cytoplasmic; perhaps conditions in the nucleus are not suitable for polymerase- $\epsilon$  interaction. Therefore, these experiments do not exclude a polymerase- $\epsilon$  interaction *in vivo*.

**Tat-polymerase fusion does not encapsidate  $\epsilon$  RNA.** Intrigued by the superficial similarities between the TAR and  $\epsilon$  structures, we next asked whether the tat-polymerase fusion could encapsidate an RNA bearing TAR in place of  $\epsilon$ .

First, we verified that the polymerase portion of the tat-polymerase fusion was active, i.e. that it could function to direct encapsidation of  $\epsilon$ . For this purpose, HepG2 cells were transfected with plasmid pE-LacZ (Fig. 2a) encoding an  $\epsilon$ -LacZ chimeric RNA, the target to be encapsidated by core, supplied by CMV-C (Fig. 2a), a CMV-driven genome with a stop codon in polymerase (known to disrupt its function), together with the tat-polymerase fusion (SVtat-pol, Fig. 2a). Following transfection, we harvested from equivalent numbers of transfected cells either the total cellular polyA<sup>+</sup> RNA or the RNA contained within purified cytoplasmic core particles. The RNA present in equal portions of each preparation was quantified by RNase protection using a riboprobe derived from LacZ sequences; the ratio of encapsidated to total polyA<sup>+</sup> RNA is a measure of the packaging efficiency.



Figure 2b, lanes 3, 4 show the results from a transfection of the  $\epsilon$  target, the core donor and the tat-polymerase fusion. The  $\epsilon$  RNA is produced and encapsidated, as we find a protected band of the expected size (425 nts) in both the sample containing polyA+ RNA (lane 4) and the sample containing RNA from core particles (lane 3). The packaging efficiency (ca. 50%) of the tat-polymerase fusion is equivalent to that of the wildtype polymerase (data not shown), and we conclude the polymerase portion of the fusion is active to direct encapsidation. Tat alone (without polymerase) is unable to direct encapsidation (lanes 1, 2).

We next asked whether the tat-polymerase fusion could direct the encapsidation of chimeric RNA bearing TAR at its 5' end. We therefore transfected the tat-polymerase fusion and the core donor together with either pTAR-LacZ (lanes 5, 6), which contains the TAR hairpin immediately adjacent to the initiation site of transcription, or p(30)TAR-LacZ (lanes 7, 8) in which TAR sequences are preceded by the 5'-terminal 30 nts of HBV pgRNA (as is  $\epsilon$  in pregenomic RNA). The tat-polymerase fusion is unable to direct the encapsidation of these chimeric RNAs. [We have also constructed an R17 bacteriophage coat protein-polymerase fusion and have shown it to be unable to direct encapsidation of an RNA bearing at its 5' end the R17 coat protein operator (data not shown)].

One model for encapsidation is that assembling core subunits recognize polymerase bound to an RNA target. If this were the case, we might expect tat-polymerase tethered to TAR (via tat) to be a suitable substrate for core encapsidation. Since it is not, we suggest that to effect encapsidation core might itself directly recognize elements of  $\epsilon$  structure, or recognize polymerase only once it has been modified in some way by binding specifically an  $\epsilon$  target.

**Tat-polymerase C-terminal truncations are stably expressed.** We next set out to localize a packaging domain of polymerase by truncating the polymerase portion of the fusion and assaying for RNA packaging activity. The expression of the truncated tat-

polymerase fusion proteins can be monitored by their ability to transactivate the wildtype HIV-1 LTR promoter (via the tat portion interacting with TAR); such expression data allows us to rule out the possibility that truncations that abolish encapsidation function do so by simply destabilizing the protein.

Therefore, we introduced into the tat-polymerase fusion a series of UAA ochre nonsense mutations in the polymerase reading frame so as to express the N-terminal 788, 652, 533 or 195 amino acids of polymerase (depicted in Fig. 3a). To determine protein expression, these truncation constructs were cotransfected along with the HIV-TAR-CAT reporter construct into HepG2 cells and CAT activity was determined 48 hours later. The results of a typical experiment are reported in Fig. 3b as fold-transactivation of the HIV promoter as compared to transfection of the HIV-TAR-CAT reporter alone. Each of the truncated fusion products (lanes 4-7) is expressed and transactivates the HIV-1 LTR promoter as well as the full-length tat-polymerase fusion (row 3) and tat alone (row 2).

**A single tat-polymerase truncation directs encapsidation of  $\epsilon$ .** We next asked how the tat-polymerase truncations functioned in encapsidation (Fig. 4). Along with the  $\epsilon$  target (pE-LacZ) and the core donor (pCMV-C) we transfected each of the truncated tat-polymerase fusions and determined levels of encapsidated RNA. When a stop codon is present at aa 788 (lanes 5, 6), 652 (lanes 7, 8), or 533 (lanes 9, 10), RNA is not encapsidated. However, when the stop codon is present at aa 195 (in the spacer region of polymerase; lanes 11, 12), encapsidation now proceeds at levels nearly equivalent to the full-length fusion (lanes 3, 4).

That only the shortest fusion (tatpol.195) was active for encapsidation was somewhat surprising, so we considered several alternative possibilities for its packaging activity. The straightforward interpretation is that the N-terminal 195 aa fragment of polymerase is indeed sufficient to direct RNA encapsidation. However, encapsidation may also result from (1) high-level nonsense suppression creating a full-length read-through

fusion product, (2) translational reinitiation downstream of the aa 195 stop codon producing N- and C-terminal polymerase fragments which together are competent to package, (3) an RNA splicing event which effectively removes the stop codon from the transcript, or (4) recombination between the core donor and tat-polymerase input DNAs, re-creating the full length polymerase open reading frame.

**N-Terminal polymerase fragment is not sufficient to direct encapsidation.** If the N-terminal 195 aa polymerase fragment alone is indeed sufficient to direct RNA encapsidation, then a second stop codon introduced farther downstream in the same construct should not interfere with packaging. We therefore placed a second termination codon downstream of the aa195 stop, at aa626 (depicted in Fig. 5a). While the singly-mutated tatpol.195 directs RNA packaging (Fig. 5b, lanes 3, 4), the doubly-mutated tatpol.195/626 which bears both stop codons does not direct encapsidation (lanes 5, 6). Thus, we conclude polymerase coding sequence downstream of aa195 is required in the tatpol.195 construct, and that the N-terminal 195 aa fragment alone is not sufficient to direct encapsidation.

To explore the remaining alternative explanations for the packaging activity of tatpol.195, we next introduced the aa 195 stop codon mutation into the polymerase sequences of the core donor construct (depicted in Fig. 6a). While tatpol.195 together with the original core donor (which has a single stop codon at aa 626) encapsidates RNA (Fig. 6b, lanes 3, 4), tatpol.195 together with the doubly-mutated core donor (containing both stop codons) does not package RNA (lanes 5, 6). Therefore we must conclude that the core donor is contributing polymerase sequences to effect encapsidation.

This contribution may be in the form of recombination between the two constructs, restoring the full-length polymerase open reading frame; however the strength of this signal is much greater than we would expect if it were due to recombination. Alternatively, encapsidation may be the result of intragenic complementation; for example, the core donor

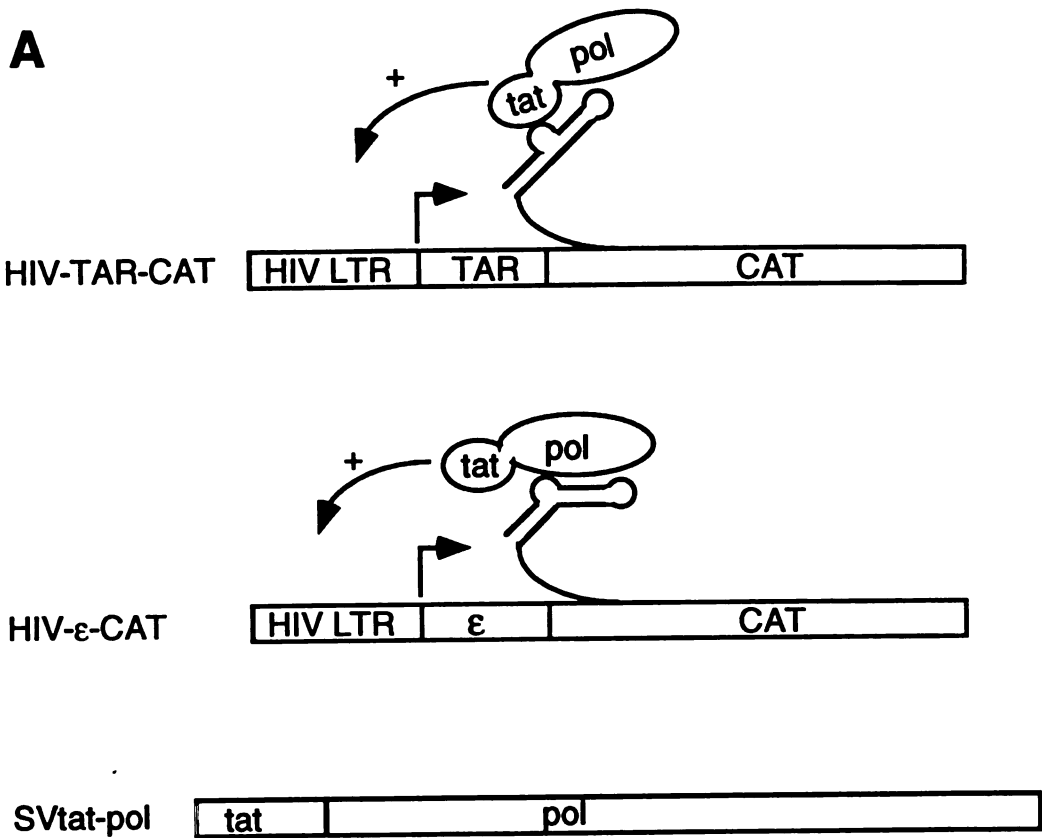
may supply an N-terminal polymerase fragment (which would be disrupted by stop codon 195) while tatpol.195 supplies a C-terminal fragment (which would be disrupted by downstream stop codon 626) through either translational reinitiation or RNA splicing. These two fragments together would effect RNA encapsidation.

One possible example of intragenic complementation, in DHBV polymerase, has in fact been reported. Wu et al.(17) introduced a frameshift mutation into the spacer region of DHBV polymerase so as to potentially create an N-terminal polymerase fragment and a C-terminal fragment (in this case fused to Pre-S) which together appear to mediate encapsidation and DNA synthesis. It remains to be determined whether our own results, as described above, are the consequence of similar intragenic complementation.

#### **ACKNOWLEDGMENTS**

We thank B. M. Peterlin for contributing plasmids and A. Frankel for helpful discussions.

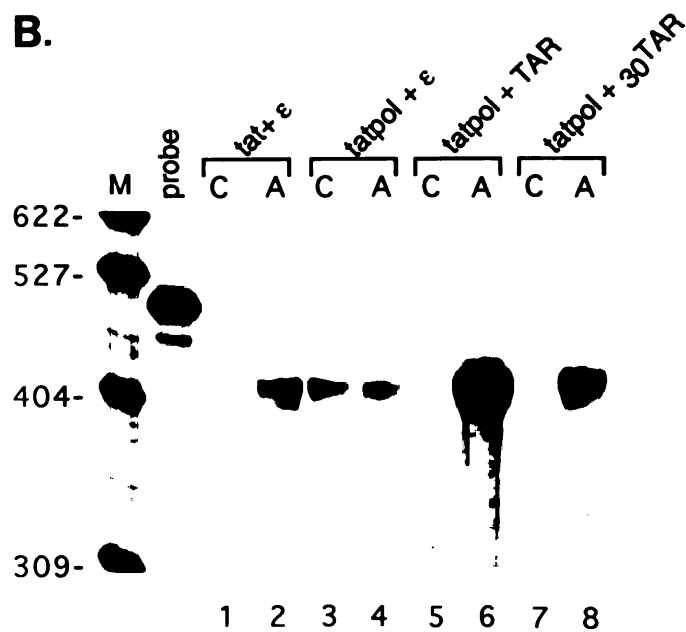
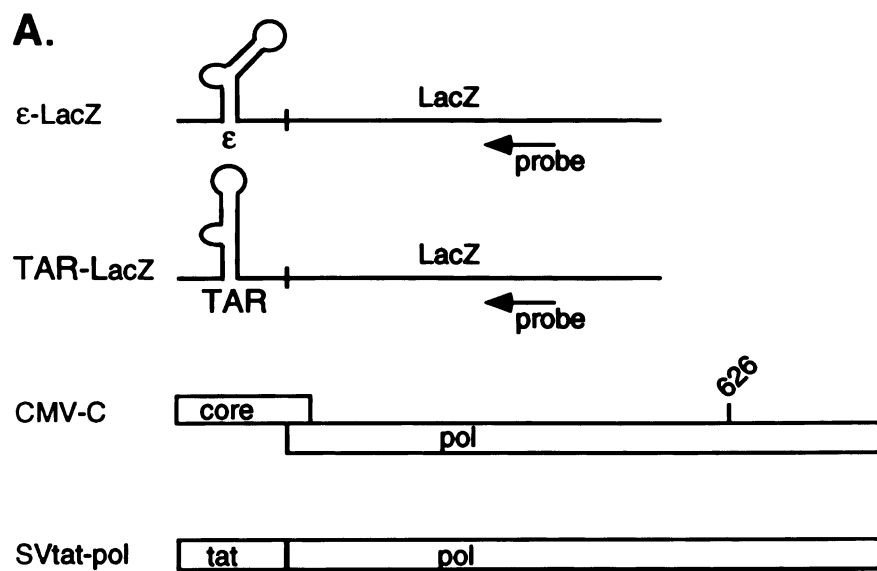
**FIG. 1 Tat-polymerase transactivation of TAR and  $\epsilon$  targets. (a) Schematic representation of HIV-TAR-CAT and HIV- $\epsilon$ -CAT targets which contain the chloramphenicol acetyltransferase gene downstream of either TAR or  $\epsilon$  sequences, driven by the HIV-1 LTR promoter. pSVtat-pol expresses the tat-polymerase fusion protein. (b) Representative transactivation data from a single experiment. CAT activity was determined 48 hours post-transfection by a phase separation method. Results are reported as fold-transactivation of target by effector as compared to transfection of the target alone.**



**B**

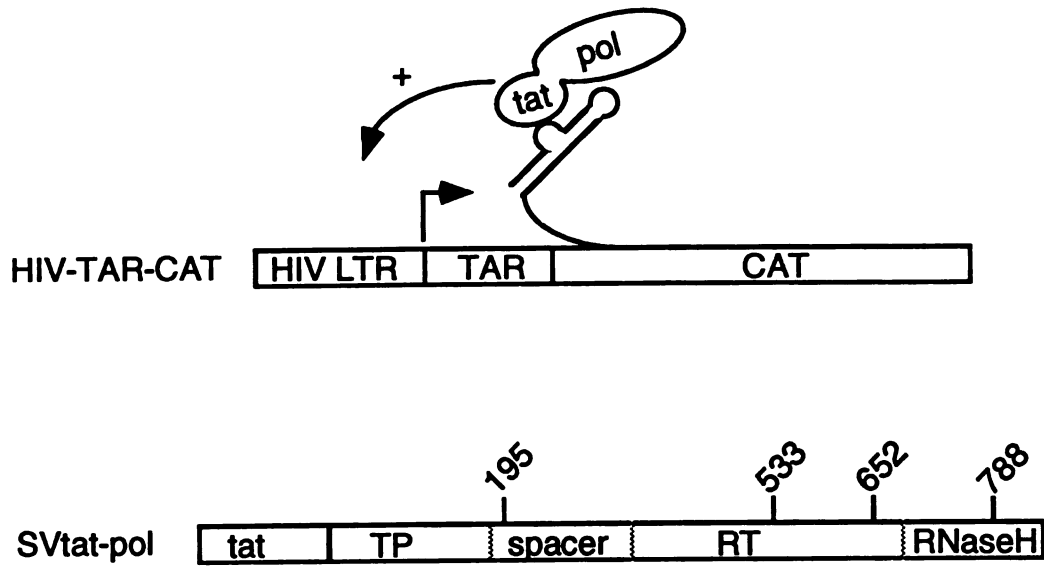
<u>EFFECTOR</u>	<u>TARGET</u>	<u>FOLD-TRANSACTIVATION</u>
1. tat	TAR	170
2. tatpol	TAR	340
3. tat	$\epsilon$	1
4. tatpol	$\epsilon$	2

**FIG. 2 Tat-polymerase encapsidation of  $\epsilon$  and TAR RNA.** (a) Schematic representation of the RNA targets pE-LacZ, from which is transcribed a chimeric RNA bearing  $\epsilon$  fused to LacZ, and pTAR-LacZ from which is transcribed an RNA bearing TAR fused to LacZ. pCMV-C expresses core protein and pSVtat-pol expresses the tat-polymerase fusion protein. (b) RNase protection analysis of total polyA+ and encapsidated RNA. From equivalent numbers of transfected cells, we prepared either total polyA+ RNA or RNA extracted from purified cytoplasmic core particles. RNA from equal portions of each preparation was quantified by RNase protection using a riboprobe schematically depicted by arrow. When annealed to complementary LacZ sequences and digested with RNase, this probe generates a 425 nt protected fragment. M: size standards; probe: undigested LacZ probe; lanes 1-8: core RNA (C; lanes 1, 3, 5, 7) and polyA+ RNA (A; lanes 2, 4, 6, 8) from cells transfected with pCMV-C along with either pSVtat and pE-LacZ (lanes 1, 2), pSVtat-pol and pE-LacZ (lanes 3, 4), pSVtat-pol and pTAR-LacZ (lanes 5, 6), or pSVtat-pol and pTAR(30)-LacZ (lanes 7, 8).



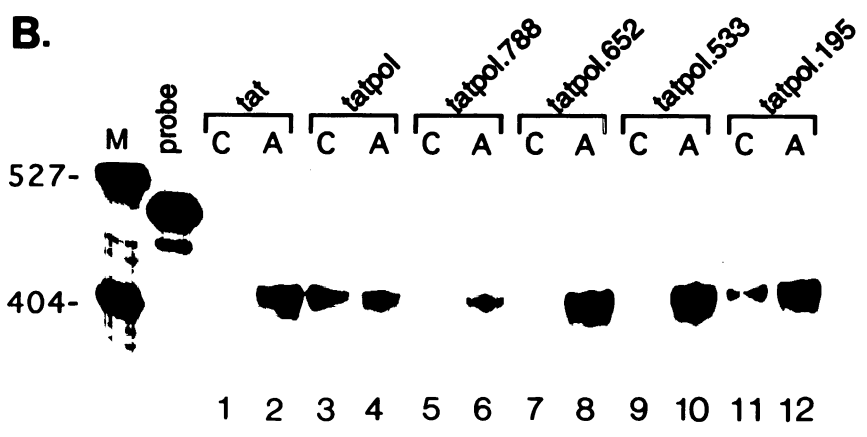
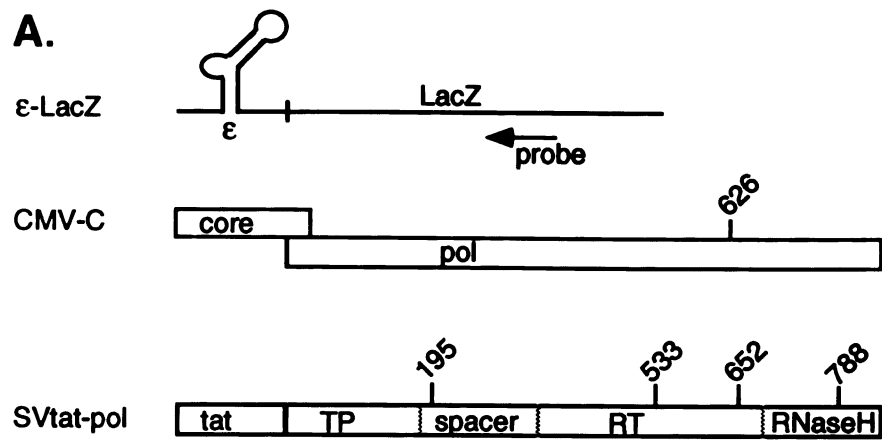


**FIG. 3 Transactivation of TAR by tat-polymerase truncations.** (a) Schematic representation of HIV-TAR-CAT which contains the chloramphenicol acetyltransferase gene downstream of TAR, driven by the HIV-1 LTR promotor. Tat-polymerase truncations express the N-terminal 788 (tatpol.788), 652 (tatpol.652), 533 (tatpol.533), or 195 (tatpol.195) aa of polymerase. (b) Representative transactivation data from a single experiment. CAT activity was determined 48 hours post-transfection by a phase separation method. Results are reported as fold-transactivation of target by effector as compared to transfection of the target alone.

**A****B**

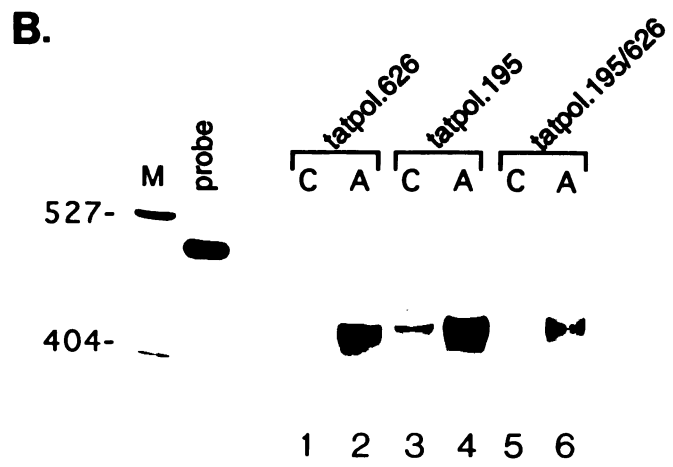
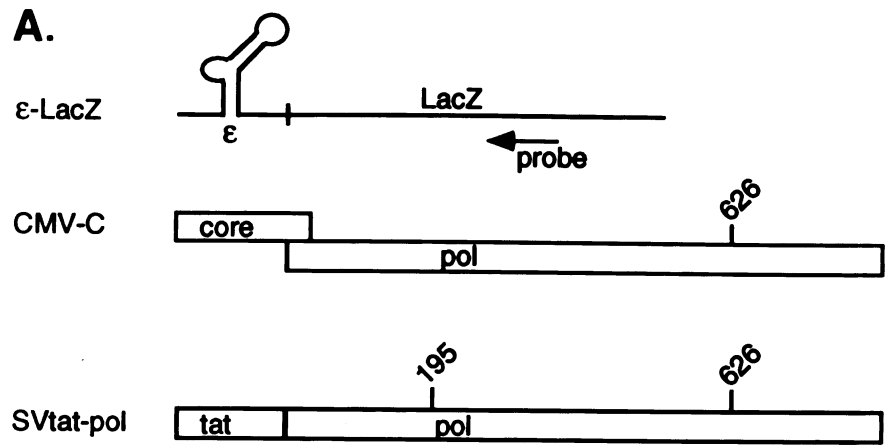
<u>EFFECTOR</u>	<u>TARGET</u>	<u>FOLD-TRANSACTIVATION</u>
1. none	TAR	-
2. tat	TAR	85
3. tatpol	TAR	245
4. tatpol.788	TAR	135
5. tatpol.652	TAR	190
6. tatpol.533	TAR	205
7. tatpol.195	TAR	215

**FIG. 4  $\epsilon$  RNA encapsidation directed by tat-polymerase truncations.** (a) Schematic representation of the RNA target pE-LacZ, from which is transcribed a chimeric RNA bearing  $\epsilon$  fused to LacZ. pCMV-C expresses core protein and pSVtat-pol truncations express the N-terminal 788, 652, 533, and 195 aa of polymerase. (b) RNase protection analysis of total polyA<sup>+</sup> and encapsidated RNA. Methods are as described in the legend of Fig 2b. M: size standards; probe: undigested LacZ probe; lanes 1-12: core RNA (C; lanes 1, 3, 5, 7, 9, 11) and polyA<sup>+</sup> RNA (A; lanes 2, 4, 6, 8, 10, 12) from cells transfected with pE-LacZ and pCMV-C along with either pSVtat (lanes 1, 2), pSVtat-pol (lanes 3, 4), pSVtat-pol.788 (lanes 5, 6), pSVtat-pol.652 (lanes 7, 8), pSVtat-pol.533 (lanes 9, 10), or pSVtat-pol.195 (lanes 11, 12).

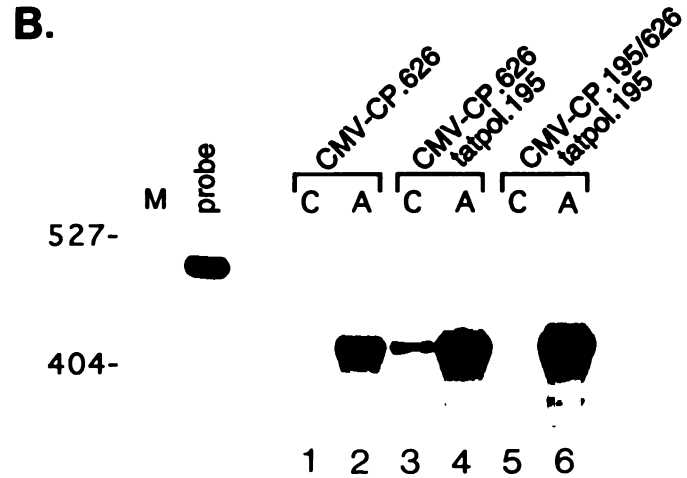
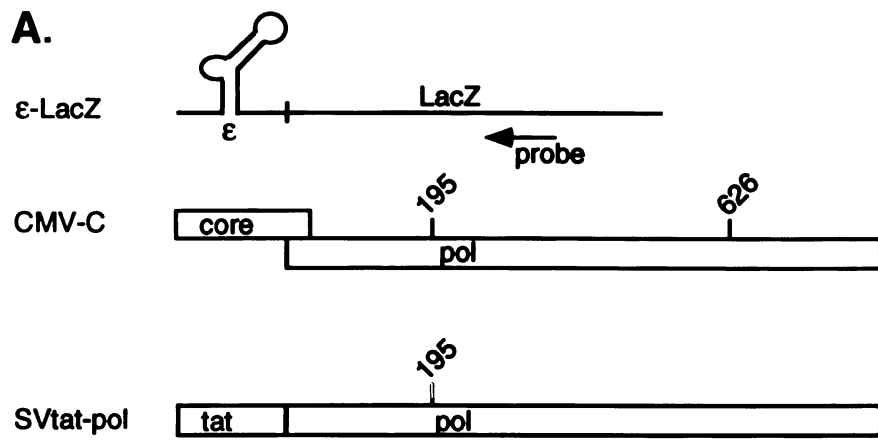


**FIG. 5  $\epsilon$  RNA encapsidation directed by doubly-mutated tat-polymerase.**

(a) Schematic representation of the RNA target pE-LacZ, core donor pCMV-C, and pSVtat-pol.195/626 which contains stop codons at both aa 195 and aa 626. (b) RNase protection analysis of total polyA+ and encapsidated RNA. Methods are as described in the legend of Fig 2b. M: size standards; probe: undigested LacZ probe; lanes 1-6: core RNA (C; lanes 1, 3, 5) and polyA+ RNA (A; lanes 2, 4, 6) from cells transfected with pE-LacZ and pCMV-C along with either pSVtat-pol.626 (lanes 1, 2), pSVtat-pol.195 (lanes 3, 4), or pSVtat-pol.195/626 (lanes 5, 6).



**FIG. 6  $\epsilon$  RNA encapsidation directed by tat-pol.195 and doubly-mutated core donor construct.** (a) Schematic representation of the RNA target pE-LacZ, the doubly-mutated core donor pCMV-CP195/626 containing stop codons at both aa 195 and aa 626 of polymerase, and pSVtat-pol.195. (b) RNase protection analysis of total polyA<sup>+</sup> and encapsidated RNA. Methods are as described in the legend of Fig 2b. M: size standards; probe: undigested LacZ probe; lanes 1-6: core RNA (C; lanes 1, 3, 5) and polyA<sup>+</sup> RNA (A; lanes 2, 4, 6) from cells transfected with pE-LacZ along with either pCMV-CP.626 (lanes 1, 2), pSVtat-pol.195 and pCMVCP.626 (lanes 3, 4), or pSVtat-pol.195 and pCMV-CP.195/626 (lanes 5, 6).





## REFERENCES

1. Bartenschlager, R., M. Junker-Niepmann, and H. Schaller. 1990. The P gene product of the hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J. Virol.* 64: 5324-5332.
2. Bartenschlager, R. and H. Schaller. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* 11: 3413-3420.
3. Carroll, R., L. Martarano, and D. Derse. 1991. Identification of lentivirus tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimeras. *J. Virol.* 65: 3460-3467.
4. Cullen, B. R. 1990. The HIV-1 tat protein: An RNA sequence-specific processivity factor? *Cell* 63: 655-657.
5. Frankel, A. D. 1992. Activation of HIV transcription by tat. *Curr. Opin. Genet. Develop.* 2: 293-298.
6. Hirsch, R., R. Colgrove, and D. Ganem. 1988. Replication of duck hepatitis B virus in two differentiated human hepatoma cell lines after transfection with cloned viral DNA. *Virology* 167: 136-142.
7. Hirsch, R., J. Lavine, L. Chang, H. Varmus, and D. Ganem. 1990. Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* 344: 552-555.
8. Hirsch, R., D.D. Loeb, J.R. Pollack, and D. Ganem. 1991. *Cis* - Acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. *J. Virol.* 65: 3309-3316
9. Junker-Niepmann, M., R. Bartenschlager, and H. Schaller. 1990. A short *cis*-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.* 9: 3389-3396.
10. Lavine, J., R. Hirsch, and D. Ganem. 1989. A system for studying the selective encapsidation of hepadnaviral RNA. *J. Virol.* 63: 4257-4263.
11. Nassal, M. 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J. Virol.* 66: 4107-4116.
12. Pollack, J.R., and D. Ganem. 1993. An RNA stem-loop structure directs hepatitis B virus RNA encapsidation. *J. Virol.* 67: 3254-3263.
13. Roychoudhury, A.F. Faruqui, and C. Shih. 1991. Pregenomic RNA encapsidation analysis of eleven missense and nonsense polymerase mutants of human hepatitis B virus. *J. Virol.* 65: 3617-3624.

14. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. **Molecular Cloning: A laboratory Manual**, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
15. Selby, M.J. and B.M. Peterlin. 1990. Trans-activation by HIV-1 tat via a heterologous RNA binding protein. *Cell* 62: 769-776.
16. Valenzuela, P., M. Quiroga, J. Zaldivar, P. Gray, and W.J. Rutter. 1980. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. *ICN/UCLA Symp. Mol. Cell. Biol.* 18: 57-70.
17. Wu, T.T., L.D. Condeary, L. Coates, C. Aldrich, and W. Mason. 1991. Evidence that less-than-full-length pol gene producers are functional in hepadnavirus DNA synthesis. *J. Virol.* 65: 2155-2163.

## **CHAPTER 4**

**DUCK HEPATITIS B VIRUS POLYMERASE BINDS AN RNA STEM-  
LOOP STRUCTURE TO INITIATE BOTH VIRAL RNA ENCAPSIDATION  
AND DNA SYNTHESIS**

**Jonathan R. Pollack<sup>1</sup> and Don Ganem<sup>2,3</sup>**

**Department of Biochemistry and Biophysics<sup>1</sup>, Howard Hughes Medical Institute<sup>2</sup>, and  
Departments of Microbiology and Immunology and Medicine<sup>3</sup>, University of California  
Medical Center, San Francisco, California 94143-0502.**

## **ABSTRACT**

Hepadnaviruses encode a polymerase (P) protein with key roles in both reverse transcription and genomic RNA encapsidation. Genetic analysis of *cis*-acting signals required for viral replication implicates an RNA stem-loop structure in both RNA packaging and the initiation of reverse transcription, a process in which P protein also serves as the primer. We now show that DHBV polymerase binds specifically and with high affinity to this RNA stem-loop structure. Mutational analysis indicates that all mutations in the RNA target that inhibit the P protein - RNA interaction inhibit both *in vivo* RNA packaging and *in vitro* DNA priming to comparable extents. However, certain mutations in the loop region of the RNA have minimal impact on P protein - RNA binding but are nonetheless severely defective for packaging and DNA synthesis. Thus, P protein - RNA complex formation is necessary but not sufficient to initiate these activities. In addition, examination of RNA binding by truncated P proteins indicates that the C-terminus of the protein, although required for RNA encapsidation *in vivo*, is dispensable for RNA binding and DNA priming.

## INTRODUCTION

Hepadnaviruses are small hepatotropic viruses that cause hepatitis and predispose carriers to the development of hepatocellular carcinoma. These viruses include the prototype human hepatitis B virus (HBV) along with the related woodchuck hepatitis virus (WHV), ground squirrel hepatitis virus (GSHV), duck hepatitis B virus (DHBV) and heron hepatitis B virus (HHBV). All hepadnaviruses contain 3-3.2 kb DNA genomes whose replication proceeds by reverse transcription of an RNA intermediate (Summers and Mason, 1982). Reverse transcription occurs principally within subviral core particles composed of the viral nucleocapsid (core) protein (C), polymerase (P), and the RNA template for the reaction, the RNA pregenome (pgRNA).

The hepadnaviral polymerase is required for several steps of viral replication. Besides its enzymatic functions in reverse transcription (reverse transcriptase, RNase H), polymerase is essential for the specific encapsidation of pgRNA, a role which is independent of its enzymatic activity (Bartenschlagger *et al.*, 1990; Hirsch *et al.*, 1990) and appears to require all of its subdomains [terminal protein, spacer, reverse transcriptase, RNase H (Bartenschlagger *et al.*, 1990; Hirsch *et al.*, 1990; Roychoudhury *et al.*, 1991)]. Polymerase also functions as a protein primer for reverse transcription (Wang and Seeger, 1992), resulting in the covalent linkage of a short DNA sequence (up to 4 nucleotides) to P protein (Wang and Seeger, 1993). This short DNA sequence is templated by RNA sequences within the encapsidation signal ( $\epsilon$ ). The newly made, P-linked DNA sequence must then dissociate from the template and reanneal at the 3' end of pgRNA where reverse transcription is continued.

The encapsidation signal ( $\epsilon$ ), present at the 5' end of pgRNA, includes an RNA stem-loop structure that mediates the selective packaging of this transcript (Junker-Niepmann *et al.*, 1990; Pollack and Ganem, 1993). Identical  $\epsilon$  sequences are also present

at the 3' end of pgRNA (due to the terminally redundant nature of the transcript) but cannot direct encapsidation from that position (Hirsch *et al.*, 1990; Junker-Niepmann *et al.*, 1990). The encapsidation signal consists of a lower stem, a 6 nucleotide bulge, and an upper stem with a 6-7 nucleotide loop (Junker-Niepmann *et al.*, 1990; Knaus and Nassal, 1993; Pollack and Ganem, 1993). In HBV, the stems and bulge are structural requirements for encapsidation while sequence specificity is important particularly in the loop (Pollack and Ganem, 1993). The  $\epsilon$  stem-loop structure is conserved phylogenetically among mammalian and avian hepadnaviruses (Junker-Niepmann *et al.*, 1990). However, DHBV  $\epsilon$  sequences are not sufficient to direct RNA encapsidation (Hirsch *et al.*, 1991); the presence of a second, auxiliary region of RNA located some 900 nts downstream of  $\epsilon$  is required to effect encapsidation (Calvert and Summers, personal communication).

Encapsidation of  $\epsilon$ -containing RNA requires polymerase. It has also been shown that encapsidation of the viral polymerase itself requires  $\epsilon$ -containing RNA (Bartenschlager and Schaller, 1992). Because of this coupling between the encapsidation of polymerase and  $\epsilon$ -containing RNA, and because the polymerase is believed to be present in a 1:1 stoichiometry with pgRNA (1-2 copies of each per virion), it has been suggested that encapsidation is initiated by P protein recognition of the encapsidation signal (Bartenschlager and Schaller, 1992). It has also been suggested that this same recognition mediates DNA priming (Wang and Seeger, 1993).

In this study we demonstrate that DHBV polymerase binds the DHBV encapsidation signal specifically and with high affinity. We also show by mutational analysis of the  $\epsilon$  stem-loop that the RNA requirements for polymerase binding correlate well with requirements for RNA encapsidation *in vivo* and DNA priming *in vitro*, with one notable exception: loop mutations which have little effect on P protein binding greatly reduce both packaging and priming. Finally we show that the C-terminus of polymerase, which is essential for RNA encapsidation, is dispensable for RNA binding and DNA priming.

## RESULTS

**Binding of DHBV polymerase to the encapsidation signal.** In order to demonstrate binding of DHBV polymerase to the encapsidation signal ( $\epsilon$ ), we performed a streptavidin-biotin mediated co-precipitation assay.  $^{35}\text{S}$ -methionine labelled DHBV polymerase was translated in rabbit reticulocyte lysate in the presence of synthetic DHBV  $\epsilon$  RNA transcripts into which biotinylated uridine had been incorporated. P protein/RNA complexes were then detected by precipitation of biotinylated RNA substrates with streptavidin-linked agarose beads. The RNA target was present in the reaction mix during translation to allow for the possibility of co-translational binding. [Pilot experiments verified that none of the RNA substrates had an inhibitory effect on polymerase translation *in vitro* (data not shown).]  $^{35}\text{S}$ -labelled DHBV polymerase translated in reticulocyte lysate appears as an 80-85 kD doublet (Fig. 1B, lane 2). The higher MW species corresponds to the full-length 768 aa protein. The lower MW species corresponds to *in vitro* translation initiation at a second in-frame methionine codon, 43 aa downstream (data not shown).

To demonstrate specific P protein/ $\epsilon$  interaction we performed a binding reaction with either a DHBV  $\epsilon$  ( $\text{d}\epsilon$ ) or an HBV  $\epsilon$  ( $\text{h}\epsilon$ ) RNA substrate. Although DHBV  $\epsilon$  and HBV  $\epsilon$  share phylogenetically conserved RNA structures [Fig. 1A, (Junker-Niepmann *et al.*, 1990)], DHBV and HBV polymerases *in vivo* direct encapsidation of only their own respective encapsidation signal (Pollack and Ganem, 1993; and data not shown). DHBV polymerase was precipitated in the presence of biotinylated DHBV  $\epsilon$  (Fig. 1B, lane 3), but not biotinylated HBV  $\epsilon$  (lane 4) or non-biotinylated DHBV  $\epsilon$  RNA substrate (lane 5). The faint signal that often appears with negative-control RNA substrates is due to non-specific binding of polymerase to the streptavidin-agarose beads (data not shown). Since the amount of polymerase protein loaded in lane 2 (Fig. 1B) represents one-tenth the amount analyzed in each binding reaction, and since binding occurs in RNA substrate excess (88 nM, see below), we estimate that 5-10% of the total polymerase protein produced *in vitro* is



functional for RNA binding. We have also found that polymerase- $\epsilon$  interaction will occur with the same efficiency and specificity when the RNA target is added post-translationally in the presence of 0.5 mM cycloheximide (not shown).

In order to determine the dissociation constant of the DHBV polymerase- $\epsilon$  interaction, we performed an RNA titration (Fig. 1C, inset) using increasing amounts of DHBV  $\epsilon$  RNA substrate. Polymerase binding was saturated at an RNA concentration of 177 nM, and half-maximal binding occurred at 14 nM RNA. Since a 10-fold reduction in the polymerase concentration did not significantly lower the RNA concentration at half-maximal binding (indicating that polymerase concentration is significantly below the RNA concentration at half-maximal binding; data not shown), this value represents a reasonable estimate of the dissociation constant ( $K_d$ ) for the P protein- $\epsilon$  interaction. This value has ranged from 8 nM to 26 nM upon replicate experiments.

**Binding of polymerase to  $\epsilon$  stem-loop mutants.** We next determined the RNA structure and sequence specificity of polymerase- $\epsilon$  interaction by assaying P protein binding to a series of RNA stem-loop mutants (Fig. 2, and summarized in Table 1).

To test the importance of the lower stem, we introduced 3 consecutive nucleotide changes on either the left (LowerL) or right side (LowerR) of this stem in order to disrupt potential base-pairing. In the doubly-mutated construct (LowerL/R) potential for base-pairing in this region is restored. Polymerase did not bind the LowerL substrate (Fig. 2, lane 5), while introduction of the compensatory changes in the right side of the stem restored binding (LowerL/R, lane 7). A small amount of binding (ca. 10% of the WT level) was observed with the LowerR RNA substrate (lane 6). Closer examination of LowerR revealed that the nucleotide substitutions introduced to disrupt the lower stem can be accommodated in an alternative lower stem structure by base-pairing to alternative partners in this region; such accommodation is not possible for LowerL (data not shown).

We next determined the consequences of mutating the bulge and loop for binding. Deletion of the six nucleotide bulge ( $\Delta$ bulge, lane 8) eliminated binding, while changing four consecutive residues in the bulge (Bulge2-5, lane 9) did not reduce binding. Altering residues 3 and 4 of the seven nucleotide loop (Loop3-4; lane 14) reduced polymerase binding approximately 4-fold, while a mutant altering residues 5 and 6 of the loop (Loop5-6; lane 15) bound P protein with wildtype efficiency. In other RNA-protein interactions, single unpaired nucleotides often contribute to binding specificity (Frankel *et al.*, 1991 ). DHBV  $\epsilon$  contains two unpaired U residues in the upper stem, and one unpaired U residue in the lower stem. Deleting one of the unpaired U residues in the upper stem ( $\Delta$ U1, lane 16) reduced binding 4- to 5-fold, while deleting either of the two other unpaired U residues ( $\Delta$ U2 and  $\Delta$ U3; lanes 17 and 18) had no effect.

**RNA packaging of  $\epsilon$  stem-loop mutants.** As RNA encapsidation in the human hepatitis B virus requires both structural and sequence-specific contributions of the HBV encapsidation signal (Pollack and Ganem, 1993), we wished to determine  $\epsilon$  structure and sequence requirements for DHBV RNA encapsidation. We therefore introduced the stem-loop mutations described above into the 5' copy of  $\epsilon$  in overlength DHBV genomes, which were then transfected into LMH avian hepatoma cells. Three days post-transfection we harvested from equivalent numbers of cells either total cellular polyA+ RNA or the RNA contained within purified cytoplasmic core particles. The RNA present in equal portions of each preparation was quantified by RNase protection using a riboprobe complementary to pgRNA sequences; the ratio of encapsidated to total polyA+ RNA is a measure of the packaging efficiency.

The results of an encapsidation assay are shown in Figure 3. The wildtype DHBV RNA pregenome was encapsidated, as a protected band of the expected size (361 nt) appears in sample lanes of both polyA+ RNA (lane 5) and RNA isolated from core particles (lane 4). The efficiency of encapsidation (ca. 50%) was comparable to our previously

reported results (Hirsch *et al.*, 1990). As expected, the reaction is P-protein dependent: a viral genome carrying an early frameshift mutation in polymerase (pol442, lanes 6, 7) was unable to encapsidate pgRNA.

For the most part, the packaging efficiency of our stem-loop mutants exhibited a strong correlation with the ability of these mutants to bind P protein. LowerL was not encapsidated (lanes 8, 9), while the compensatory mutations introduced in LowerL/R restored RNA encapsidation (lanes 12, 13) to near-wildtype levels; therefore we conclude that the lower stem is required for encapsidation. LowerR (lanes 10, 11) was encapsidated, although at 5-fold reduced levels, again the likely result of alternative base-pairing which retains a lower stem structure (as discussed above). Deletion of the bulge ( $\Delta$ bulge; lanes 14, 15) abolished RNA packaging, while changing residues within the bulge (Bulge2-5; lanes 16, 17) was tolerated. Both mutants containing substituted loop residues (Loop3-4; lanes 25, 26 and Loop5-6; lanes 27, 28) were not packaged, and of the single unpaired U residue deletion mutants, only  $\Delta$ U1 (lanes 29, 30) showed reduced in packaging efficiency (ca. 5-10% WT).

The features of  $\epsilon$  important for DHBV RNA encapsidation are nearly identical to those previously reported for HBV encapsidation (Pollack and Ganem, 1993). The RNA structure and sequence requirements for DHBV packaging also closely parallel those for DHBV P protein-RNA binding, with the notable exception of the loop mutants. This is particularly evident in Loop5-6, which bound polymerase at wildtype levels but was not encapsidated.

**DNA priming mediated by  $\epsilon$  stem-loop mutants.** It has been shown recently that P-protein-mediated priming of DNA synthesis occurs at RNA sequences within the encapsidation signal (Wang and Seeger, 1993). The sixth nucleotide of the bulge (C2576) within the motif 5'-UUAC-3' templates the covalent linkage of a dG to DHBV polymerase. Up to three more nucleotides are polymerized (pol-GTAA), followed by a translocation to

the 3' end of pgRNA where DNA synthesis continues. Because DNA priming occurs at RNA sequences within the encapsidation signal it has been suggested that polymerase might use a common RNA recognition event for the biochemically distinct steps of DNA priming and RNA encapsidation (Wang and Seeger, 1993). We therefore tested the ability of our stem-loop mutants to mediate DNA priming. Following cotranslational binding, a portion of the reaction was removed and incubated with  $^{32}\text{P}$ -dGTP; covalent  $^{32}\text{P}$ -dG-polymerase linkage was then detected by SDS-polyacrylamide gel electrophoresis.

Wildtype DHBV  $\epsilon$  mediated DNA priming: an 80-85 kD  $^{32}\text{P}$ -labelled species (actually a doublet) corresponding to P protein appears in Fig. 4, lane 1. HBV  $\epsilon$  did not support DNA priming by the DHBV polymerase (lane 2).  $^{32}\text{P}$ -linkage clearly results from the polymerase activity of P protein since DHBV  $\epsilon$  is unable to mediate DNA priming when the conserved reverse transcriptase motif YMHD of DHBV polymerase is mutated to YMHA (data not shown); this mutant is also non-functional for DNA synthesis *in vivo* (Chang *et al.*, 1990).

The ability of stem-loop mutants to support DNA priming correlated exactly with their *in vivo* RNA packaging function. LowerL (lane 3) was unable to mediate DNA priming, while this activity was restored in LowerL/R (lane 5). LowerR (lane 4) functioned in priming, though at 3-fold reduced levels; again, this activity is the probable result of an alternate lower stem formation. The presence of the bulge (lane 6) but not its specific sequence (lane 7) was required for primase activity. [Note that we intentionally left intact residue C2576 of the bulge which templates  $^{32}\text{P}$ -dG linkage.] Nucleotide substitutions in the loop (lanes 10, 11) drastically reduced priming activity (to ca. 1-5% WT levels), and deletion of only one of the single unpaired U residues ( $\Delta\text{U1}$ , lane 12) impaired priming, about 3 fold.

Although the structure and sequence requirements for DNA priming parallel those for RNA packaging *in vivo*, when correlated with the requirements for P protein-RNA interaction an instructive difference is observed. While all mutations that inactivate P

protein-RNA binding inactivate priming, loop mutations (especially Loop5-6) which do not affect RNA binding nonetheless inactivate priming.

**RNA binding by C-terminal truncations of polymerase.** *In vivo* it is believed that expression of the entire polymerase coding region, including all of its subdomains (terminal protein, spacer, reverse transcriptase, RNase H), is required to direct RNA encapsidation. We wondered whether the entire coding sequence was required for RNA binding *in vitro* or whether we could localize a discrete RNA binding domain of the protein. We therefore linearized the polymerase cDNA template in different locations so as to produce 3'-truncated RNAs encoding C-terminally deleted versions of P protein. The ability of these truncated P proteins to bind either DHBV  $\epsilon$  or HBV  $\epsilon$  was then determined using the streptavidin-biotin mediated precipitation assay (Fig. 5).

Full-length DHBV polymerase is a 768 aa protein. C-terminal truncations of polymerase at aa 728 (Fig. 5, lanes 8-10) and aa 568 (lanes 12-14) were competent to bind DHBV  $\epsilon$ . Truncations at aa 472 (lanes 15-17) and 371 (lanes 18-20) did not bind DHBV  $\epsilon$ . Truncations at aa 728 and aa 568, but not at aa 472, were also competent to prime DNA synthesis (data not shown). Therefore amino acid residues beyond position 568 appear dispensable for RNA binding and DNA priming.

The lower migrating species of the doublet in each of these binding reactions represents *in vitro* translation initiation from the second methionine residue (aa 44) of polymerase. Because this species binds RNA and primes DNA synthesis, we conclude that the N-terminal 43 amino acids of polymerase are also dispensable for these activities.

Encapsidation *in vivo* is not dependent on the polymerase activity of P protein, since P protein containing mutations that alter the conserved reverse transcriptase motif YMDD to YMHA (aa 511-514) directs RNA encapsidation but not DNA synthesis *in vivo* (Chang *et al.*, 1990; Hirsch *et al.*, 1990). *In vitro* this same polymerase mutant was competent to bind RNA (lanes 5-7), but does not prime DNA synthesis (data not shown).

Thus the polymerase activity of P protein is dispensable for both  $\epsilon$  binding and RNA encapsidation.

## **DISCUSSION**

**The polymerase-RNA interaction.** We demonstrate that DHBV polymerase binds specifically and with high affinity ( $K_d \cong 14$  nM) to DHBV  $\epsilon$ . That DHBV polymerase does not bind the structurally similar HBV  $\epsilon$  is a strong indication of the specificity of the interaction. Mutational analysis of the DHBV  $\epsilon$  stem-loop reveals that the requirements for binding include the lower stem and bulge structures, a single unpaired U residue (U1), and to a lesser extent loop nucleotides (i.e. Loop3-4). Binding is not dependent on the polymerase activity of P protein. Although the ability of the polymerase to covalently bind a G residue templated by the bulge indicates an intimate interaction between the P polypeptide and the RNA, we cannot formally exclude the possibility that the polymerase- $\epsilon$  interaction is indirect, i.e. that a cellular factor may participate in the binding. Experiments are underway to investigate this possibility. Since DHBV polymerase contains no currently known RNA binding motif (e.g. RNP motif, KH motif, RGG box, zinc finger, arg-rich; Mattaj, 1993), it is of great interest to characterize further the nature of this RNA-protein interaction.

**Polymerase binding initiates encapsidation.** Polymerase specifically binds the encapsidation signal, and the RNA structure and sequence specificity of this interaction correlates well with that of RNA encapsidation (except for the loop region, see below). Moreover, neither RNA binding nor RNA encapsidation requires the enzymatic activity of polymerase. Because the specificity of RNA recognition resides within polymerase, and because RNA lesions that block P binding block encapsidation, we conclude that polymerase binding to  $\epsilon$  initiates encapsidation. Core protein, which has only non-specific

RNA binding activity (Birnbaum and Nassal, 1990), likely recognizes the polymerase- $\epsilon$  complex to further viral assembly.

The ability of P protein to bind its RNA target provides a ready explanation for how hepadnaviruses ensure the incorporation of polymerase into the virion. Retroviruses solve this problem by synthesizing their polymerases as gag-pol fusion proteins; these are then incorporated into the capsid through interactions between the gag portion of the fusion and other gag subunits. In hepadnaviruses, however, polymerase is produced independently of core (Chang *et al.*, 1989; Schlicht *et al.*, 1989). But since P protein-RNA complex formation is required for packaging to proceed, the noncovalent interactions that direct the RNA into nucleocapsids will automatically co-deliver the polymerase, already bound to the origin of (-) strand DNA synthesis. However, the molecular details of this RNA encapsidation process are still unclear. A simple model would be that assembling C proteins might noncovalently interact with RNA-bound P protein (see below). If so, such interactions must not be possible with free polymerase, since  $\epsilon$  deletion mutants do not encapsidate P protein chains (Bartenschlager and Schaller, 1992).

In HBV,  $\epsilon$  sequences alone are sufficient to mediate encapsidation. This is not the case in DHBV, where an RNA region some 900 nt downstream of  $\epsilon$  is also required for RNA encapsidation (Hirsch *et al.*, JV and Calvert and Summers, personal communication). A similar requirement for auxiliary encapsidation sequences has been reported in several retroviruses (Linial and Miller, 1990). How these downstream sequences are involved in the encapsidation process is uncertain. Clearly, they are not required for pgRNA recognition by polymerase, since DHBV  $\epsilon$  alone can bind P protein. However, these downstream sequences may provide additional P protein binding sites, be involved in the recruitment of core protein or cellular factors, or even participate in RNA:RNA interactions that generate structures required for efficient encapsidation.

**Polymerase binding initiates DNA synthesis.** As discussed in the introduction, it has been recently reported that DNA priming by P protein occurs within  $\epsilon$  sequences, suggesting a possible link between the RNA packaging and DNA priming reactions (Wang and Seeger, 1993). We have now shown that the RNA stem-loop requirements for DNA priming *in vitro* exactly match those for RNA packaging *in vivo*. We therefore conclude that a single RNA recognition event mediates both of these functions.

RNA packaging *in vivo* is not dependent on prior DNA synthesis (Bartenschlagger *et al.*, 1990; Hirsch *et al.*, 1990). It is not known whether DNA priming *in vivo* is dependent on prior RNA encapsidation, although more extensive DNA elongation appears to be limited to encapsidated templates. Because *in vitro* DNA priming and synthesis can occur in the absence of core protein (Wang and Seeger, 1992), these events may occur either prior to or concomitantly with RNA packaging *in vivo*. The *cis*-preference polymerase exhibits for encapsidating (and therefore replicating) the pgRNA from which it has been translated suggests that all these events might in fact occur co-translationally *in vivo*. However, *in vitro* we observe no strict requirement for cotranslational binding of P protein to  $\epsilon$  sequences.

**Polymerase-RNA interaction is necessary but not sufficient for RNA packaging and DNA priming.**

All RNA mutations that we have tested that block polymerase-RNA binding block both packaging and DNA priming, strongly implying that stem-loop binding by polymerase is required for these reactions. However, the phenotype of loop mutations reveals that while P protein - RNA binding is necessary for these steps, it is not sufficient. The Loop3-4 and Loop5-6 mutants, which bind polymerase at 25% and 100% WT levels, respectively, demonstrate no detectable RNA packaging activity, and severely reduced (1-5% WT) levels of DNA priming. A critical role for loop nucleotides in HBV encapsidation has also been described (Pollack and Ganem, 1993).



Why might this be so? That loop mutants inhibit RNA packaging might seem at first to implicate a possible role of the loop in C protein recognition. However, loop mutations are equally disruptive to DNA priming, which occurs independently of core protein. We therefore propose that additional factor(s), perhaps of host origin, may be required to effect both the RNA packaging and DNA priming functions. It is attractive to speculate that such factor(s) might contact the loop region (nt 3-6) of  $\epsilon$ , thereby being brought into close proximity with bound P protein (Fig. 6). Such proposed factor(s) might then activate the RNA packaging and DNA priming activities of P protein, for example by inducing a conformational change or by catalyzing a covalent modification (or both). That such factors would be present in the rabbit reticulocyte lysate is consistent with the observation that hepadnaviral replication can take place in cells of many tissues and species if the promoter driving pgRNA expression, which is normally liver-specific, is replaced by a more broadly active heterologous promoter (Seeger *et al.*, 1989). Thus, most cells would have whatever host machinery might be required to support pgRNA packaging and reverse transcription. Precedents exist for the binding of both viral and cellular factors to RNA stem-loop structures involved in viral replication (Andino *et al.*, 1990; Marciniak *et al.*, 1990).

**C terminus of polymerase is dispensible for RNA binding but not RNA packaging.** By producing C-terminally truncated versions of polymerase *in vitro* we have demonstrated that the C terminus of polymerase (beyond aa 568) is not required for RNA binding and DNA priming. *In vivo*, however, several deletion mutations and even single missense mutations introduced in the C terminus of polymerase (beyond aa 568) abolish RNA packaging (Bartenschlagger *et al.*, 1990; Hirsch *et al.*, 1990; Roychoudhury *et al.*, 1991; Chen *et al.*, 1992). It is a formal possibility that all these mutations merely destabilize the P polypeptide *in vivo*. However, we favor a model in which the C terminus of polymerase, encompassing the RNase H domain of the protein, functions at a step in

RNA encapsidation subsequent to RNA binding. Perhaps this region of polymerase functions in contacting core protein to promote viral assembly. Because the same C-terminal deletions do not inactivate DNA priming, however, this region cannot be involved in interactions with the putative host factors mentioned above, unless separate factors participate in the activation of packaging and priming.

## **MATERIALS AND METHODS**

**Materials.** Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions. Radionucleotides were purchased from Amersham Corp.

**Plasmids.** DHBV nucleotide positions are numbered from the unique EcoRI site of DHBV 3 (Sprengel *et al.*, 1985). In this numbering scheme, nt 2530 is the transcription initiation site of pgRNA (Buscher *et al.*, 1985).

pT7dpol contains DHBV nts 170-3021 cloned into the SmaI site of the polylinker of pBS(-) (Stratagene). When linearized with AflII and transcribed *in vitro* with T7 RNA polymerase pT7dpol produces a transcript containing the DHBV polymerase coding region but no  $\epsilon$  sequences.

p $\epsilon$ BS was constructed by cloning DHBV nts 2526-2845 into the HindIII and NsiI sites of the polylinker of pBS(-). When linearized with EcoRV and transcribed *in vitro* with T3 RNA polymerase p $\epsilon$ BS produces a 132 nt transcript containing DHBV  $\epsilon$ . Stem-loop mutations (Table 1) were introduced into p $\epsilon$ BS by site-directed mutagenesis using standard methods (Kunkel *et al.*, 1987) and confirmed by DNA sequencing. pE-BS was constructed as previously described (Pollack and Ganem, 1993) and produces a 172 nt transcript containing HBV  $\epsilon$ , spanning HBV nts 1815-1986 [HBV adw2 (Valenzuela *et al.*, 1980; pgRNA initiates at nt 1815).

Overlength stem-loop mutant genomes used for encapsidation assays were constructed in two steps. First, the AflII-XbaI (nt 2526-2662) DNA fragment of D0.5G (nts 1658-3021 in pBS-) was replaced with the corresponding fragment (containing the appropriate mutation) of p $\epsilon$ BS. Second, an NsiI (nt 2845) genome monomer was cloned into the NsiI site of D0.5G in a head to tail orientation to produce an overlength D1.5G (nt

1658-3021/0-3021) mutant genome. The presence of the stem-loop mutation in final constructs was confirmed by DNA sequencing.

Polymerase mutant 442 (Chang *et al.*, 1990) contains a single nucleotide insertion resulting in the frameshifting and early termination of polymerase at aa 442; this mutant is non-functional for RNA packaging and DNA synthesis (Chang *et al.*, 1990; Hirsch *et al.*, 1990). Polymerase mutant YMHA (Chang *et al.*, 1990) contains a double missense mutation in the conserved reverse transcriptase motif YMDD (altered to YMHA); this mutant packages RNA but is non-functional for DNA synthesis (Chang *et al.*, 1990; Hirsch *et al.*, 1990).

pDHBV-RX was constructed by cloning DHBV nts 2662-3021 into the polylinker of pBS(-). When linearized with HindIII and transcribed *in vitro* with T7 RNA polymerase a riboprobe complementary to DHBV pgRNA can be produced.

***In Vitro* transcription.** RNAs for *in vitro* translation, binding and DNA primase assays were produced using the MEGAscript kit (AMBION, Inc.) according to the manufacturer's directions. For production of biotinylated RNAs, 0.75 mM Biotin-11-UTP (ENZO diagnostics, Inc, NY, NY), one-tenth of the non-biotinylated UTP concentration, was added to the transcription reaction.

***In vitro* translation/binding.** RNA-protein binding reactions were performed co-translationally. <sup>35</sup>S methionine-labelled DHBV polymerase was translated from 1 ug T7dpol RNA in a total volume of 25 ul using nuclease-treated reticulocyte lysate according to the manufacturer (Promega Biotec), in the presence of 100 ng biotinylated RNA substrate at 30° C for 90 min. All translation/binding reactions were carried out in the presence of at least a 20-fold molar excess of tRNA (from the 50 mg/ml calf thymus tRNA added to stimulate translation in the reticulocyte lysate, plus the endogenous reticulocyte tRNA pool).

**Streptavidin precipitation of polymerase-RNA complexes.** Following translation/binding, polymerase-RNA complexes were precipitated using streptavidin-agarose as described by Scherly *et al.* (1989). 500 ul Ipp150 [150 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.1% NP-40, 10 ug/ml yeast RNA, 0.04 U/ml RNasin (Promega), 1 mM Phenylmethylsulfonylfluoride] was added to the completed translation reaction along with 25 ul packed streptavidin-agarose beads (BRL). The solution was rocked at 4°C for 60 min, pelleted by a 15 sec centrifugation and washed three times with 500 ul Ipp150. The pellet was resuspended in 50-100 ul 2x SDS-PAGE sample buffer, vortexed briefly and boiled for 5 min. After a brief centrifugation, 10 ul of supernatant was loaded onto an 8% SDS-polyacrylamide gel. After electrophoresis, the gel was treated with Enhance (DuPont), dried and exposed to Kodak XOMAT film at -70°C overnight. Signal intensity was quantitated by phosphorimager (Molecular Dynamics).

***In vitro* DNA priming assays.** DNA priming reactions were performed essentially as described by Wang and Seeger (1992). Following *in vitro* translation/binding, 5 ul of the reaction was removed and added to 5 ul of a mixture containing 100 mM Tris-HCl (pH 7.5), 30 mM NaCl, 20 mM MgCl, 28 uM each dATP, dCTP and dTTP, and <sup>32</sup>P-dGTP (2.4 uM, 400 Ci/mmol). Reactions were incubated at 30°C for 30 min and stopped by the addition of 9 volumes 2x SDS-PAGE sample buffer. Samples were then boiled for 3 min, and 10 ul loaded on a 10% SDS-polyacrylamide gel. After electrophoresis, the gel was dried and subjected to autoradiography (-70°C overnight) as above.

**Cell Culture and Transfections.** LMH avian hepatoma cells were grown in H21/F12 (1:1) mix supplemented with 10% fetal calf serum and passaged every 2-3 days at a 1:3 dilution. DNA transfections were performed by the calcium phosphate coprecipitation method exactly as described previously (Hirsch *et al.*, 1988).

**RNA preparation and RNase protection assay.** PolyA+ total cellular RNA was purified 72 hrs post-transfection as previously described (Hirsch *et al.*, 1988). RNA within cytoplasmic core particles was isolated 72 hrs post-transfection as previously described (Lavine *et al.*, 1989) by polyethylene glycol precipitation of core particles followed by proteinase K digestion, phenol extraction and precipitation of nucleic acid. RNase protection analysis was performed as previously described (Hirsch *et al.*, 1990) on polyA+ RNA or core RNA prepared from equivalent numbers of transfected LMH cells (1/2 of a 60 mm plate). Synthesis of [ $\alpha$ -<sup>32</sup>P]CTP-labeled RNA probes was carried out as previously described (Hirsch *et al.*, 1990); Probe DHBV-RX is a 388 nt riboprobe with 27 nonhybridizing polylinker nts and 361 nts of DHBV complementary to pgRNA, spanning nts 2662-3021.

#### **ACKNOWLEDGMENTS**

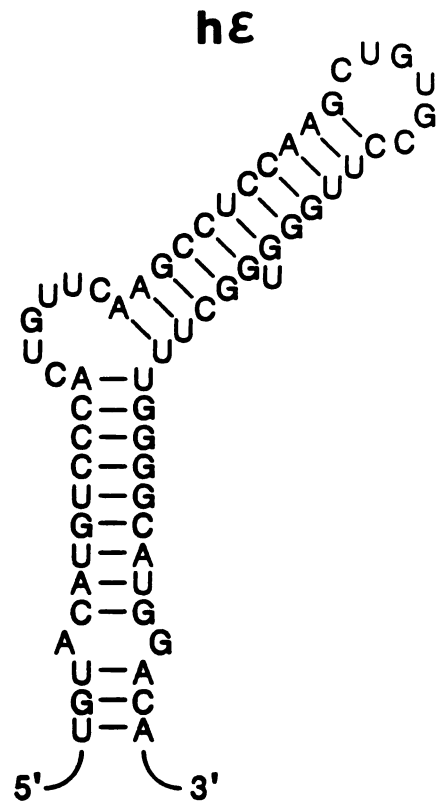
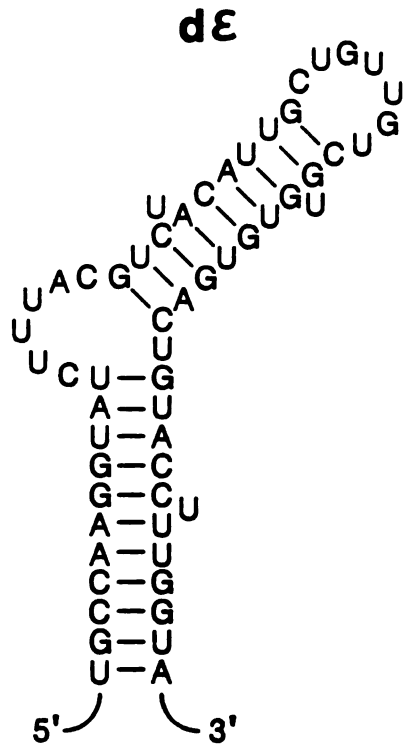
This work was supported by grants from the National Institutes of Health. We wish to thank Sandy Johnson, Steve Chuck, Chris Chang, Sophie Roy and Ian Taylor for helpful discussions, and Sophie Roy and Ian Taylor for critical reading of the manuscript.

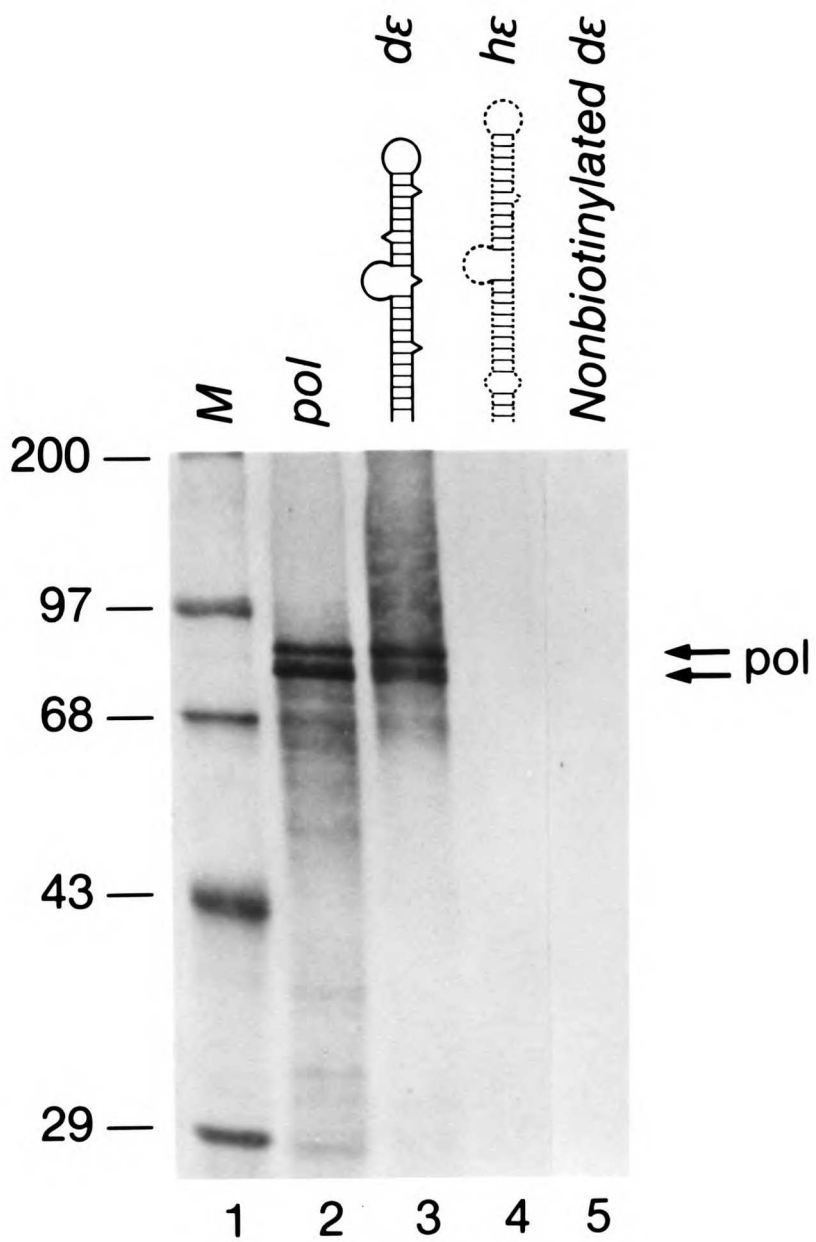
**TABLE 1 DHBV  $\epsilon$  stem-loop mutations.** The primary sequence of the wildtype (WT) DHBV  $\epsilon$  stem-loop region is shown at top, annotated above with arrows indicating the sequence components of the upper stem, lower stem, loop and bulge. Subsequent rows show the mutant designations along with their nucleotide changes. (.) represents no change of nucleotide; (-) represents deletion of a nucleotide. The RNA binding, RNA packaging, and DNA priming efficiency of mutants is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable signal.

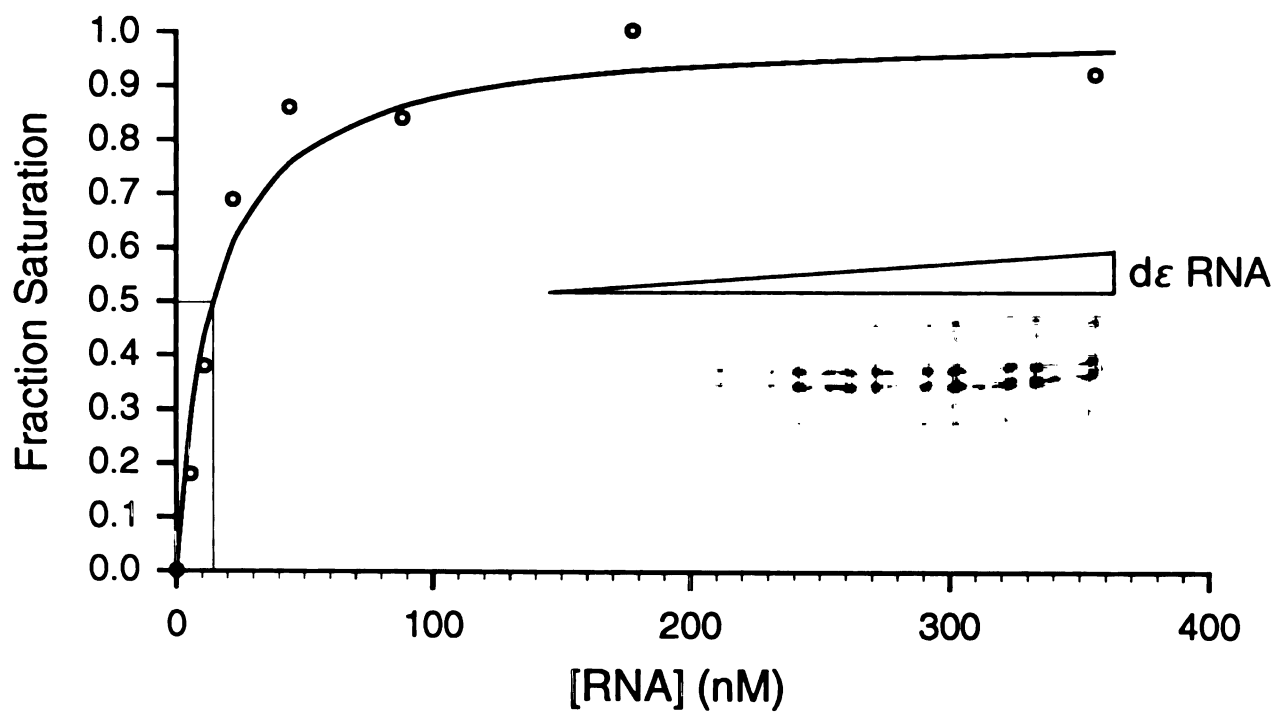
mutation	RNA binding			RNA packaging			DNA priming		
	RNA binding	RNA packaging	DNA priming	lower stem	bulge	upper stem	loop	upper stem	lower stem
WT	++	++	++	UGCCAAGGUAUCUUUACGUCUACAUGCUGUUGUCGUGUGACUGUACCUUGGUA					
LowerL	-	-	-	.....CAU.....					.....AUG.....
LowerR	+/-	+	+	.....CAU.....					.....AUG.....
LowerL/R	++	++	++	.....CAU.....					.....AUG.....
Δbulge	-	-	-	.....-----.....					.....AUG.....
Bulge2-5	++	++	++	.....CGCU.....					.....AUG.....
Loop3-4	+	-	+/-	.....AC.....					.....AUG.....
Loop5-6	++	-	+/-	.....CA.....					.....AUG.....
ΔU1	+	+/-	+	.....					.....AUG.....
ΔU2	++	++	++	.....					.....AUG.....
ΔU3	++	++	++	.....					.....AUG.....



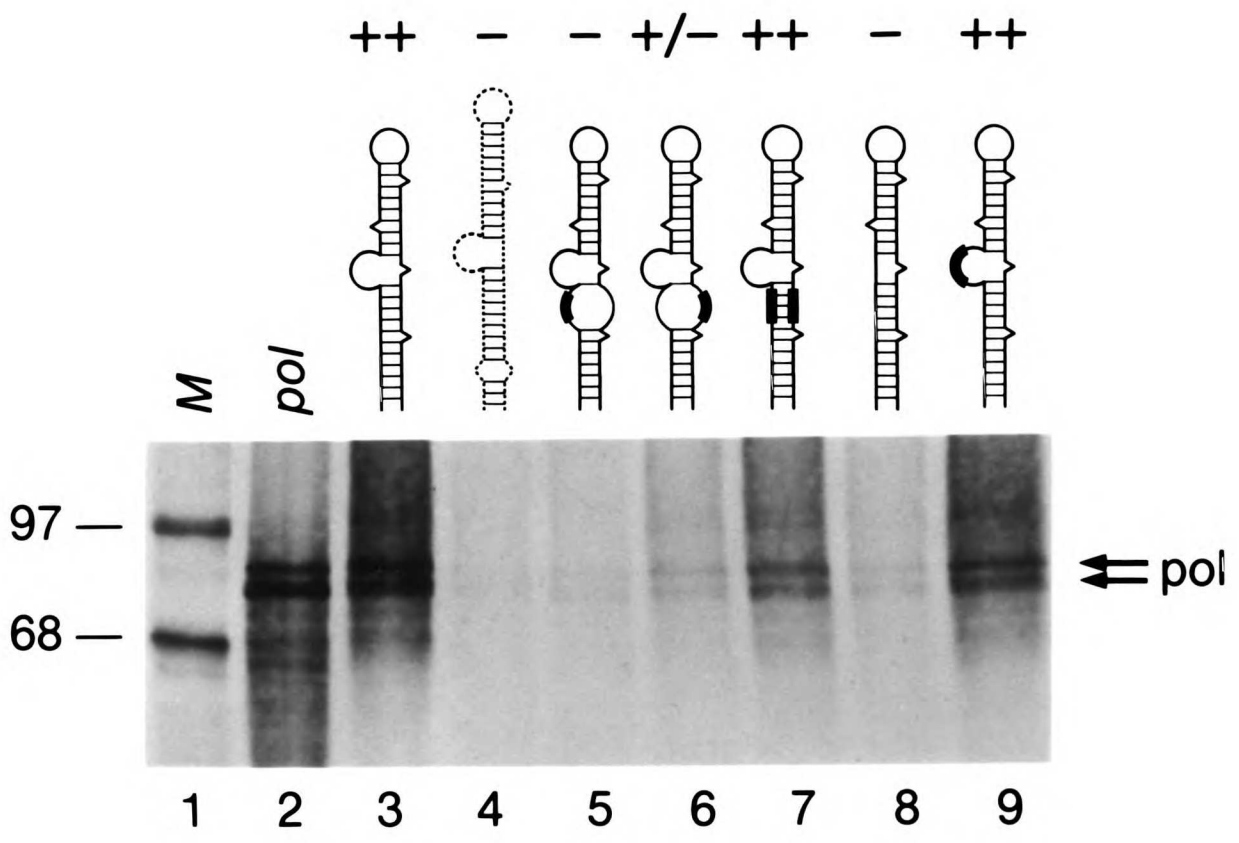
**FIG. 1 DHBV Polymerase-RNA binding assay.** (A) DHBV (DHBV 3)  $\epsilon$  and HBV (HBV adw2)  $\epsilon$  RNA structures based on phylogenetic analysis and enzymatic probing (Junker-Niepmann *et al.*, 1990; Knaus and Nassal, 1993; Pollack and Ganem, 1993). (B)  $^{35}\text{S}$ -labelled DHBV polymerase protein was translated in reticulocyte lysate in the presence of substrate RNAs containing biotinylated U residues. Polymerase-RNA complexes were precipitated with streptavidin-agarose beads, released by boiling in SDS-PAGE sample buffer, and analyzed by SDS-PAGE electrophoresis. Lane 1: MW size standards; lane 2: one-tenth total protein analyzed per binding assay; lanes 3, 4, 5: polymerase binding to DHBV  $\epsilon$  (d $\epsilon$ ), HBV  $\epsilon$  (h $\epsilon$ ), and non-biotinylated DHBV  $\epsilon$ , respectively. (C) RNA titration of polymerase binding with increasing amounts of biotinylated DHBV  $\epsilon$  RNA, ranging from 0 to 355 nM. Fraction saturation is defined as the ratio of protein bound at each RNA concentration to the maximal protein bound. The curve drawn was determined by the best fit of the equation: fraction saturation =  $[\text{RNA}] / (k_d + [\text{RNA}])$  using the non-linear least squares method furnished in Sigmaplot (Jandel Scientific, Corte Madera, CA). The calculated  $K_d$  for the line is 14 nM and the point of 50% fractional saturation is indicated.

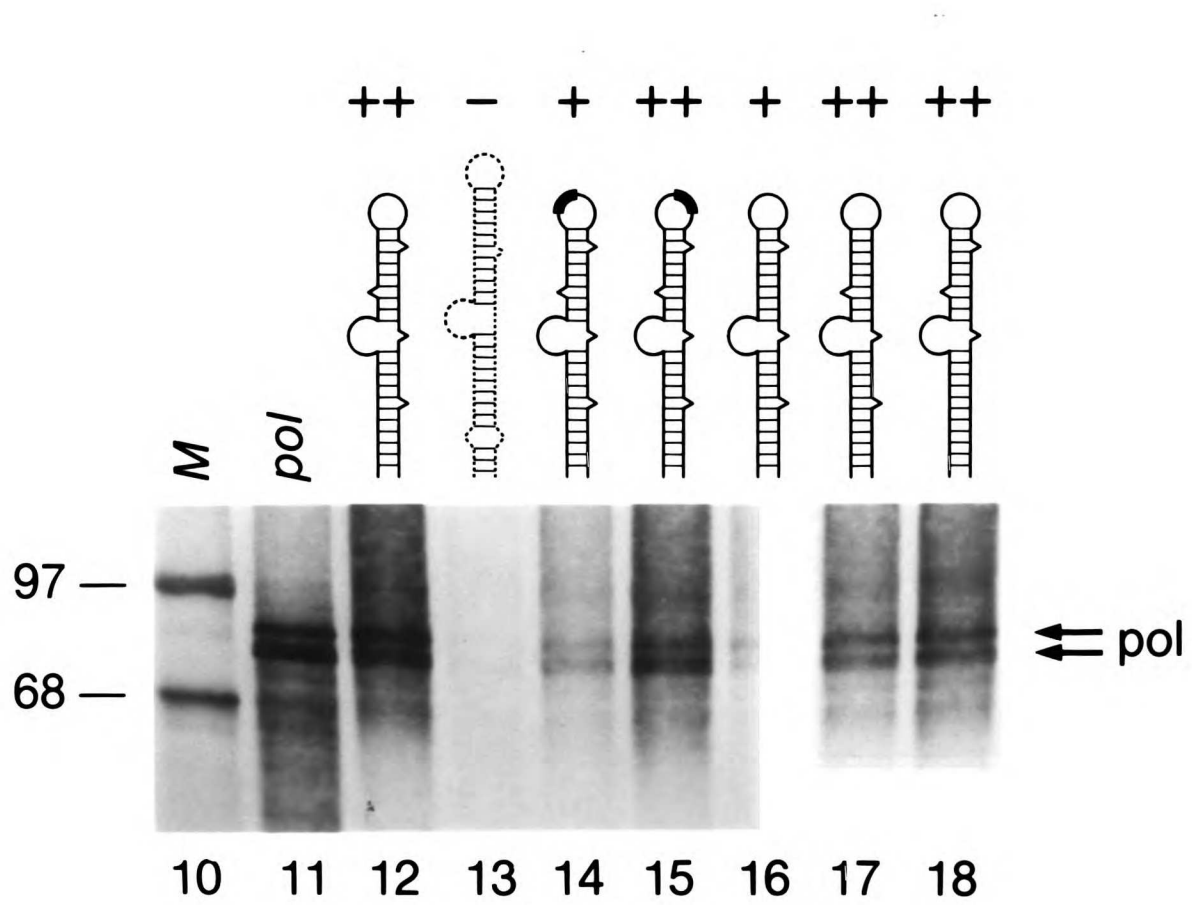






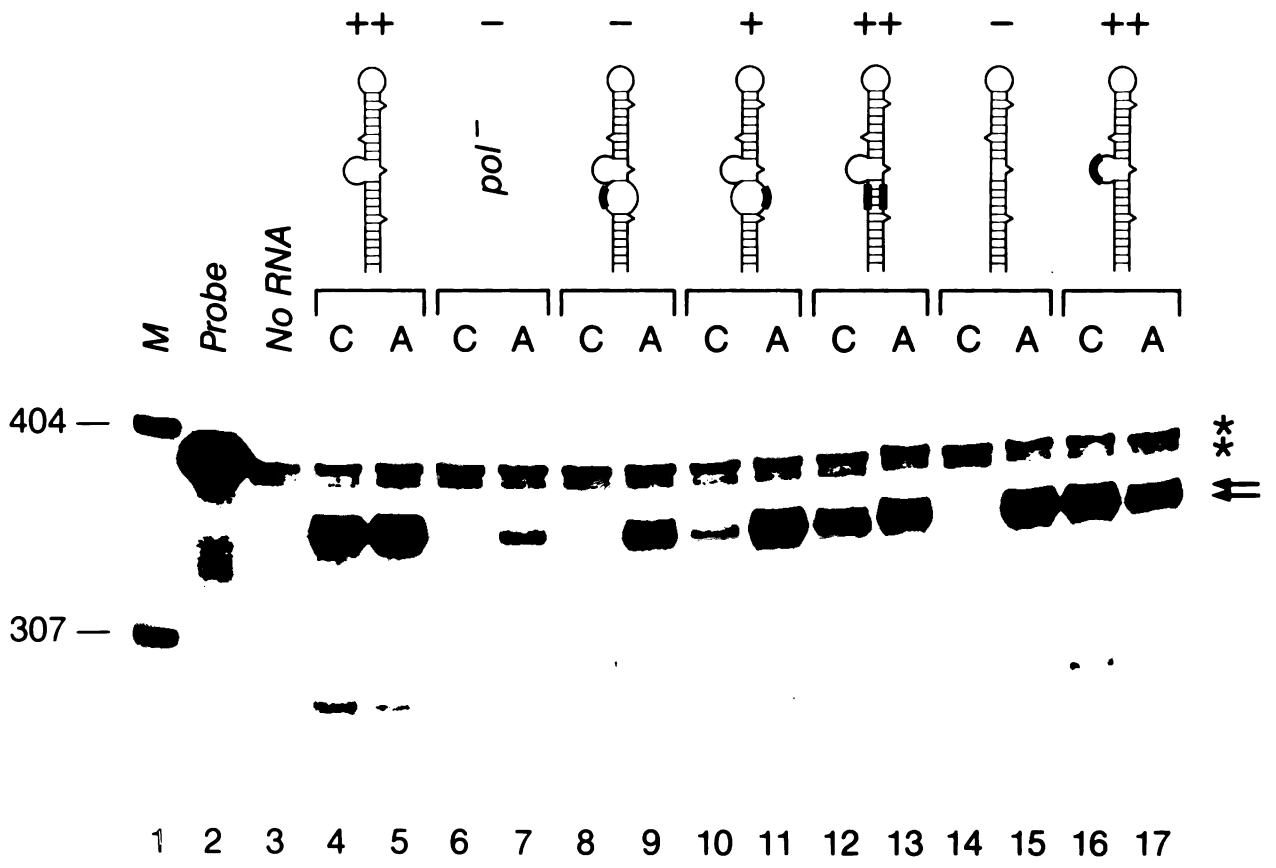
**FIG. 2 RNA binding assay of DHBV  $\epsilon$  stem-loop mutants.** Steptavidin-biotin based precipitation assays on mutant RNA substrates were performed as described in the legend to Fig. 1B. Lanes 1, 10: MW standards; lanes 2, 11: one-tenth total protein analyzed per binding assay; lanes 3, 12 and lanes 4, 13: polymerase binding to DHBV  $\epsilon$  and HBV  $\epsilon$  respectively; lanes 5, 6, 7, 8, 9, 14, 15, 16, 17, 18: polymerase binding to LowerL, LowerR, LowerL/R,  $\Delta$ bulge, Bulge2-5, Loop3-4, Loop5-6,  $\Delta$ U1,  $\Delta$ U2, and  $\Delta$ 3 respectively. The relative binding efficiency of mutants is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -: binding level not above negative control (HBV  $\epsilon$ ) RNA.

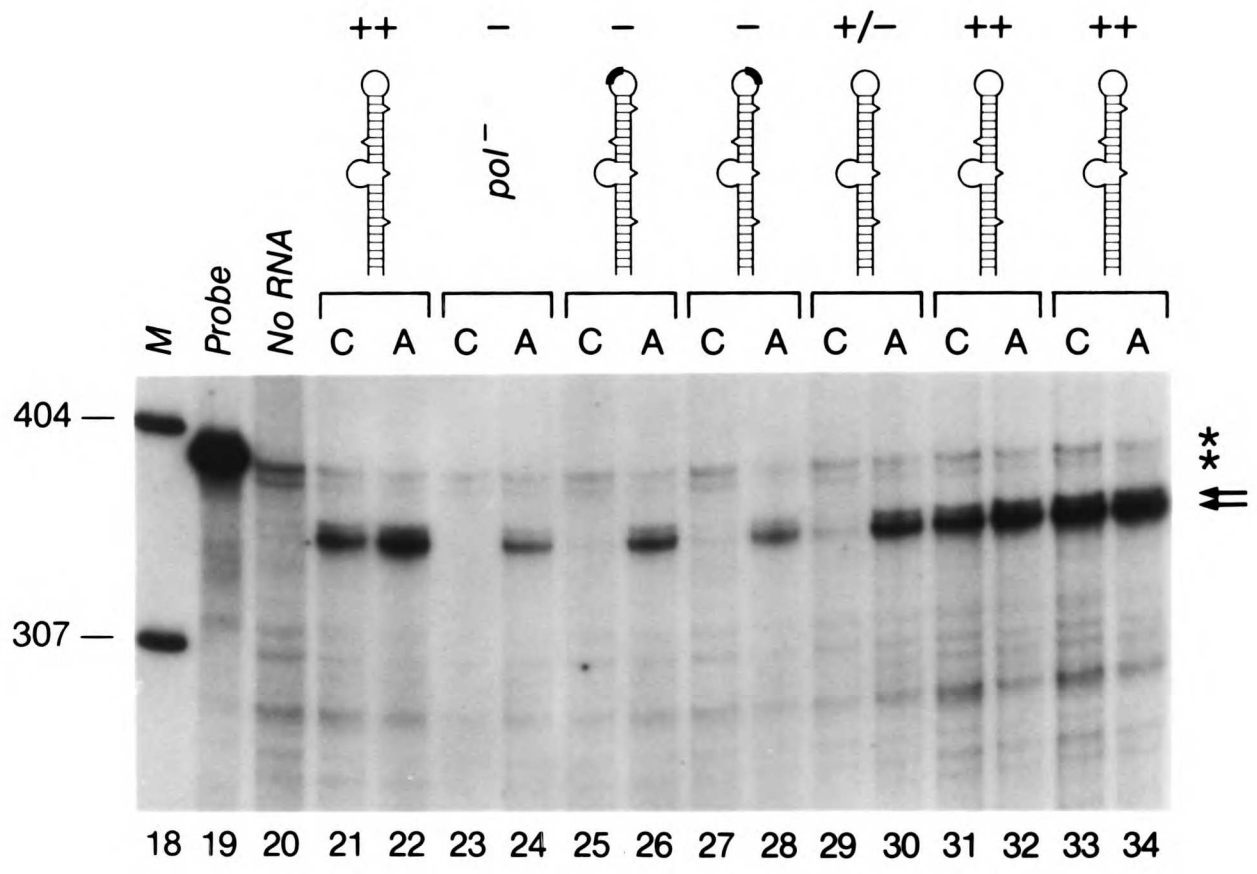




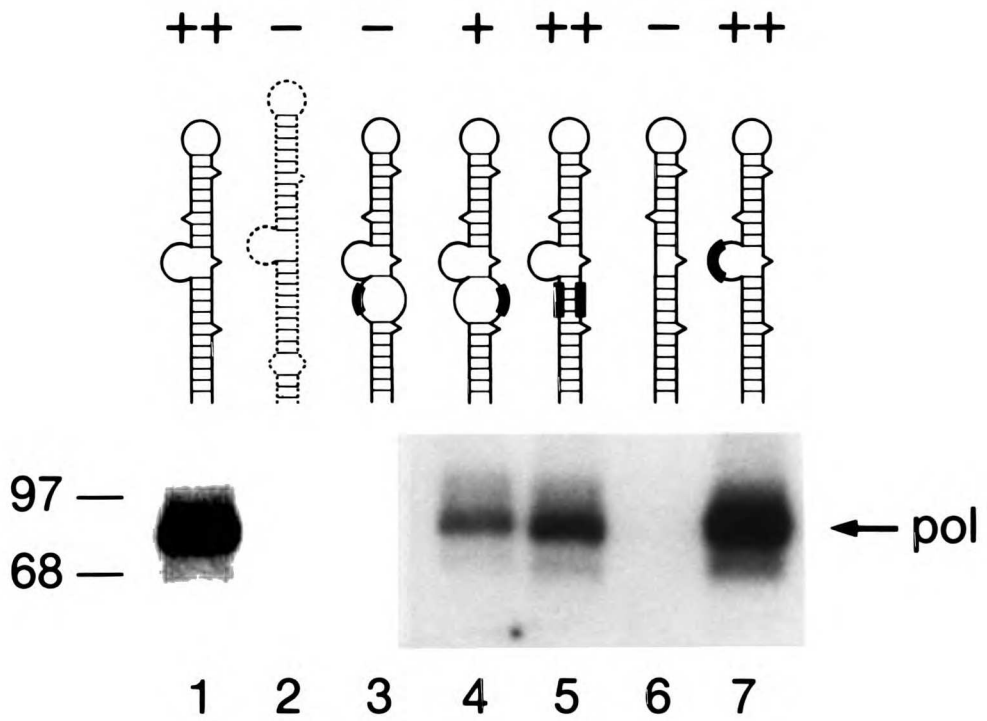
**FIG. 3 Encapsidation assay of DHBV  $\epsilon$  stem-loop mutants.** From equivalent numbers of transfected cells we isolated either total poly(+) RNA or RNA extracted from cytoplasmic core particles. RNA from equal portions of each preparation was quantified by RNase protection using a DHBV antisense riboprobe that, when annealed to complementary DHBV pg RNA sequence and digested with RNase, protects a 361 nt fragment (which appears as a doublet). Lane 1, 18: MW standards; lanes 2, 19: undigested probe; lanes 3, 20: probe digested in the absence of input RNA; lanes 4/5, 6/7, 8/9, 10/11, 12/13, 14/15, 16/17, 21/22, 23/24, 25/26, 27/28, 29/30, 31/32, and 33/34: core RNA (C)/ poly A(+) RNA (A) from cells transfected with WT, pol442, LowerL, LowerR, LowerL/R,  $\Delta$ bulge, Bulge2-5, WT, pol442, Loop3-4, Loop5-6,  $\Delta$ U1,  $\Delta$ U2, and  $\Delta$ 3 respectively. Arrows indicate relevant protected fragment; asterisks indicate incompletely digested input probe. The encapsidation efficiency of mutants is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -:no detectable encapsidated RNA.



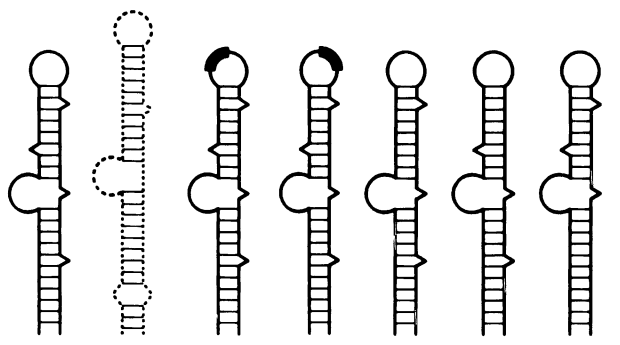




**FIG. 4 DNA priming mediated by DHBV  $\epsilon$  stem-loop mutants.** Following translation, polymerase and RNA template were incubated in the presence of  $^{32}\text{P}$ -dGTP. The covalently linked  $^{32}\text{P}$ -dG-polymerase product of the priming reaction was analyzed by SDS-PAGE. Lanes 1, 8 and lanes 2, 9: priming reaction mediated by DHBV  $\epsilon$  and HBV  $\epsilon$  respectively; lanes 3, 4, 5, 6, 7, 10, 11, 12, 13, and 14: priming reaction mediated by LowerL, LowerR, LowerL/R,  $\Delta$ bulge, Bulge2-5, Loop3-4, Loop5-6,  $\Delta$ U1,  $\Delta$ U2, and  $\Delta$ 3 respectively. The relative binding efficiency of DNA priming is indicated as follows: ++: 50-100% of WT, +: 10-50% of WT, +/-: 1-10% of WT, -: no detectable priming.



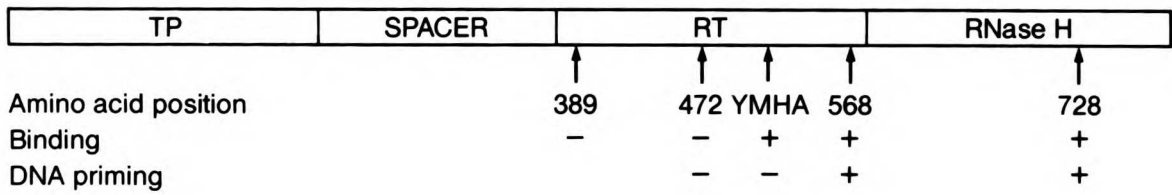
++ - +/- +/- + ++ ++



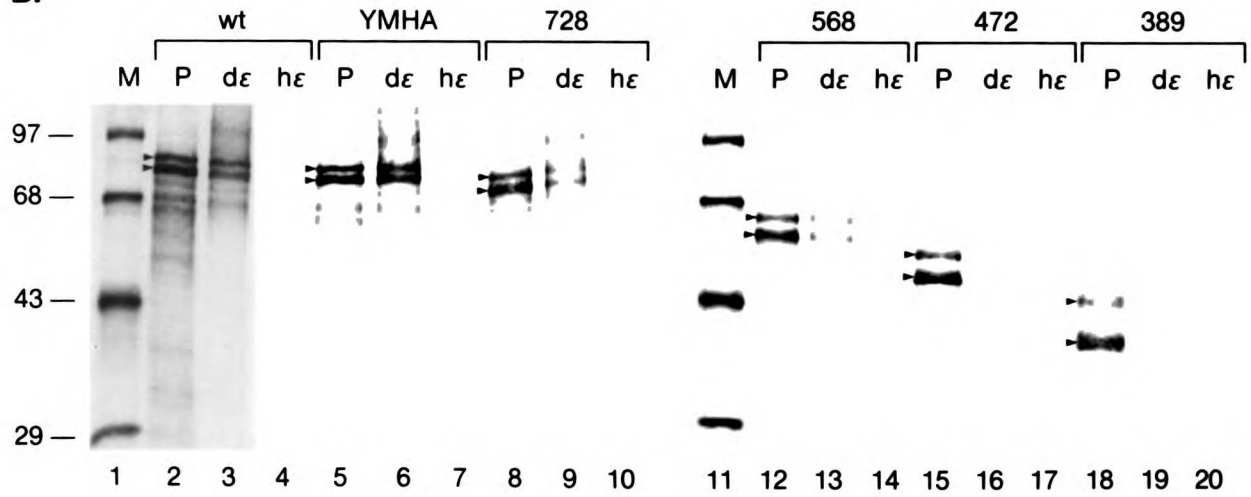
8 9 10 11 12 13 14

**FIG. 5 RNA binding of C-terminally truncated polymerase proteins. (A)** Schematic representation of the domain structure of the 786 aa DHBV polymerase protein. Polymerase protein was truncated at aa positions 728, 568, 472, and 389, as indicated by arrows. YMHA (aa 513/514) is a double missense mutation in the conserved reverse transcriptase motif YMDD. Below each truncation is indicated its ability to bind RNA and prime DNA synthesis. (B) Truncation products were tested for binding to DHBV  $\epsilon$  and HBV  $\epsilon$  using the streptavidin-biotin based precipitation assay described in the legend to Fig. 1B. Lanes 1, 11: MW standards; lanes 2-4, 5-7, 8-10, 12-14, 15-17, and 18-20: binding by WT polymerase, polYMHA, pol728, pol568, pol472, and pol389, respectively. P lanes indicate one-tenth total protein analyzed per binding reaction; d $\epsilon$  and h $\epsilon$  lanes indicate binding to DHBV  $\epsilon$  and HBV  $\epsilon$ , respectively. Major truncation products are indicated by arrowheads in each P lane.

**A.**



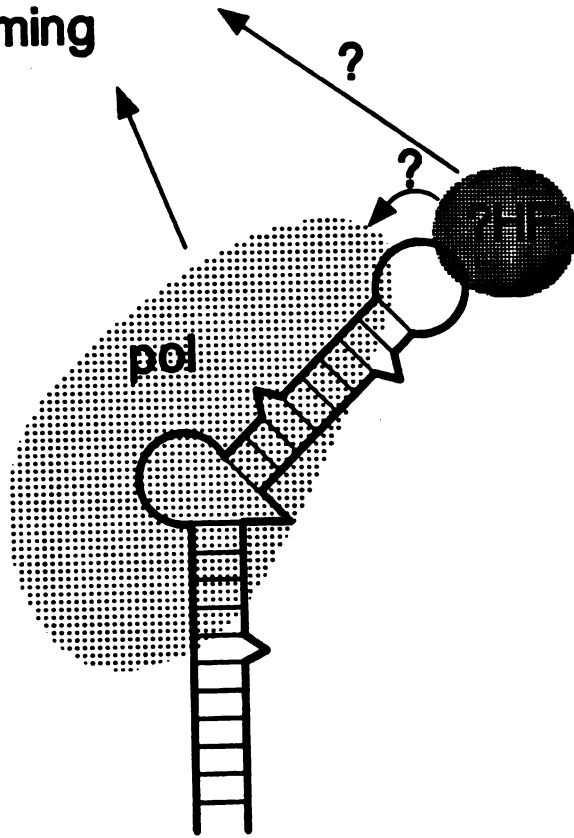
**B.**



**FIG. 6 Polymerase- $\epsilon$  interaction initiates viral RNA encapsidation and DNA synthesis.** Polymerase (pol) is shown binding the  $\epsilon$  stem-loop RNA structure. A proposed host factor (HF) is diagrammed contacting the loop region. This proposed complex mediates RNA encapsidation and DNA priming. A detailed explanation of the model is contained in the text.



- RNA packaging
- DNA priming



## REFERENCES

- Andino,R., Reickhof,G.E. and Baltimore,D. (1990) *Cell*, **63**, 369-380.
- Bartenschlager,R., Junker-Niepmann,M. and Schaller,H. (1990) *J. Virol.* ,**64**, 5324-5332.
- Bartenschlager,R. and Schaller,H. (1992) *EMBO J.* ,**11**, 3413-3420.
- Birnbaum,F.and Nassal,M. (1990) *J. Virol.*, **64**,3319-3330.
- Buscher,M.W.,Reiser,W.,Will,H. and Schaller,H. (1985) *Cell* ,**40**,717-724.
- Chang,L.-J., Pryciak,P.,Ganem,D. and Varmus,H.E. (1989)*Nature* ,**337**,364-368.
- Chang,L.-J., Hirsch,R., Ganem,D. andVarmus,H.E. (1990) *J. Virol.*,**64**,5553-5558.
- Chen,Y.,Robinson,W.S. and Marion,P.L. (1992) *J. Virol.*,**66**,1282-1287.
- Chiang,P.-W.,Jeng,K.S., Hu,C.P. and Chang,C. (1992)*Virology* ,**186**, 701-711.
- Frankel,A. D.,Mattaj, I.W. and Rio,D.C. (1991) *Cell*,**67**,1041-1046.
- Hirsch,R., Colgrove,R.and Ganem,D. (1988)*Virology*, **167**,136-142.
- Hirsch,R.,Lavine,J.,Chang,L.,Varmus,H. and Ganem,D.(1990) *Nature*,**344**,552-555.
- Hirsch,R.,Loeb,D.D.,Pollack,J.R.and Ganem,D. (1991) *J. Virol.* ,**65**,3309-3316
- Junker-Niepmann,M.,Bartenschlager,R. and Schaller,H.(1990) *EMBO J.* ,**9**,3389-3396.
- Knaus,T. and Nassal,M. (1993) *Nuc. Acid. Res.*,**21**, 3967-75.
- Kunkel,T.A.,Roberts,J.D. and Zavour,R.A. (1987) *Meth. Enzymol.*,**154**,376-382.
- Lavine,J., Hirsch,R. and Ganem,D. (1989) *J. Virol.*,**63**,4257-4263.
- Linial,M.L. and Miller, A.D. (1990) *Curr. Top. Microbiol. Immunol.* ,**157**, 125-152.
- Mattaj,I. W. (1993) *Cell* ,**73**, 837-840.
- Marciniak,R.A.,Garcia-Blanco,M.A. and Sharp,P.A. (1990) *PNAS* ,**87**,3624-3628.
- Nassal,M. (1992) *J. Virol.*,**66**, 4107-4116.
- Pollack,J. R. and Ganem, D. (1993) *J. Virol.*,**67**, 3254-3263.
- Roychoudhury,A.,Faruqui,F. and Shih,C. (1991) *J. Virol.*,**65**, 3617-3624.
- Scherly,D.,Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.* ,**8**, 4163-4170.

- Schlicht,H.-J.,Radziwill,G. and Schaller,H. (1989)*Cell* ,**56**,85-92.
- Seeger,C., Baldwin, B. and Tennant, B.C. (1989) *J. Virol.* ,**63**, 4665-4669.
- Sprengel,R.,Kuhn,C.,Will,H. and Schaller,H. (1985) *J. Med.Virol.*,**15**,323-333.
- Summers,J. and Mason,W.S. (1982) *Cell* ,**29**,403-415.
- Valenzuela,P.,Quiroga,M., Zaldivar,J.,Gray,P. and Rutter,W.J. (1980) *ICN-UCLA Symp. Mol. Cell Biol.*,**18**,57-70.
- Wang,G.-H. and Seeger, C. (1992) *Cell* ,**71**, 663-670.
- Wang,G.-H. and Seeger, C. (1993) *J. Virol.*,**67**,6507-6512.

## **CHAPTER 5**

## CONCLUSION

In this thesis I have characterized the hepadnaviral packaging signal and defined protein-RNA interactions required for the initiation of RNA packaging.

Using nucleases to probe the structure of  $\epsilon$  *in vitro* and *in vivo*, I have determined that HBV  $\epsilon$  forms an RNA stem-loop structure with a lower stem, a six nucleotide bulge, and an upper stem with a six nucleotide loop. This structure is consistent with phylogenetic predictions. By mutating the stem-loop sequences and assaying levels of encapsidated RNA in transfected hepatoma cells, I have determined the structure and sequence requirements for RNA packaging. These include a structural role of the lower and upper stem, the presence of the bulge independent of its sequence, and specific sequences particularly in the loop.

Although *in vivo* attempts to demonstrate P protein -  $\epsilon$  interactions using tat-polymerase chimeras were largely unsuccessful, I was able to successfully detect  $\epsilon$  recognition by P protein *in vitro*. Using as an assay the streptavidin-mediated co-precipitation of  $^{35}\text{S}$ -labelled DHBV polymerase (expressed in reticulocyte lysate) and biotinylated RNA substrate, I have demonstrated a specific and high affinity ( $K_d \cong 14 \text{ nM}$ ) interaction between P protein and  $\epsilon$ . Mutational analysis of DHBV  $\epsilon$  reveals the requirements of binding to include the lower stem and bulge structures, independent of sequence. I have also demonstrated a strict dependence of the RNA packaging reaction on binding: all mutations that disrupt binding eliminate packaging. Because the specificity of RNA packaging resides within P protein, I conclude that RNA packaging is initiated by P protein recognition of  $\epsilon$ .

As discussed in the introduction, it has been recently reported that DNA priming by P protein occurs within  $\epsilon$  sequences (8), suggesting a possible link between RNA packaging and DNA priming. I have now demonstrated that the RNA stem-loop requirements for DNA priming *in vitro* exactly match those for RNA packaging *in vivo*,

supporting the suggestion that a single RNA recognition event mediates both of these functions. This is yet one further example of the remarkable parsimony displayed by this small yet complex virus. While it is known that packaging *in vivo* is not dependent on prior DNA synthesis, it remains to be determined whether DNA priming might precede or occur concomitantly with packaging *in vivo*.

I have also shown that mutations in the loop region that have little impact on binding drastically reduce RNA packaging and DNA priming. From this result I conclude that P protein binding, while necessary, is not sufficient to mediate these functions. I speculate that a protein of host origin may interact with loop sequences to activate both RNA packaging and DNA priming functions. Precedents exist for the co-occupation of RNA structures involved in viral replication by viral and cellular factors, for example on HIV-1 TAR (6) and at the 5' end of the polio virus genome (1).

Finally, I show that the C terminus of P protein, while required for RNA packaging *in vivo*, is dispensable for RNA binding and DNA priming *in vitro*. I suggest that the C terminus of polymerase is required to mediate a step in packaging subsequent to RNA binding, perhaps in the recognition of core protein to promote nucleocapsid assembly.

These findings invite instructive comparisons with other agents. In retroviruses the nucleocapsid region of the gag precursor polypeptide mediates specific recognition of the packaging signal. Retroviral pol, produced as a gag-pol fusion, is targeted to virions by its covalent linkage to gag, and encapsidation of pol and the viral RNA occur independently of one another.

In hepadnaviruses, polymerase is not produced as a fusion with core, and packaging of P protein and viral RNA are co-dependent. Now I have demonstrated that P protein mediates the specific recognition of the packaging signal. Requiring polymerase interaction with  $\epsilon$  to initiate packaging ensures its co-packaging with the viral RNA, thus

solving the problem of encapsidating a polymerase that is not produced as a fusion with the capsid protein.

The double-stranded RNA virus of yeast, known as L-A, shares several features in common with retroviruses. In the L-A viral particle plus strand RNA is packaged along with an RNA-dependent polymerase. The packaging signal has been defined as a discrete RNA stem-loop structure on plus strand RNA (4). The polymerase is synthesized as a gag-pol precursor by ribosomal frameshifting (2), and is therefore likely incorporated into viral particles through gag/ gag-pol interactions. However, viral RNA recognition appears to be mediated through a discrete region of the pol portion of the gag-pol fusion (3, 5), and polymerase expression is required for RNA packaging (5).

Therefore the yeast L-A virus has evolved a hybrid packaging mechanism, using features of both retroviruses and hepadnaviruses. Like retroviruses, pol is targeted to the viral particles by virtue of its fusion to the capsid protein. However, like hepadnaviruses, selective recognition of the RNA genome is mediated by polymerase (cf. Fig. 1).

**Future directions.** Many new experimental directions are suggested by the results I have described. First, the nature of the P protein- $\epsilon$  interaction needs to be characterized in more detail. For example, it remains to be seen whether the interaction of P protein with  $\epsilon$  is direct or indirect. Since in our system the binding always occurs in a reticulocyte lysate, it is formally possible that a host protein is required to mediate P protein -  $\epsilon$  interaction. Since P protein must intimately contact the bulge region of the structure, which functions in DNA priming to template the covalent linkage of a short DNA sequence to P protein, a direct interaction of P protein with  $\epsilon$  seems most likely.

Also, since P protein contains no previously described RNA binding motifs (7), it is of great interest to characterize the molecular details of P protein-RNA interaction. Clearly the C terminus of polymerase is dispensable for  $\epsilon$  binding, but further mutational studies will be needed to define the minimal binding domain

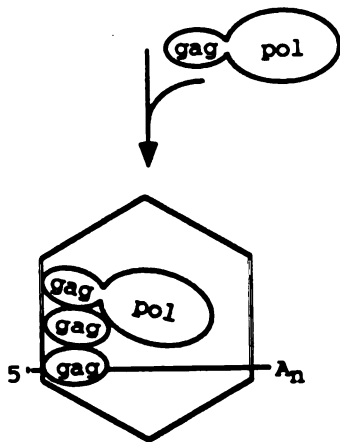
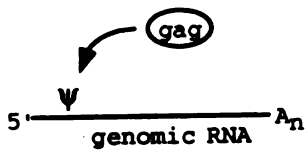
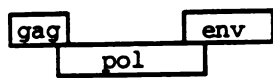
The steps of RNA packaging subsequent to RNA binding also remain to be elucidated. Does core recognize a P protein- $\epsilon$  binary complex? If so, what is the assembly pathway? Do core monomers, dimers or some other particulate intermediates mediate the recognition? What role does the proposed loop-binding host protein play in encapsidation? To this end, the same biotinylated  $\epsilon$  RNAs used in P protein binding assays could be used to affinity-purify the loop-binding protein from reticulocyte lysates for further study.

Finally, the P protein- $\epsilon$  interaction, because it is specific and because it is required for two separate steps in hepadnaviral replication (RNA packaging and DNA priming), represents an attractive target for future antiviral therapy. The development of antivirals targeted to specific RNA/protein interactions has already been described (9).

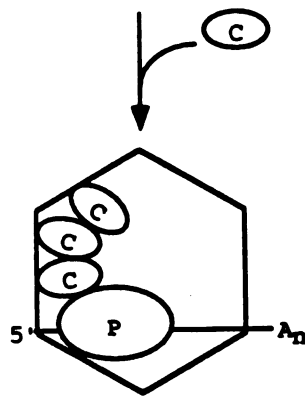
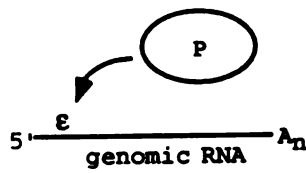
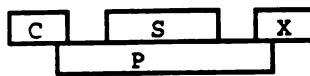


**Figure 1. Comparison of packaging strategies.** Schematic comparison of the encapsidation strategies employed by retroviruses, hepadnaviruses, and the double-stranded RNA yeast L-A virus. The genomic organizations are depicted at the top.  $\psi$ ,  $\epsilon$ , and VBS (viral binding site) are the *cis*-packaging signals for retroviruses, hepadnaviruses, and L-A virus, respectively. For further details, see text.

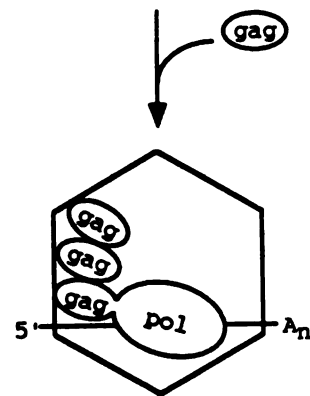
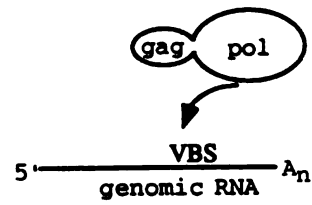
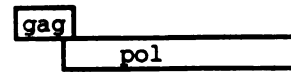
Retroviruses



Hepadnaviruses

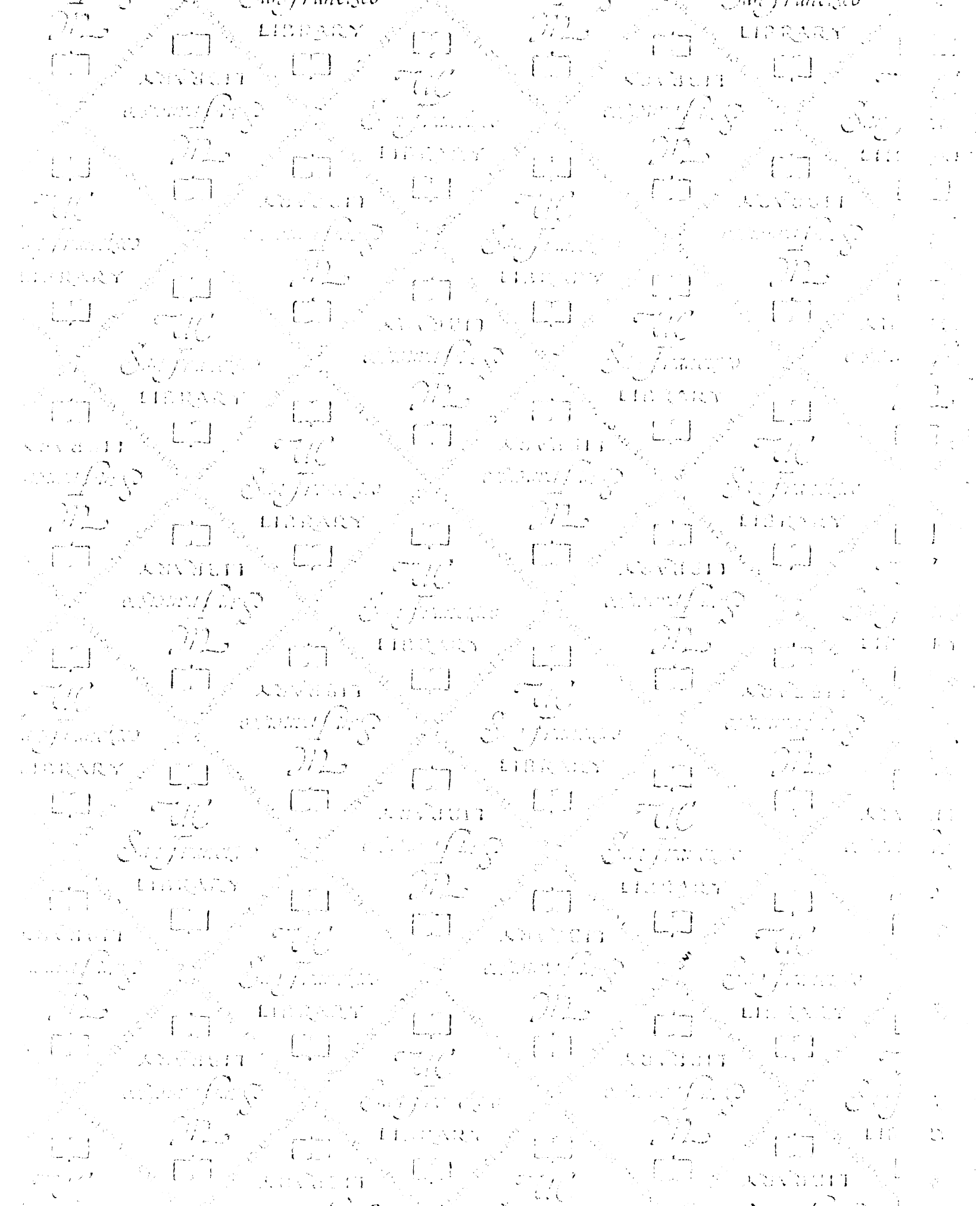


Yeast L-A virus



## REFERENCES

1. Andino, R., G.E. Reickhof, and D. Baltimore. 1990. A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell* 63: 369-380.
2. Dinman, J. D., T. Icho, and R.B. Wickner. 1991. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc. Natl. Acad. Sci. USA* 88: 174-178.
3. Fujimura, T. and R.B. Wickner. 1988. Gene overlap results in a viral protein having an RNA binding domain and a major coat protein domain. *Cell* 55: 663-671.
4. Fujimura, T., R. Esteban, L. Esteban, and R.B. Wickner. 1990. Portable encapsidation signal of the L-A double-stranded RNA virus of *S. cerevisiae*. *Cell* 62: 819-828.
5. Fujimura, T., J.C. Ribas, A.M. Makhov, and R.B. Wickner. 1992. Pol of gag-pol fusion protein required for encapsidation of viral RNA of yeast L-A virus. *Nature* 359: 746-9.
6. Marciniak, R.A., M.A. Garcia-Blanco, and P.A. Sharp. 1990. Identification and characterization of a HeLa nuclear protein that specifically binds to the trans-activation-response (TAR) element of human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA* 87: 3624-3628.
7. Mattaj, I. W. 1993. RNA recognition: a family matter? *Cell* 73: 837-840.
8. Wang, G.-W. and C. Seeger. 1993. Novel mechanism for reverse transcription in hepatitis B viruses. *J. Virol.* 67: 6507-6512.
9. Zapp, M.L., S. Stern, and M.R. Green. 1993. Small molecules that selectively block RNA binding of HIV-1 rev protein inhibit rev function and viral production. *Cell* 74: 969-978.



624368



3 1378 00624 3680

LIBRARY  
USE ONLY

