uOttawa

L'Université canadienne
Canada's university

#

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Abrahem Abrahem
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biochemistry)
GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Formation of an RNA Polymerase II Pre-Initiation Complex on an RNA Promoter Derived from Hepatitis *Delta* Virus in Vitro

TITRE DE LA THÈSE / TITLE OF THESIS

Martin Pelchat
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Earl Brown
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Ken Dimock

Martin Holcik

Odette Laneuville

Karl Andrew White
York University

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

FORMATION OF AN RNA POLYMERASE II PRE-INITIATION COMPLEX ON AN RNA PROMOTER DERIVED FROM HEPATITIS *DELTA* VIRUS IN VITRO

By

Abrahem Abrahem F.

Thesis submitted to the

Faculty of Graduate and Postdoctoral studies

In partial fulfillment of the requirements

Of Doctor of Philosophy in Biochemistry

Biochemistry, Microbiology and Immunology

Faculty of Medicine

University of Ottawa

Your file Votre référence
ISBN: 978-0-494-69125-0
Our file Notre référence
ISBN: 978-0-494-69125-0

ABSTRACT

Hepatitis *delta* virus (HDV) is the smallest known RNA pathogen able to infect humans. It is a subviral agent that depends on hepatitis B virus (HBV) infection. As HDV does not possess its own polymerase, the virus is believed to depend upon the host DNA-dependent RNA polymerases (RNAPs) for its transcription and replication. Although RNAPs are able to use RNA as a template, how they recognize RNA promoters is unknown. In this study, I used a 199-nts HDV RNA fragment, corresponding to the right terminal stem-loop extremity of the genomic polarity (R199G; nts 1540 to 60) as a model to investigate the recognition of an RNA promoter by RNAP II. Inhibition of the transcription reaction using an antibody specific to the C-terminal domain (CTD) of the largest subunit of RNAP II, and the direct binding of purified RNAP II to the RNA promoter confirmed the involvement of RNAP II in this process. RNA affinity chromatography established the formation of an RNAP II pre-initiation complex on R199G, and that the formed complex contains the core RNAP II subunit and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and TFIIS. I also found that the addition of free nucleotide triphosphates (NTPs) to the RNAP II pre-initiation complex formed on R199G caused the formation of an active initiation complex. As a result of the RNAP II complex activation, a short HDV RNA transcript was transcribed from R199G, and its 5' extremity was found to be located near the expected transcription start site of HDV mRNA (i.e. 1630). I also established the direct binding of the TATA-binding protein (TBP) to the RNA promoter, and suggested that this protein might be required for RNAP II complex nucleation on the RNA promoter. My findings represents a milestone in the study of the unconventional use of an RNA promoter

by RNAP II, and provide a better understanding of the events leading to the formation of the RNAP II pre-initiation complex on an RNA promoter.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to take this opportunity to thank the many people, who have helped me and contributed positively towards completing this work,

First, I would like to thank my supervisor, Dr. Martin Pelchat for his support, guidance, patience and encouragement. Also, I would like to thank my co-supervisor Dr. Earl Brown and my research advisory committee members, Dr. Jocelyn Cote and Dr. Jonathan Lee for their guidance and support.

I am also very grateful to my lab mates whom provided me with help and valid comments: Lynda Rocheleau, Dorota Sikora, Valerie Greco, Teodora Dimitrijevic and Youser Al-Ali.

I am very grateful to my good brother Mohamed Abou-Farha and to my family back home in Libya, specially my father Fathallah and my mother Azza for their support and encouragement and I am very grateful to my wife K. Younes for her patience and support. This appreciation is extended to my children: Fathallah, Saleh, Suhaib, Shima and Islam and to my brothers and sisters; Mabroka, Abdulmonem, Raja, Mohamed, Salem, Karema, Ali, Afaf, Ahmed Essa, Ezia, Asma and their families.

Also would like to thank Mr. Moubarak Alshamek (the head of the people's congress in Libya) for his support and assistance.

I am thankful to my good friend Alhaj Saad Buhalfia for all the support he has given me throughout my study and for all moments that he made me smile in a stressful days

Finally, I am praying to the ultimate God (in Arabic, Allah) for his guidance and making all this work possible.

LIST OF ABBREVATIONS

ADAR1	Adenosine Deaminase that acts on RNA1
ADAR2	Adenosine Deaminase that acts on RNA2
ASBVd	Avocado Sunblotch Viroid
Armi	*Armitage*
A3	HDAg-positive cell line
BRE^d	Downstream BRE
BREs	TFIIB recognition elements
BRE^u	Upstream BRE
CAK	CDK-activating complex
CDC	Central for disease control and prevention
CDK	Cyclin-dependent kinases
CCD	Coiled coil domain
CEVd	Citrus exocortis viroid
CIP	Calf intestinal phosphatase
CMV	Cytomegalovirus
CTD	Carboxy-terminal domain
DEPC	Diethylpyrocarbonate
DHFR	Murine dihydrofolate reductase
DdRP	DNA dependent RNA polymerase
DTT	Dithiothreitol
DPE	Downsream promoter elements
DSIF	DRP sensitivity inducing factor
ECL	Enhanced chemiluminescence

E. coli	*Escherichia coli*
EMSA	Electrophoretic mobility shift assay
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GST	Glutathione-S-transferase
HBV	Hepatitis B virus
HBsAg	Hepatitis B virus surface antigen
HDAg-L	Hepatitis *delta* Large antigen
HDAg-S	Hepatitis delta Small antigen
HDV	Hepatitis *delta* virus
HepG2	Human hepatoma cell line
HeLa	HeLa cervical cancer cell line
Huh7	Human hepatoma cell line7
Inr	Initiator
KDa	Kilodalton
L	Large
M	Middle
mRNA	Messenger Ribonucleic Acid
NE	Nuclear extract
NELF	Negative elongation factor
NLS	Nuclear localizing signal
NTPs	Nucleotide triphosphates
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction

PIC	Pre-initiation complex
PMG	Major pelvic ganglia cells
PNK	Polynucleotide kinase
PSTVd	Potato Spindle Tuber Viroid
P-TEFb	Positive transcription elongation factor b
RBD	RNA binding domain
RdRP	RNA dependent RNA polymerase
RIP	RNA immunoprecipitation
RNAi	RNA interference
RPB1	RNAP II large subunit 1
RPB2	RNAP II subunit 2
RNA	Ribonucleic acid
RNAPs	RNA polymerases
RNAP I	RNA polymerase I
RNAP II	RNA polymerase II
RNAP III	RNA polymerase III
rRNA	Ribosomal ribonucleic acid
RISC	RNA induced silencing complex
SDS	Sodium dodecyl sulphate
siRNA	Small interfering RNA
TAFs	TBP-associated factors
TAR	Transactivation respose RNA
TBP	TATA binding protein
TdT	Terminal deoxynucleotidyl transferase

TET	Tetracycline
TFIIA	Transcription factor two A
TFIIB	Transcription factor two B
TFIID	Transcription factor two D
TFIIE	Transcription factor two E
TFIIF	Transcription factor two F
TFIIH	Transcription factor two H
TFIIS	Transcription factor two S
tRNA	Transfer ribonucleic acid
UV	Ultraviolet
WHO	World health organization
δAg	Delta antigen
5'RACE	5' Rapid Amplification of cDNA Ends

CHAPTER ONE

INTRODUCTION

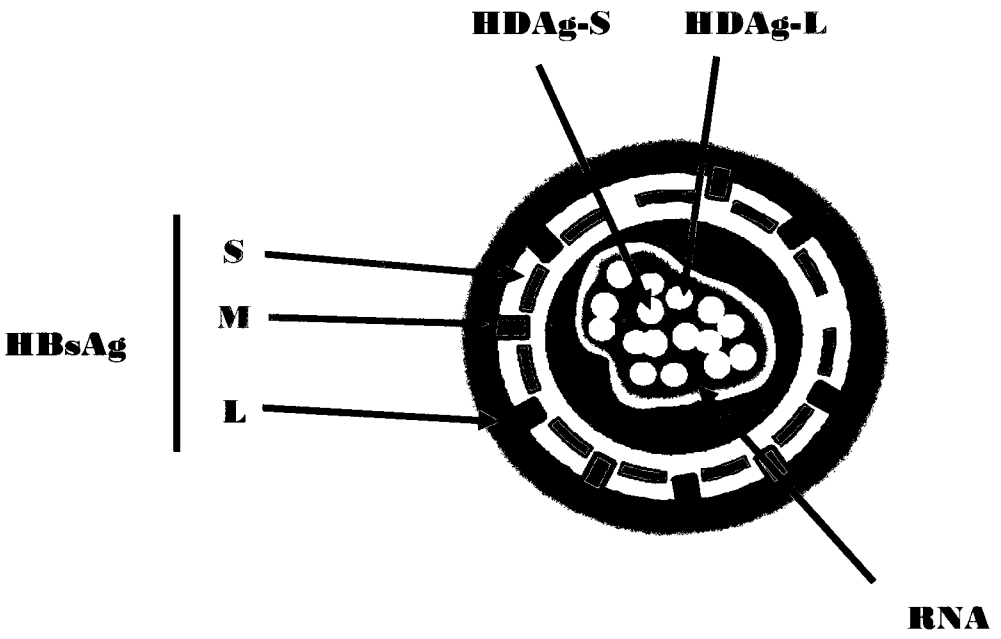
1.1 Hepatitis *delta* virus (HDV)

Viruses were believed to be the smallest infectious organisms until the late 1960s [1]. This belief was changed by the discovery of small replicable RNA pathogens with genomes ranging from 250 to 1700 nts in length. Such infectious agents are divided into four groups: viroids, satellite RNAs, viroid-like plant satellites and hepatitis *delta* virus (HDV). These RNA pathogens are capable of causing infections and diseases either plants or humans [1, 2]. The RNA pathogen to be investigated in this thesis is HDV. HDV is a defective virus that was discovered in 1977 in an Italian patient who had a chronic hepatitis B infection [2]. It is made of a spherical particle of about 36 nm in diameter that is coated with HBV envelope proteins (HBsAg). This protein envelope encapsidates the HDV RNA genome as well as the small *delta* antigen (HDAg-S), a 195 amino-acid protein of about 24 kDa, and the large *delta* antigen (HDAg-L), a 214-amino-acid protein of about 27 kDa (Figure 1); both proteins are encoded by the HDV RNA genome [3]. Due to the unavailability of an RNA-dependent RNA polymerase (RdRP) in the human cell, DNA-dependent RNA polymerase(s) (RdRP) is believed to carry out HDV RNA transcription [4, 5, 6].

1.1.1 HDV genetic diversity and epidemiology

HDV is genetically diverse and it is classified into eight different clades based on sequence analysis. Clade I is the most widely distributed among the eight clades, and is

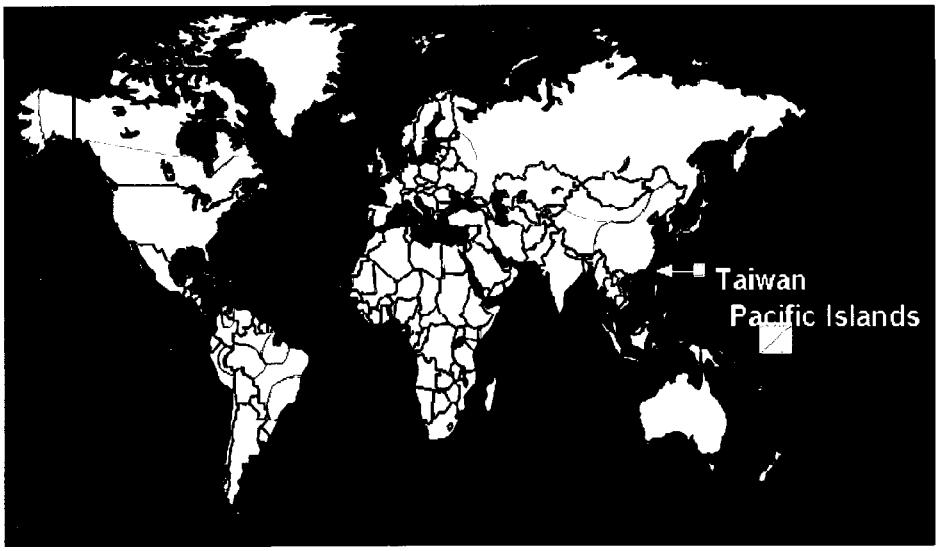
Figure 1: Composition of the HDV particle. The HDV particle contains a circular strand of genomic RNA and both HDAg-S and HDAg-L. It is coated with HBsAg envelope proteins which is composed of small (S), middle (M), and large (L) (adapted with permission from http://www.hivandhepatitis.com/).


HDAg-S
HDAg-L
S
M
L
HBsAg
RNA

present throughout Europe, North America, Africa, and some regions of Asia [8, 9]. Clade II is mostly present in Taiwan, Japan, and parts of Russia [10, 11, 12, 13]; Clade III is the most-prevalent genotype in South America (Venezuela, Peru, and Colombia) [10, 11]. The presence of five other genetically variable HDV clades was discovered in HDV samples from West and Central Africa [14]. In this study, HDV RNA extracted from 25 patients who had been previously diagnosed with HDV infection was reverse transcribed and the resulting cDNA was sequenced [14]. Those patients had either been born in Africa or had traveled to Africa. Phylogenetic analyses indicated that approximately 70% of the characterized isolates formed highly divergent groups. These highly divergent sequences resulted from extensive mutational changes that occur in the course of HDV RNA replication due to the infidelity of the RdRP activity in the host [15].

The identification of these different clades pointed to the epidemiology of this virus in different parts of the world. It is also endemic in developing countries, as it is transmitted through transfusions of unscreened or poorly screened blood and blood products [16]. It spreads at lower frequencies among drug users and from an infected mother to her infant during birth [16]. This was demonstrated by examining HBsAg positive individuals from various parts of the world for the presence of a specific antibody for HDV [17]. The antibody was detected in HBsAg-positive Italian descendents and polytransfused HBsAg patients from all around the world [17]. These data indicated that the infection was associated with HBV, and spread among patients through parental and parenteral routes [17]. Due to the association of both HDV and HBV infections, and the similarities of HDV infection risk factors to those of HBV, HDV spreads in areas where HBV is common (Figure 2).

Figure 2: Geographic distribution of HDV. From: Centers for Disease Control and Prevention (CDC), Atlanta, GA USA (http:// www.cdc.gov/ncidod/disease/ hepatitis/ slideset/hep_d/slid_6.htm) (permission is obtained from the CDC).



HDV Prevalence

- High
- Intermediate
- Low
- Very Low
- No Data

1.2 HDV biology

1.2.1 Requirement of HBsAg for HDV packaging and infectivity

Association between HDV and HBV infections is explained by the use of HBsAg for coating of both viruses. HDV takes advantage of the fact that HBV is quite inefficient in particle assembly. In fact the serum of an infected patient may contain an excess of 1000- to 100,000 fold of empty HBV subviral particles as compared to complete infectious particles [18]. These HBV subviral particles are used by HDV to coat itself [18]. HBsAg consists of three HBV envelope proteins, which are classified according to their molecular weight into small (S, 24 KDa), middle (M, 33 KDa) and large (L, 42.3 KDa) proteins. The M protein is made of the S protein (226 amino acid residues) and the pre-S2 polypeptides that consists of 55 amino acid residues. The L protein is a multi-subunit protein that is made of the M protein and pre-S1 polypeptides, which is made of 108 to 119 amino acid residues [19, 20].

The role of these HBV envelope proteins in the assembly and infectivity of HDV particles was investigated using human hepatoma cell lines (Huh7) stably expressing either the S, SM or SML proteins [21]. These cells were transfected with plasmids containing a head to tail trimer of HDV cDNA under the control of simian virus 40 late promoter. Each of these three cell lines produced HDV particles coated with one of the three HBsAg envelope proteins. These HDV particles were then harvested and used to infect chimpanzee primary hepatocytes. Chimpanzee primary hepatocytes were used instead of the Huh7 cell line because of the resistance of the Huh7 cell line to HDV infection [21]. It was found that only particles generated from cells expressing all three proteins (SML) were able to infect chimpanzee primary hepatocytes [21]. The detection of both HDAg-S and HDAg-L proteins by western blot analysis and HDV RNAs by northern blot analysis in the generated HDV

particles pointed to the transcription\translation of HDAg subgenomic HDV mRNA and replication of HDV RNA in the infected cells. These data clearly demonstrate the dependency of HDV infectivity of chimpanzee primary hepatocytes on the presence of all three proteins from the HBV virus [21].

Since HDV is dependent on HBsAg envelope proteins for coating and infectivity, HDV infection occurs only as a co-infection or a super-infection to HBV infections [17]. Co-infection with both viruses usually results in a severe form of acute hepatitis disease which leads to an increased risk of liver failure. Super-infection occurs when chronic HBV patients are infected with HDV, this might lead to liver cirrhosis or hepatocellular carcinoma.

Because HDV infection can only occur as a co-infection or a super-infection to HBV, the protection from HDV infection can be achieved by vaccination against HBV [22]. Due to the efficacy of vaccinations against HBV, the numbers of HDV infections in Western Europe were reduced between years 1970 and 1980 [22]. Starting from year 1990 no decline was reported in the numbers of infected individuals. These numbers were stabilized as a result of influx of infected migrants from HDV-endemic areas [23]. The numbers of globally infected HDV individuals are estimated by the World Health Organization (WHO) to be around 10 million people [24]

1.2.2 HDV RNAs and their structures

Even though HDV infection is associated with HBV, HBsAg is the only protein required for HDV particle assembly and infectivity. During the course of an HDV infection, three major viral RNAs are detected in infected liver cells. The first one is a single-stranded

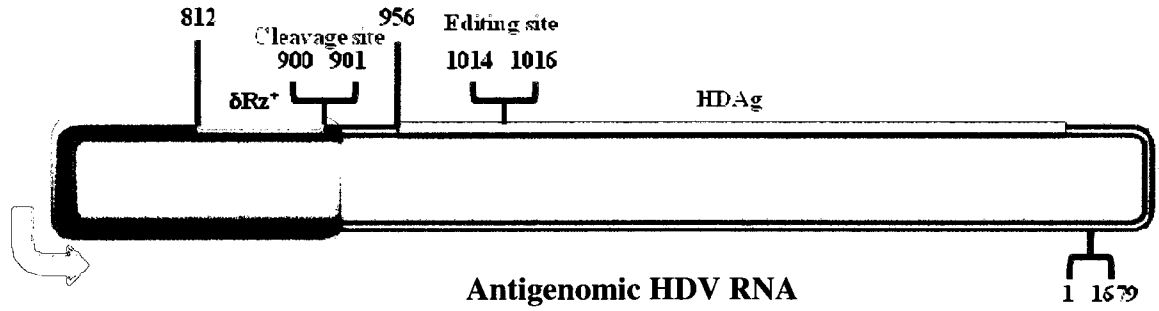
circular genomic RNA of approximately 1700 nts that folds upon itself to form an unbranched, rod-like structure using 74% intra-molecular base-pairing [25, 26]. The second RNA is the complementary sequence (i.e. antigenomic RNA) that also forms an unbranched rod-like structure [26]. The third species is an extra subgenomic 800 nts long mRNA that results from transcription of genomic RNA, and encodes for HDAg (Figure 3) [28]. In addition, both genomic and antigenomic HDV RNAs contain an 85 nts self cleaving motif (i.e. *delta* ribozyme motif) (Figure 4). During replication, the *delta* ribozyme motifs self-cleave the multimeric HDV RNA molecules to generate HDV RNA monomers [27].

1.2.3 HDV Replication

HDV RNA replication is a very complex process; it involves the replication of two different HDV RNAs and the transcription of two subgenomic RNAs, which encode for the two HDAg proteins. Similarities between those two proteins make it difficult to differentiate between their localization in the infected cells. In order to understand the HDV RNA replication mechanism and to better interpret the results obtained from any replication system, several precautions have to be taken into consideration: first, a good infection system that supports HDAg localization and HDV RNA replication has to be used; second, when using an expression system, it has to somehow mimic the real infection in terms of controlling the number of HDV RNA copies introduced into the cells, as this will avoid overwhelming the cell system with the supplied HDV RNA; third, a good detection method

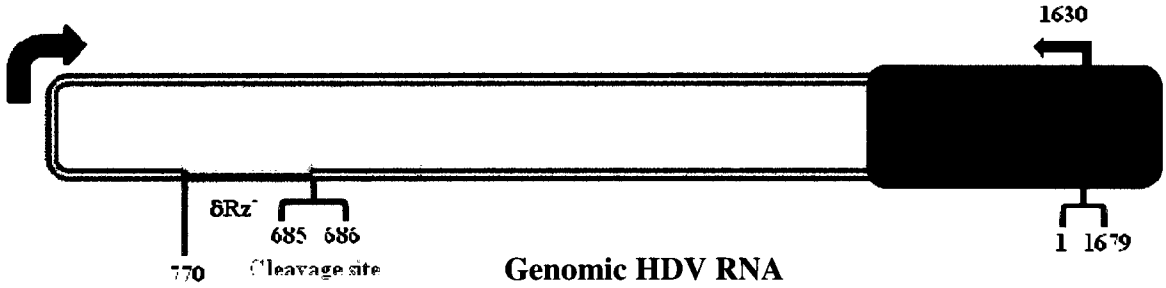
Figure 3: Schematic representation of the three HDV RNAs accumulating during an infection. The antigenomic **(A)** and genomic **(B)** HDV RNAs are represented in counterclockwise and clockwise orientation respectively. The *delta* ribozyme motifs are presented in light green colored boxes with nucleotides 685/686 and 900/901 indicating the cleavage sites for genomic and antigenomic RNAs respectively. The brown box shows a 213-nt RNA template that corresponds to the antigenomic left terminal tip region (L213AG; nts 687 to 900) used in this study. The blue box shows an RNA fragment containing 199 nts corresponding to nucleotides 1541 to 60 of the right genomic region (R199G) used in this study. **(C)** The light blue colored box shows subgenomic mRNA with a 5' cap and a 3' polyadenylated end.

A)



Antigenomic HDV RNA

B)



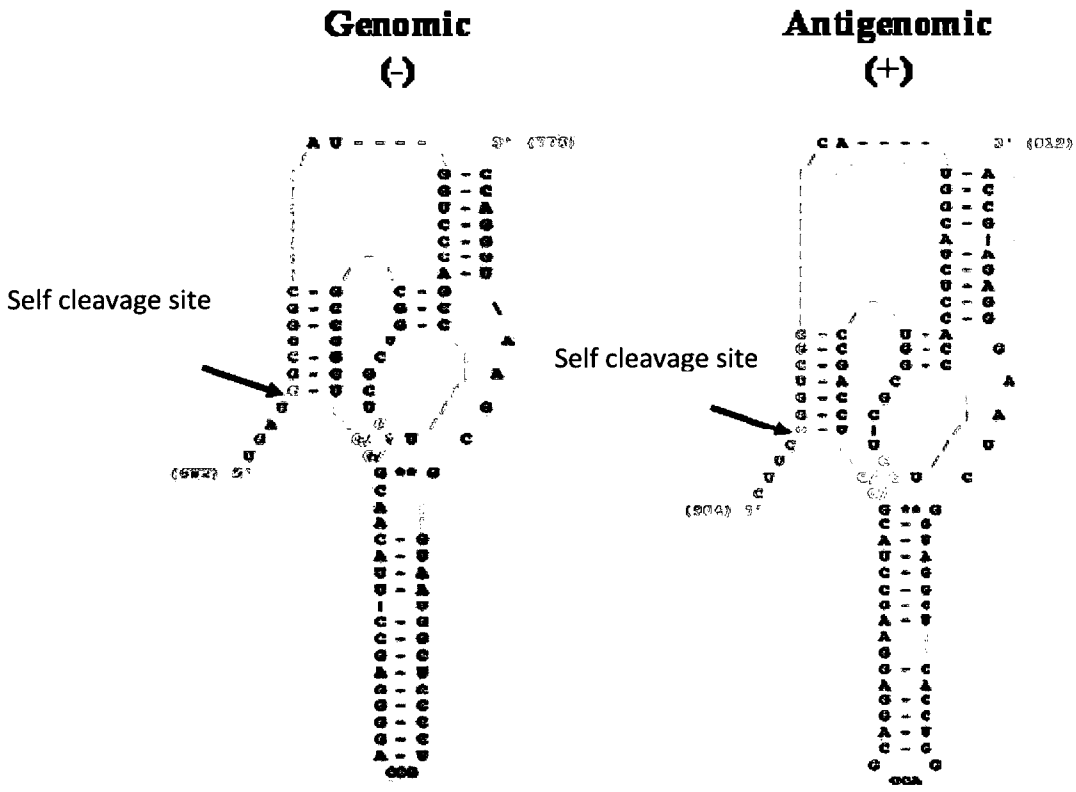
Genomic HDV RNA

C)

3' Poly (A) RNA *Delta* antigen open reading frame 5' Cap

[DV subgenomic mRNA

Figure 4: Secondary structures of both genomic and antigenomic *delta* ribozyme motifs. The self cleavage sites are indicated by black arrows (adapted with permission from http://subviral.med.uottawa.ca/cgi-bin/rz2d2.cgi?fkSequence=1137&display=1). The red colored base paring represent the p1.1 loop. The pink colored nucleotide represents the cleavage site. The blue colored nucleotides represent the substrate strand.


Genomic
(-)
Antigenomic
(+)
Self cleavage site
Self cleavage site

that differentiates between HDV RNA accumulation and replication should be used; and fourth, good antibodies that differentiate between the two HDAg proteins should be used for protein localization experiments. Unfortunately, a system that takes all these points into consideration has not been developed yet. Until a good HDV infection system and replication detection methods are developed, I base our discussions on the data available in the field.

HDV requires HBsAg to coat itself, and to infect liver hepatocyte cells. Once inside the cells, HDV RNA is released from its envelope and enters the nucleus to start its replication cycle (Figure 5A). In fact, both genomic and antigenomic HDV RNA accumulate in the nucleus of transfected cells during HDV RNA replication [28]. HDV RNA replicates by a double rolling circle mechanism. The replication begins using a circular genomic HDV RNA as a template to generate a multimeric intermediate of antigenomic polarity. The multimeric products are then self-cleaved by intrinsic *delta* ribozyme motifs present in the antigenomic polarity to produce unit-length linear antigenomic RNAs [4]. These antigenomic RNA monomers are ligated to form circular antigenomic RNAs. Subsequently, these RNAs are then used as templates to repeat the transcription, cleavage and ligation processes, resulting in the generation of circular genomic RNA monomers (Figure 5B). This replication cycle repeats itself to generate more HDV RNAs.

A cellular ligase is hypothesized to be involved in the ligation process based on a study using a panel of HDV RNA substrates bearing 5' hydroxyl and 2', 3'- cyclic phosphate termini. 2', 3'-cyclic phosphate is the product of *delta* ribozyme cleavage during HDV replication [29]. The ligation of these substrates was found to occur in host cells, but not in

Escherichia coli, pointing to the involvement of the host in the ligation reaction [30]. A role for HDAg proteins in the enhancement of both HDV RNA cleavage and ligation processes was presented. This finding pointed to HDAg's capability to act as a chaperone [31].

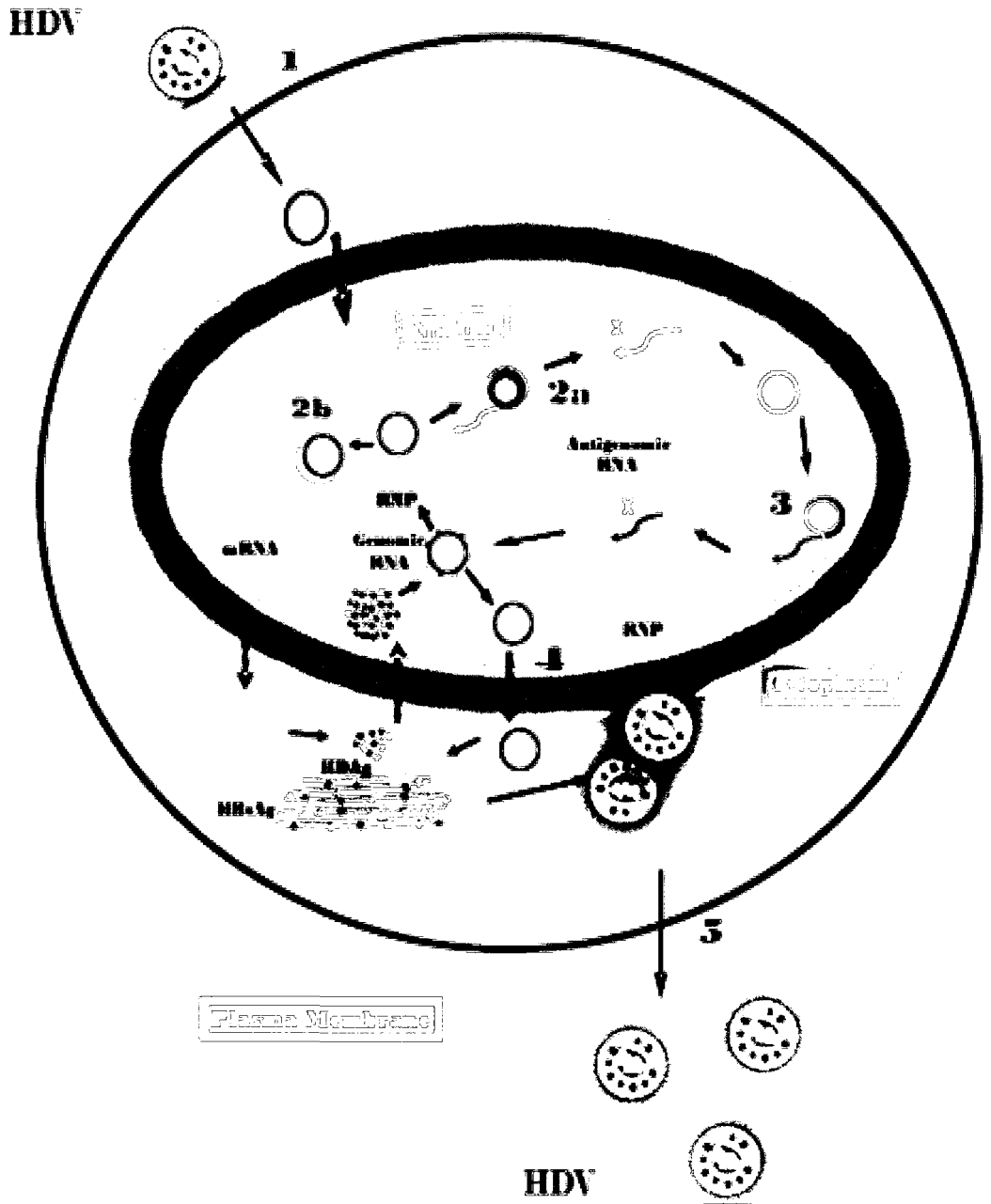
As HDV does not encode its own polymerase and can replicate in the absence of HBV, it has to use host cell machinery to replicate. RNAP II has been implicated in HDV RNA replication [56]. Recently, more complexity was added to the field by the association of both RNAP I and RNAP III with HDV RNA [32]. Even though the association of those two polymerases with HDV RNAs has been confirmed, the inhibition of mRNA transcription from genomic RNA and the accumulation of genomic RNA by low concentrations of α-amanitin pointed to the involvements of RNAP II in those two processes (See section 1.4) [33, 34].

1.2.4 Transcription of HDV mRNAs

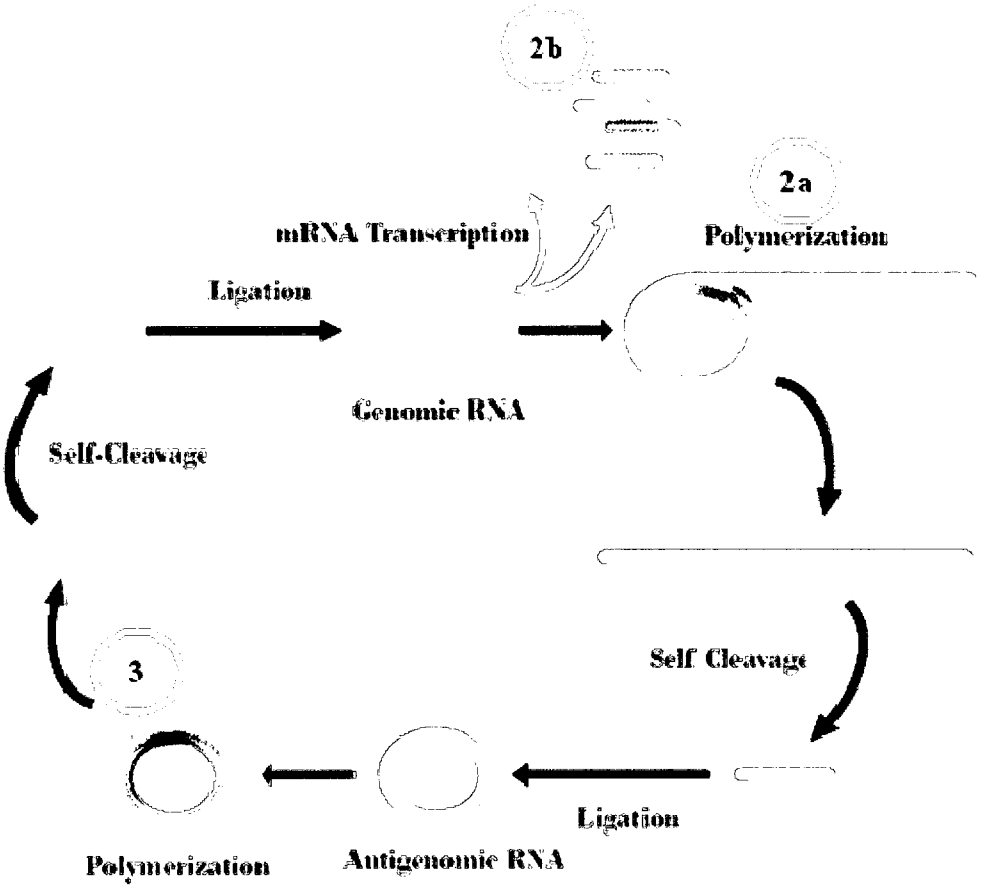
During the replication process, an HDV mRNA of about 800 nts that encodes HDAg is transcribed from the genomic RNA [33]. This small subgenomic RNA has the hallmarks of mRNAs transcribed by RNAP II. It is post-transcriptionally modified by 5' capping and 3' polyadenylation [22, 35]. Its polyadenylation signal (AAUAAA) is located 15-20 nts upstream from the polyadenylation site [28]. This mRNA is translated into the HDAg-S which is required for HDV RNA accumulation. During replication, some antigenomic RNA is edited by an adenosine deaminase acting on RNA (ADAR1) at position 1015. In this editing process, the UAG stop codon of the HDAg-S sequence is edited to UIG. As inosine (I) in the UIG pairs with cytosine (C) during replication, the replicated genomic RNA contains ACC, which subsequently is transcribed to UGG. This editing results in the

Figure 5: Stages of HDV infection and its rolling-circle mechanism in the infected cell. **A) (1)** HDV attaches to the cell membrane and enters into liver cell where the genomic RNA is translocated to the nucleus, **(2a)** genomic RNA serves as a template for the synthesis of antigenomic multimers, which then self-cleaves and are ligated to form circular antigenomic RNA monomers; **(2b)** the genomic RNA is used for mRNA synthesis; **(3)** the antigenomic RNA monomers are used to synthesize genomic RNA multimers, which self-cleave and are ligated to generate circular genomic RNA monomers; **(4)** HDV ribonucleoproteins (RNPs) are exported from the nucleus to the cytoplasm, where particle assembly takes place; and, **(5)** the produced HDV particles are released to infect other cells. **B)** Magnification of the HDV replication cycle showing steps **(2a, 2b and 3).**

A)


HDV
1
2b
2a
Antigenomic RNA
RNP
3
mRNA
Genomic RNA
RNP
4
HDAg
HBsAg
5
Plasma Membrane
HDV

B)


2b
2a
mRNA Transcription
Polymerization
Ligation
Genomic RNA
Self-Cleavage
Self Cleavage
3
Ligation
Polymerization
Antigenomic RNA

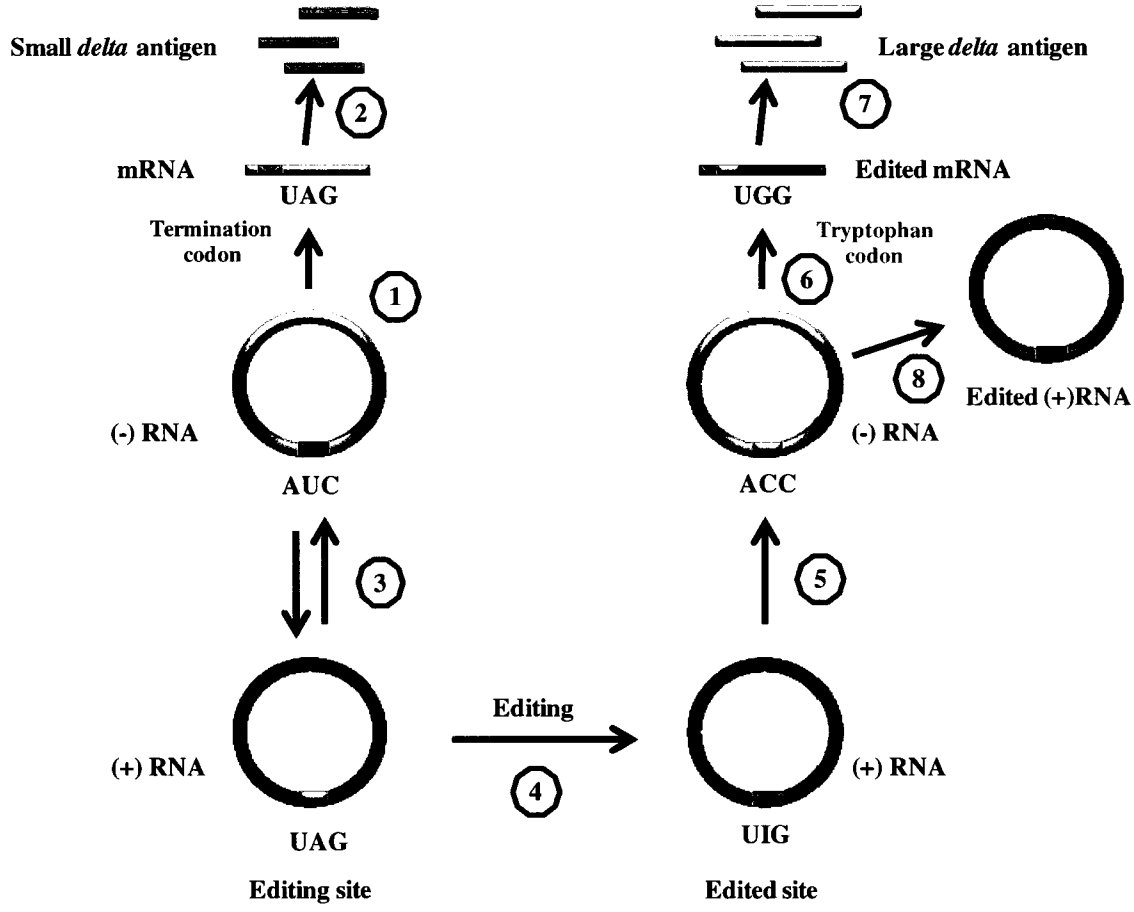
changes of the stop codon (UAG) into a tryptophan codon (UGG) leading to the addition of an extra 19 amino acids to the HDAg-S to finally make the HDAg-L (Figure 6) [36].

Recently, an antigenomic HDV RNA that folds into alternative branched secondary structure around the amber/W site in clade III HDV from Peruvian and Ecuadorian isolates was described. This branched structure makes the antigenomic HDV RNA a better substrate for the ADAR1 editing than unbranched antigenomic HDV RNA, which subsequently leads to transcription of subgenomic mRNA that encodes for HDAg-L [37]. In several others studies, HDV RNA editing levels have been shown to negatively affect the level of HDV RNA replication, as some of the antigenomic RNAs are committed to the editing process [36, 38, 39, 40]. HDAg-S and HDAg-L proteins play essential roles in HDV RNA transcription and particle assembly respectively [41].

1.3 Functional domains of HDAg-S and HDAg-L

Various functional domains have been identified in both HDAg-S and HDAg-L. HDAg is required for transport of HDV RNA to the nucleus. This was proven when transfected genomic HDV RNAs were found to predominantly localize to the cytoplasm [34]. Co- transfection of Huh7 cells with both genomic HDV RNA and HDAg resulted in the nuclear localization of HDV RNA. In order for this protein to transport the HDV genomic RNA to the nucleus, it requires a nuclear localization signal (NLS). A NLS can be identified as a short stretch of amino acids in a nuclear protein that has the capability to transport proteins, DNA or RNA into the nucleus. Deletion of this motif usually results in

Figure 6: HDV RNA editing by double-stranded RNA adenosine deaminase. **(1)** Small *delta* antigen mRNA synthesis. **(2)** mRNA translation to HDAg-S. **(3)** Transcription of full-length genomic RNA to full-length antigenomic RNA. **(4)** Editing of a specific A residue of full-length antigenomic RNA at position 1015 to inosine (I) by a double-stranded RNA adenosine deaminase. **(5)** During RNA replication, I in the UIG pairs with a C residue forming ACC in the full antigenomic RNA. **(6)** Transcription of the edited genomic RNA to antigenomic mRNA will contain a UGG codon (tryptophan) instead of the stop codon. **(7)** Translation of the newly transcribed mRNA to HDAg-L that contains an extra 19 amino acid. **(8)** Replication of antigenomic edited RNA from genomic RNA.


Small *delta* antigen
2
mRNA
UAG
Termination codon
1
(-) RNA
AUC
3
(+) RNA
UAG
Editing site
Editing
4
Large *delta* antigen
7
Edited mRNA
UGG
Tryptophan codon
6
8
Edited (+)RNA
(-) RNA
ACC
5
(+) RNA
UIG
Edited site

the disruption of nuclear import [42, 43]. In HDAg, the NLS is located from the 66th to the 75th amino acid in both proteins (Figure 7) [44]. The deletion of the NLS region led to a significant accumulation of antigens in the cytoplasm [34]. However, because some antigens managed to get translocated to the nucleus in the absence of the NLS domain no functional defect was observed in HDAg-L and HDAg-S antigens. This finding suggests that these antigens are transported by other proteins with NLS or through another NLS domain that might exist within HDAg [45].

In addition to its involvement in nuclear localization, HDAgs have RNA binding capability through RNA binding domains (RBD) [46]. RBD are present in a large variety of proteins, which are required for post-transcription modification of mRNA. Beside the post-transcriptional modification, RBD are required for several other processes that include nuclear export, transport, and localization and mRNA transcription. HDAg RBD has been proven to be required for HDAg-S but not HDAg-L binding to HDV RNA [45]. It consists of two stretches of arginine-rich motifs (ARMs) spanning from 97 to 146 amino acids, KERQDHRRRKA (97th to the 107th) and EDEKRERRIAG (136th to the 146th) [46]. The deletion of a stretch or site-specific mutations of RBD resulted in a total loss of *in vitro* RNA-binding ability. These two stretches are separated by 29 amino acids (Figure 7). Replacing the spacing sequence with a shorter one reduced HDAg's RNA binding capability *in vitro* [46].

HDAg also has a coiled coil domain (CCD). This domain is required for protein-protein dimerization that is required for both HDAg-S activation of transcription, and HDAg-L inhibition of HDV RNA replication. HDAg-L binds to HDAg-S and inhibits its role in HDV RNA replication [45]. The CCD stretches from the 12th to the 60th amino acid

[45] (Figure 7). Several amino acid mutations, which caused structural changes in the CCD of both HDAg-L and HDAg-S antigens, led to an inhibition in the ability of HDAg-S to activate transcription, and abolished the ability of HDAg-L to inhibit HDV replication [47, 48].

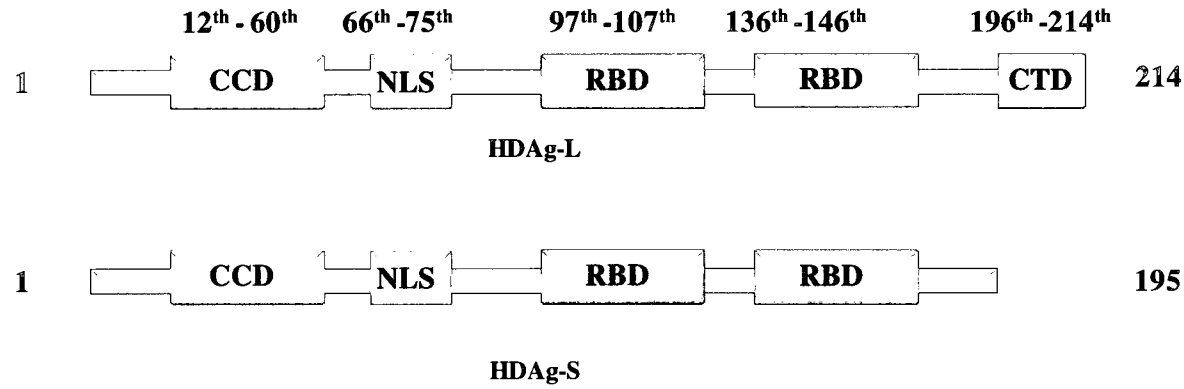
Finally, HDAg-L has a CTD, which is composed of an extra 19 amino acids not found in HDAg-S. HDAg-L is required for HDV particle assembly through binding to L antigen in the HBsAg. It has also been shown that deletions of any 5 amino acids from 196 to 210 in the CTD of HDAg-L abolished its ability to establish HDV particle assembly. Deletion of 33 amino acids from 163 to 195 upstream of the HDAg-L CTD did not have any effect on HDV particle assembly. This indicated the importance of the HDAg-L CTD for HDV packaging [49].

1.3.1 The role of HDAg in HDV biology

1.3.1.1 The role of HDAg-S in transcription

It is well documented that HDV utilizes the human transcription machinery for its transcription [4, 5, 6]. Nonetheless, in most cases accumulation of HDV RNA cannot be detected without the presence of HDAg-S. The involvement of HDAg-S in HDV accumulation was suggested when the monkey kidney COS7 cells, and Huh7 cells were transfected with a dimeric HDV cDNA construct that contained a 2- base pair deletion at position 1434 and 1435. This deletion introduced a frame shift at amino acid 53 in the HDAg-S open reading frame, which consequently led to a 40-fold reduction in HDV RNA accumulation [5]. HDV accumulation was restored to normal levels by co-transfection with a plasmid expressing wild-type HDAg-S [5]. The involvement of HDAg-S was also studied

Figure 7: Schematic diagram of functional domains found in HDAg-L and HDAg-S. The domains are shown in different colors. Coiled-coil domain (CCD 12^{th} to the 60^{th} amino acids) is represented by the green box, Nuclear localizing signal (NLS 66^{th} to the 75^{th} amino acids) is presented by orange box, RNA-binding domains (RBD; 97^{th} to the 107^{th} and136^{th} to the 146^{th}) is represented by light purple boxes, the CTDs (196^{th} to the 214^{th}) which is present exclusively in the HDAg-L), are represented by the red box.


12th - 60th
66th -75th
97th -107th
136th -146th
196th -214th
1
CCD
NLS
RBD
RBD
CTD
214
HDAg-L
1
CCD
NLS
RBD
RBD
195
HDAg-S

by stably transfecting HDV RNA into HDAg negative and positive human hepatoma cell lines (HepG2). HDV accumulation was only observed in cells expressing HDAg-S [50]. This experiment was also repeated *in vitro* using nuclear homogenates from both cell lines. Under these conditions, transcription from HDV RNA was observed from both nuclear extracts [50]. In this case and in other studies, it has been confirmed that the presence of HDAg-S is not required for the initiation of HDV RNA synthesis [34, 51], but for binding to and displacing the negative elongation factor (NELF) and the DRB-sensitivity inducing factor (DSIF) from RNAP II containing complexes [52].

NELF is a negative elongation factor that has a limited sequence similarity to HDAg. NELF interacts with RNAP II and represses its transcription function [52]. Studies have shown that transcription from DNA templates are paused in the presence of DSIF/NELF. The addition of the HDAg-S relieves the repression effect caused by the two proteins. Deletion of 8 amino acids from the CTD domain of the HDAg-S resulted in loss of the HDAg-S relieving effect; this pointed to the requirement of the HDAg-S CTD domain in the displacement process [52]. The presence of other unknown positive elongation factors that would obscure the DSIF/NELF effect might eliminate the need for HDAg-S to stimulate the transcription.

1.3.1.2 The role of HDAg-L in viral packaging and transcription inhibition

HDAg-L is believed to play an important role in HDV virion assembly. This role was investigated using a frame-shifted HDV mutant, which had a thymine residue inserted immediately after the stop codon of the HDAg-S cDNA. The insertion of a thymine resulted in the presence of two consecutive stop codons at the end of the mRNA encoding HDAg-S,

and also a frame shift immediately following those stop codons. The frame-shifted HDV mutant was capable of accumulating and expressing the HDAg-S but not the HDAg-L. No HDV particles were released from the cells where HDAg-L was lacking [41]. On the other hand, the wild type HDV clone was able to produce HDV particles, and the complementation of the former frame shifted HDV mutant by co-transfection of HDAg-L cDNA led to the recovery of HDV particles. The particles were collected and recovered from all the culture media. The HDV particles were then lysed and the HDV RNAs and proteins were extracted and subjected to Northern and Western blot analyses to detect HDV RNA, both HDAg proteins, and HBsAg in HDV [41]. HDV RNA and both HDAgs were detected in all HBsAg positive particles [41].

1.4 The involvement of host RNA polymerase(s) in HDV replication and transcription

HDV accumulates in the nucleus of mammalian host cells where it starts to replicate by a rolling circle mechanism [4, 53] (as described in section 1.2.3). In an effort to identify the RNAP(s) involved in HDV replication, α-amanitin was used to inhibit the transcription from HDV RNA templates [54]. α-amanitin is a mycotoxin known to inhibit DNA-dependent RNA synthesis in eukaryotes and to differentiate between the RNAPs responsible for the transcription from DNA templates. α-amanitin inhibits RNAP II at a concentration above 0.02 μg/ml, whereas it inhibits RNAP I and RNAP III at 200 μg/ml and 20 μg/ml, respectively [55]. Using low concentrations of α-amanitin and a specific antibody against the CTD of RNAP II, the RNA transcription from both genomic and antigenomic polarities of HDV RNA in homogenized nuclei from HepG2 cells was inhibited. These data indicated that RNAP II might be responsible for all aspects of HDV RNA synthesis [56].

In a recent study, a tetracycline (TET)-inducible HDAg-S stable cell line (293) was used to understand the HDV replication [54]. The nuclei of the TET-induced cell line were extracted and used in an *in vitro* nuclear run-on assay. When HDV genomic RNA was added to the reaction mixtures, both genomic and antigenomic HDV RNAs were detected. To investigate the role of RNAP II in HDV replication, an α-amanitin inhibition experiment was performed as follows. The nuclei of the induced cells were extracted prior to the nuclear run-on experiment to avoid α-amanitin interference with HDAg-S transcription [57]. One μg/ml α-amanitin was added to the reaction, which resulted in ~86% reduction of both genomic and antigenomic HDV RNA production as compared to the control without the α-amanitin. Transcription from glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and actin genes, which are also transcribed by RNAP II, was inhibited to the same level as HDV RNA. Under these conditions α-amanitin had no inhibitory effect on transcription from the 18S or U6 genes by RNAP I and RNAP III, respectively. Based on these results, it was concluded that RNAP II might be the only RNAP involved in HDV replication [56, 57].

However, another set of studies, using α-amanitin to determine the RNA polymerase responsible for HDV RNA transcription from RNA templates gave conflicting results. A concentration as high as 100 μg/ml was not able to inhibit the synthesis of antigenomic RNA from genomic RNA in HepG2 cell extracts [54, 4]. These findings suggest that RNAP I might be responsible for the transcriptions of antigenomic HDV RNA. A third study suggested the involvement of two different cellular machineries in the replication of genomic and antigenomic RNAs. They showed that accumulation of the genomic and antigenomic RNAs takes place in two different intranuclear compartments [6]. The genomic

RNA accumulates in the nucleoplasm and is sensitive to 10 μg/ml of α-amanitin indicating the involvement of RNAP II [6], while the antigenomic HDV RNA accumulates in the nucleolus and is resistant to higher concentrations of α-amanitin. This finding suggests the involvement of RNAP I or a similar enzyme in the synthesis of antigenomic HDV RNA [4, 6, 54].

This is also supported by the fact that HDV RNA synthesis was affected by the depletion of an RNAP I-specific transcription factor named SL1 [6]. However, the specificity of the antibody used was not determined, and the immunodepleted protein SL1 was not added back to the reaction. It was recently reported that HDV RNA can associate with the RNAP I-specific transcription factor SL1 through its component TAFI, and with RNAP III through one of its core subunits POLR3K. In this study HDV genomic and antigenomic RNA was co-immunoprecipitated with RNAP I and RNAP III using antibodies specific to those proteins [32]. However, no functional transcription assays were performed to confirm their involvement in HDV replications. Despite all of these aforementioned results, it is confirmed that the transcription of genomic RNA and subgenomic mRNA were slightly different in α-amanitin sensitivity [4, 6, 54, 56, 57], which might suggest that RNAP II is responsible for their transcription.

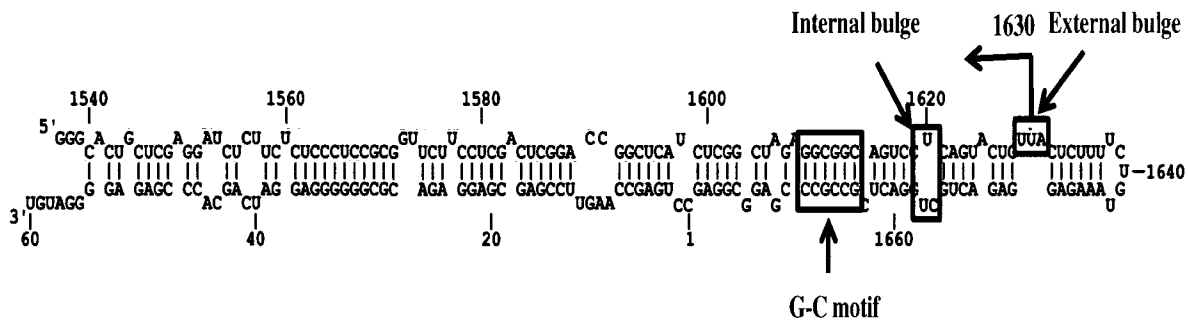
1.5 Transcription initiation on HDV RNA

Upon HDV infection, the majority of the RNA species detected are unit-length circular RNAs. The circular nature of HDV RNA makes it difficult to identify the promoter region(s) responsible for replication. In order to identify promoter regions, short HDV RNA fragments were tested in *in vitro* transcription assays using nuclear extract proteins [34, 51].

For instance, a genomic HDV RNA template (R199G), corresponding to the right extremity of the genomic HDV RNA and containing the initiation site for HDAg mRNA transcription, was used to investigate its capability to act as an RNA promoter [51] (Figure 8). *In vitro* transcription on R199G resulted in the production of a small RNA transcript, which initiates near the mRNA initiation site (position 1630) [95]. The transcribed RNA product was found to result from *de novo* synthesis. This was confirmed by RNase I digestion using a genomic HDV RNA probe that bound to the RNA product [51]. Extensive mutational studies were performed to investigate the effect of nucleotide composition and RNA structure on HDV RNA synthesis. Single and multiple point mutations were introduced to R199G. Mutations to a GC-rich region and to the internal 1654(UC) 1655, U1620 and external 1629(UUA) 1631 bulges (Figure 8) that are located in the stem loop of R199G led to a complete loss of RNA synthesis [51]. Inhibition of RNA synthesis from these mutations is believed to have occurred as a result of the change in the R199G secondary structure. On the other hand, other mutations that had no effects on the internal and external bulges structure had minor effects on HDV RNA synthesis [33]. These findings led to the speculation that the existence and the structure of the external and internal bulges in the stem loop region of R199G are necessary for HDV replication as they might include the RNAP recognition site [33, 51]. It is worth mentioning that in both studies the effect on the secondary structure was not confirmed by RNase protection assay. The above mentioned secondary structure changes were predicted using Mfold, which calculates delta G [58, 51].

In another study, various HDV RNA fragments were screened for promoter activity in an *in vitro* transcription study using HeLa nuclear extract. A region corresponding to the

Figure 8: Secondary structure of an RNA promoter located at the right terminal stem loop of HDV genomic RNA. Secondary structure of R199G used in the study showing the important motifs identified. The arrow indicates the transcription initiation site of HDAg mRNA. The blue color represent the HDAg mRNA coding sequence.


Internal bulge
1630
External bulge
1540
1560
1580
1600
1620
5'
3'
60
40
20
1
1660
U—1640
G-C motif

left terminal stem loop of the antigenomic HDV RNA (L213AG) was identified as an RNA promoter for RNAP II [34]. Transcription from the L213AG resulted in the cleavage of the template RNA followed by extension of 41 nts to the created 3' end resulting in a chimeric template/transcript product (Figure 9A) [34]. The transcription was highly sensitive to 1μg/ml of α-amanitin, which strongly suggested that RNAP II is the polymerase responsible for the transcription [34]. In support of the involvement of RNAP II, the transcription from the left terminal extremity was partially sensitive to α-amanitin when a major pelvic ganglia cell (PMG) nuclear extract that contains an α-amanitin-resistant RNAP II allele was used. The role of HDAg-S in the transcription reaction from the RNA left terminal extremity of HDV was suggested based on the pausing of the transcription by DRB-sensitivity inducing factor (DISF). This finding pointed to the role of the two repressor proteins (DSIF/NELF) in pausing transcription [52]. This pausing effect was reversed by the addition of HDAg-S to the reaction which causes the elongation of transcription to proceed [52].

Mutagenesis studies including mutations and deletions were performed on the left-terminal region, specifically within the stem loop region. These mutations were found to have slight effects on transcription [34]. The only mutations that caused loss of RNA accumulation were insertions of non-HDV sequences into and the deletion of large nucleotide segments from the conserved CUC/GAG motif (Figure 9B); CUC/GAG is a conserved motif found in a double-stranded region closest to the loop of the left antigenomic extremity [34].

Recently, pure RNAP II, TFIIS, a short HDV RNA corresponding to the left terminal antigenomic HDV RNA and NTPs were used in a crystallographic study. The role of TFIIS in the study is to induce RNA cleavage activity by enhancing the intrinsic nuclease

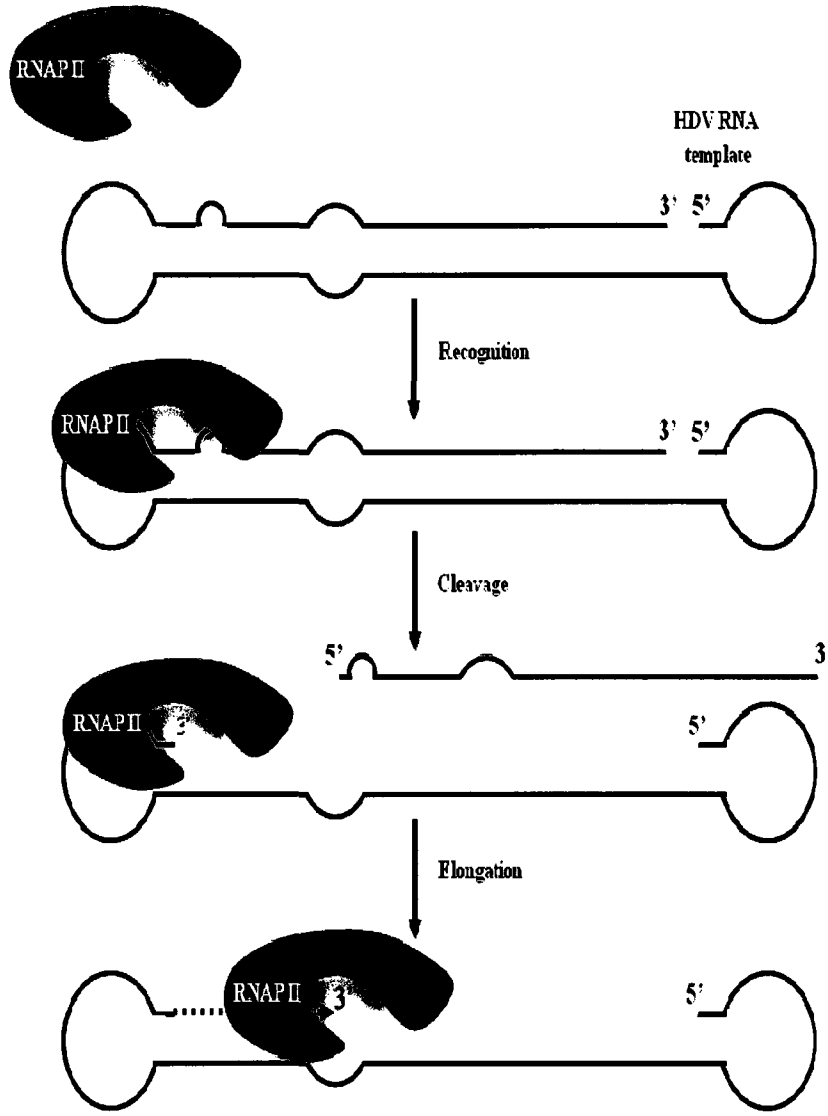
activity of RNAP II [59]. This cleavage is required to form the 3'end needed for the elongation reaction as reported in the study mentioned above [34]. Similar to the previous study, the RNAP II and TFIIS complex cleaved one HDV of the RNA strand and created a reactive stem-loop in the hybrid site that allowed the extension of 3' end of the RNA [60]. The extension of the RNA strand was impaired after the addition of a few nucleotides. The transcription impairment might indicate the requirement of other positive elongation factors or HDAg-S that would facilitate the transcription [60]. Superimposing the DNA and RNA template-product crystal structures on the RNAP II transcription active site showed that they both occupied the same site. This finding suggests that RNAP II is recognizing RNA and DNA templates in similar way [60]. Even though using the same RNAP II transcription active site, RdRP activity is slower and less processive than DdRP activity. This observation could be resulting from RNA secondary structures. The mechanism by which RNAP II is using the left terminal stem loop of antigenomic RNA as template is clearly different from that used for the transcription from right terminal stem loop of genomic polarity. This difference might be related to the presence of the mRNA transcription start site on the putative right promoter region.

1.6 DNA dependent RNA polymerase II

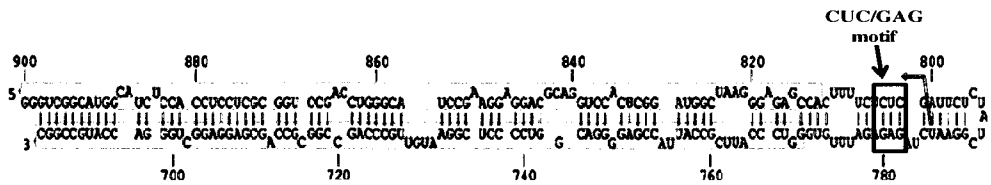
RNAP II is a large complex that consists of 12 core subunits, general transcription factors and other modulator proteins [61]. Two large subunits, named RPB1 (~220 KDa) and RPB2 (140 KDa), form the catalytic domain through which the DNA-RNA assembly takes place during transcription. RPB1 contains a unique CTD that consists of heptapeptide repeats (i. e. YSPTSPS) [62]. These repeats are conserved across organisms, with more complex

Figure 9: Simplified model for transcription from the left terminal extremity of HDV RNA and secondary structure of an RNA promoter located at the left terminal stem loop of HDV antigenomic RNA. **A) (1)** RNAP II recognizes the specific hairpin loop structure, and **(2)** the polymerase initiates the transcription by the cleavage (probably by TFIIS) of the upper strand and starts the elongation process by adding nucleotides to the 3' end of the template up to 41 nts. The model shown is reproduced with permission from the author [56]. **B)** Secondary structure of L213AG used in the study showing the important motifs identified. The green shading represents the delta ribozyme motifs.

A)


RNAP II
HDV RNA
template
3' 5'
Recognition
RNAP II
3' 5'
Cleavage
5' 3
RNAP II
5'
Elongation
RNAP II
5'

B)


CUC/GAG
motif
900
880
860
840
820
800
700
720
740
760
780

organisms having higher numbers of repeats [63]. The phosphorylation of serine residues in the CTD plays an important role in the regulation of transcription. It has been shown that RNAP II enters the pre-initiation complex in the nonphosphorylated state and becomes extensively phosphorylated during the initiation of transcription [64, 65]. During transcription, a CTD phosphatase of 150 kDa (FCP1) recycles the RNAP II by reversing the CTD phosphorylation. The nonphosphorylated RNAP II is reinstated back to the promoter to start a new transcription cycle [8]. In addition, CTD phosphorylation is required for the later stages of mRNA processing, which include polyadenylation, 5' capping and splicing [66]. It is well documented that the 5' end capping of eukaryotic mRNAs takes place immediately after the start of transcription. When nascent RNA chains reach about 25-30 nucleotides in length, the capping enzymes add 7-methylguanosine (m^7G) [67, 68]. The other important mRNA modification process is the polyadenylation of newly synthesized mRNA [69]. Splicing is another process that is coupled with the RNAP II transcription of mRNA. Many different splicing factors are involved in the splicing. Their function is to remove the non coding sequences (introns) from the transcribed mRNA. Some of the splicing factors can function as transcription factors, such as SC35 [70].

1.6.1 Composition of DNA Promoters and Formation of RNAP II Complex

The other components of the RNAP II complex are the general transcription factors. General transcription factors are accessory proteins that play critical roles in transcription, and they are highly conserved among eukaryotes [71, 72]. They interact with core promoter elements on the DNA templates, and these interactions are required for the formation of the pre-initiation complex. Core promoter elements are mostly located between ~40 base pairs

(bp) upstream and ~40 bp downstream of the transcription start site (+1) (Figure 10). However, there is no universal promoter harboring all of these identified elements. The composition of a core promoter and the events occurring prior to the activation of transcription lead to the assembly of the general transcription factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and other modulator proteins) and RNAP II to form a functional pre-initiation complex (Figure 11).

Transcription factor II D (TFIID) is a multi-subunit protein complex that consists of the TATA binding protein (TBP, 38 KDa) and more than 10 other TBP-associated factors (TAFs) that binds to DNA promoters [73, 74]. The binding of TBP to the TATA box, located between the -31 to -26 region of eukaryotic promoters, with the consensus sequences consisting of (TATA[A/T]AA[G/A]) (Figure 10) is sufficient to form a base for RNAP II binding to form the pre-initiation complex including all the general transcription factors and probably other proteins. One or two mismatches have no dramatic effects on promoter recognition, as TBP binding protein recognizes the TATA box through its phosphate backbone [75]. The requirement of a specific sequence might increase the affinity of the protein toward the DNA [75].

It is known that the TATA box is found in some, but not all, core promoters. For example, in humans, only 32% of 1031 potential promoter regions contain a putative TATA box motif, this indicates that proteins other than TBP can compensate for and initiate the transcription in the absence of a TATA box [76]. Martinez and others established the interaction of TFIID with the initiator element (Inr), which has a consensus sequence of mostly pyrimidine bases ([C/T][C/T]A_{+1}N[T/A][C/T][C/T]) (Figure 10), in a sequence-specific manner [62, 77, 78, 79]. Another important promoter element is the downstream

promoter element (DPE), which has a consensus sequence of [G/A]G[A/T][C/T]G/A/C] (Figure 10) [89]. In the absence of a TATA box, TAF1 and TAF2 cooperatively bind to the Inr and DPE elements. Mutations in the DPE or Inr elements cause the loss of TFIID binding activity [80]. From the above mentioned information it is well documented that TFIID binds to both TATA and TATA-less promoters. This binding leads to the assembly of the RNAP II transcription complex on different sets of DNA promoters [81].

The second member of the general transcription factors is TFIIA, which consists of two α (35 kDa) and two β (19 kDa) subunits [82, 83]. TFIIA binds to TBP and DNA directly upstream of the TATA box by binding the DNA phosphate backbone in the TATA box major groove [84]. Its binding is essential for stabilizing the TFIID-promoter complex to form a platform for assembly of other transcription factors [84].

Other important elements for RNAP II complex assembly are the TFIIB recognition elements (BREs). BREs are sequence motifs recognized by TFIIB (35 KDa) in a sequence-specific manner. They are located immediately upstream [upstream BRE (BRE^u)] and downstream [downstream BRE (BRE^d)] of the TATA box in TATA box-dependent promoters. They have two consensus sequences consisting of (BRE^u) [G/C]-[G/C]-[G/A] C G C C and (BRE^d) 5 –[G/A] T [T/G/A] [T/G] [G/T] [T/G] [T/G]-3 (Figure 10) [85]. TFIIB binds to the minor groove immediately downstream of the TATA box and major groove immediately upstream of the TATA box [86]. TFIIB interaction with BRE has been observed to enhance the assembly of the TBP-TFIIB-DNA complex and transcription initiation in an *in vitro* transcription system [87, 88].

TFIIE, a heterotetramer protein which consists of two α (56 KDa) and two β (34 KDa) subunits [89], recruits TFIIH to the initiation complex and plays an important role in

switching from an initiation to an elongation process by having a stimulatory effect on kinase and ATPase activity of TFIIH [81]. TFIIH is a large complex of nine subunits with molecular sizes ranging from 32 to 89 KDa. It is divided into two complexes: core (consisting of five subunits) and CDK-activating complex (CAK consisting of three subunits) [90, 91]. CAK acts to phosphorylate the CTD of the largest subunit of RNAP II [92]. TFIIH has been found to extensively phosphorylate RNAP II CTD in a TFIIE-dependent manner, it carried out the phosphorylation of RNAP II CTD in the absence of a DNA, it also phosphorylates RNAP II CTD in the late stage of pre-initiation complex assembly on DNA promoters. TFIIH has also another phosphorylation activity on TFIID, TFIIE and TFIIF. TFIIF is a heterotetramer comprised of two α (74 KDa) and two β (30 KDa) subunits. TFIIF binds to RNAP II and probably TFIIB where it participates in the formation of the pre-initiation complex. Mutation at two TFIIF-RNAP II binding sites resulted in defects in transcription start site usage [93].

Most of the above described general transcription factors are required for transcription, including the formation of the pre-initiation complex, successful initiation, and elongation processes. The initiation of transcription starts by the phosphorylation of the CTD of the RNAP II by several proteins, such as TFIIH. The RNAP II CTD phosphorylation leads to the initiation of the elongation process, in which some of the transcription factors will leave the promoter (i.e. RNAP II, TFIIB and TFIIF) [94], while the rest of the transcription factors remain on the promoter to form the platform for the formation of RNAP II re-initiation complex (i.e. TFIIA, TFIID, TFIIE, and TFIIH) [94].

Figure 10: Schematic representation of the features found on DNA promoters for RNAP II. A DNA promoter may contain some of these elements: downstream promoter element (DPE), initiator element (Inr), downstream and upstream TFIIB recognition elements (BRE^d, BRE^u), and TATA box. [62, 75, 85, 89]

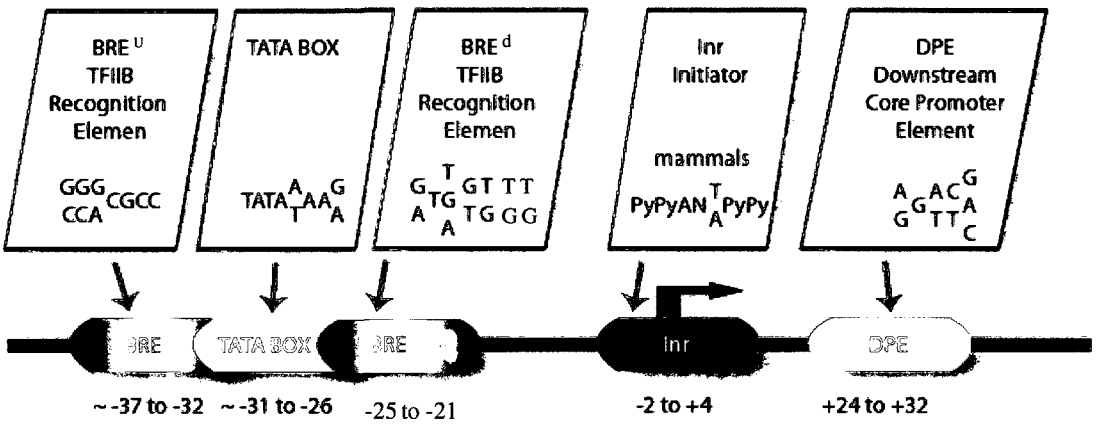
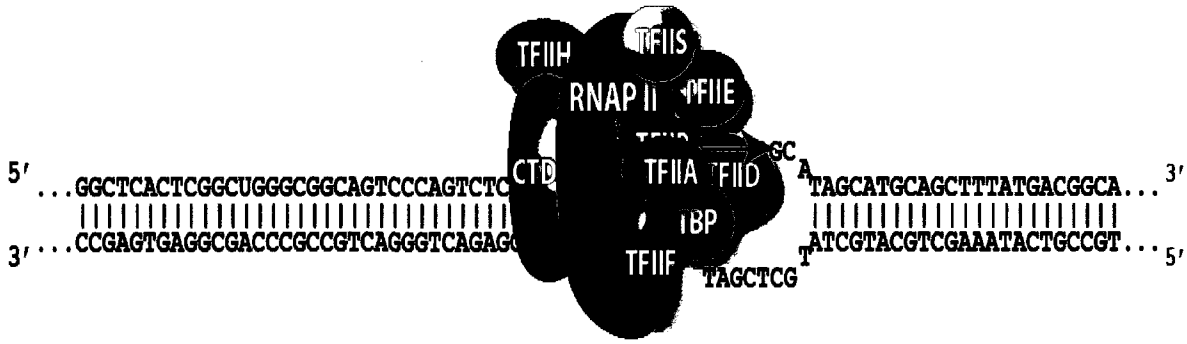

BRE u
TFIIB
Recognition
Elemen
GGG CGCC
CCA
TATA BOX
TATA A AA G
T A
BRE d
TFIIB
Recognition
Elemen
G T GT TT
A TG TG GG
A
Inr
Initiator
mammals
PyPyAN T PyPy
A
DPE
Downstream
Core Promoter
Element
A AC G
G G TT A
C
BRE
TATA BOX
BRE
Inr
DPE
~ -37 to -32
~ -31 to -26
-25 to -21
-2 to +4
+24 to +32

Figure 11: Model of RNAP II transcription from a DNA promoter showing the general transcription factors.


TFIIH
TFIIS
RNAP II
TFIIE
CTD
TFIIA
TFIID
TBP
TFIIF
5′ ...GGCTCACTCGGCUGGGCGGCAGTCCCAGTCTC
3′ ...CCGAGTGAGGCGACCCGCCGTCAGGGTCAGAG
TAGCATGCAGCTTTATGACGGCA... 3′
ATCGTACGTCGAAATACTGCCGT... 5′
TAGCTCG

1.7 RATIONALE, HYPOTHESIS AND OBJECTIVES

1.7.1 RATIONALE.

My main goal is to identify how RNAP II recognizes an RNA promoter. RNAP II is believed to carry out RNA-dependent RNA transcription of HDV RNA, specifically the transcription of the subgenomic mRNA from genomic HDV RNA. A role for RNAP II in HDV mRNA transcription was proposed based on previous data showing that an RNA fragment derived from the right terminal extremity of genomic HDV RNA (R199G) (*i*) includes the reported initiation site for HDAg mRNA transcription (i.e., position 1630; [95]); (*ii*) has an overall rod-like secondary structure conformation that is important for HDV RNA synthesis [51]; (*iii*) initiates transcription of the subgenomic mRNA that is post-transcriptionally modified with a 5'cap and a 3'poly(A) tail, which are hallmarks of transcripts made by RNAP II [35, 95]. Even though RNAP II has been shown to be involved in HDV RNA transcription, the molecular mechanism(s) underlying RNA template recognition by RNAP II has not been very well studied.

1.7.2 HYPOTHESIS

My hypothesis is that both RNA and DNA promoters are recognized in a similar way. Thus, a similarity between RNAP II pre-initiation complex formation on RNA templates and DNA promoters might exist.

1.7.3 OBJECTIVES

In order to test my hypothesis, the following research objectives were undertaken;

(i) To assess the capability of R199G to act as an RNA promoter for RNAP II using an *in vitro* transcription assay.

(ii) To investigate the formation and the composition of the RNAP II pre-initiation complex on R199G by RNA affinity chromatography and western blot analysis.

(iii) To identify the RNAP II subunit(s) directly interacting with the R199G using electrophoretic mobility gel shift assay (EMSA).

Those objectives are aimed at providing a better understanding of the events leading to RNA promoter recognition by RNAP II.

CHAPTER TWO

MATERIALS AND METHODS

2.1 HDV DNA preparation

HDV cDNA dimer, cloned into the plasmid pHDVd2, a derivative of pBluescriptKS$^+$ (Stratagene), was a kind gift of Dr. JP Perreault, Université de Sherbrooke [26]. The insert is flanked by T7 and T3 RNAP promoters. Several DNA fragments required for the synthesis of genomic (R199G) and (X-R199G) and antigenomic (L233AG) RNAs were amplified using polymerase chain reactions (PCR). The reactions were performed with 1 unit of Vent DNA polymerase (New England Biolabs; NEB) in 100 µl PCR reactions. These reactions contained site specific primers (sequences of the primers used are given in Appendix I), 1X ThermoPol Reaction Buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and 0.1 % Triton X-100 at pH 8.8), 100 nM of dNTPs and were subjected to the following cycling parameters using a Mastercycler personal thermal cycler (Eppendorf, Brinkmann intermental Inc, USA). Unless otherwise indicated, PCR parameters were the same in all reactions performed.

PCR cycles for the synthesis of genomic R199G, X-R199G and antigenomic L213AG DNA template			
Step	**Time**	**Temperature**	**Cycles**
Initial Denaturation	2 min	94 °C	1 X
Denaturation	30 sec	94 °C	35 X
Annealing	30 sec	55 °C	
Extension	30 sec	76 °C	

Final Extension	7 min	76 °C	1 X

To generate the X-R199GΔUUA DNA template, two complementary overlapping primers (sense X-199GΔUUA and antisense X-199GΔUUA) containing UUA deletions were used and the DNA was amplified by PCR as indicated below. The PCR products were separated by agarose gel electrophoresis in the presence of SYBR green I (Invitrogen, Canada). The DNA bands were visualized using Kodak Image Station 2000MM (Molecular Dynamics, Canada) and excised from the gel. The agarose gel fragments containing the desired DNA templates were dissolved in gel dissolving buffer at 50^0C for 10 min and the DNA was purified using QIAquick Gel Extraction Kit according to the manufacturer's protocol (Qiagen, Canada). 0.2 pmols of the purified DNA were subjected to a second round of PCR amplification for thirty five cycles. The primers used in the last amplification reaction were sense and antisense R199G. Using the above described approach a full-length X-R199GΔUUA DNA was generated. The PCR cycles were performed as follows;

PCR cycles for the synthesis of X-R199GΔUUA DNA template			
Step	**Time**	**Temperature**	**Cycles**
Initial Denaturation	2 min	94 °C	1 X
Denaturation	30 sec	94 °C	5 X
Annealing	30 sec	55 °C	
Extension	30 sec	76 °C	
Final Extension	7 min	76 °C	1 X

After agarose gel purification, a second round of PCR using the primers required to make the complete X-R199GΔUUA was performed as below			
Initial Denaturation	2 min	94 °C	1 X
Denaturation	30 sec	94 °C	35 X
Annealing	30 sec	55 °C	
Extension	30 sec	76 °C	
Final Extension	7 min	76 °C	1 X

A DNA fragment corresponding to the Cytomegalovirus (CMV) DNA promoter was synthesized as a positive control as follows. Two overlapping sense and antisense CMV primers were used in five PCR reaction cycles at an annealing temperature 55^0C. Using the same conditions, short DNA templates for R38G and its derivatives R38GPU, R38GSW as well as P11.60 were produced using two complementary oligonucleotides for each reaction. A total of 5 cycles of PCR at annealing temperature of 50°C were used in all PCRs. Primers used in these reactions are outlined in Appendix 1.

PCR cycles for the synthesis of (CMV) DNA promoter, R38G, R38GPU, R38GSW and P11.60			
Step	**Time**	**Temperature**	**Cycles**
Initial Denaturation	2 min	94 °C	1 X
Denaturation	30 sec	94 °C	5 X
Annealing	30 sec	55/50 °C	
Extension	30 sec	76 °C	

Final Extension	7 min	76 °C	1 X

The PCR products in all of the above experiments were run on 1.5% agarose gels in the presence of SYBR green I. PCR products were visualized using Kodak Image Station 2000MM (Molecular Dynamics, Canada). Gel bands corresponding to the PCR products were purified using QIAquick Gel Extraction Kit (Qiagen, Canada) as described above.

2.2 RNA synthesis

Transcription reactions were performed to synthesize different RNA fragments using the previously synthesized DNA templates. Sense primers were used to synthesize the aforementioned DNA templates contained T7 RNA polymerase promoter sequences. Transcription reactions of 100 µl were prepared with the desired DNA template in 20 µl of 5X transcription buffer (250 mM Tris-HCl pH 7.5, 75 mM $MgCl_2$, 25 mM dithiothreitol (DTT), 10 mM spermidine) and 2.5 µM NTPs, 20 µl of gel purified DNA template (5–10 pmol), 0.1 U/µl pyrophosphatase (Sigma Aldrich, Canada) and 1U/µl of T7 RNAP. The transcription reactions were incubated at 37°C for 2 h.

To generate the full length genomic RNA transcript, the plasmid pHDVd2 was subjected to *Hind III* (NEB, Canada) digestion prior to its use in the reaction. The digested plasmid was transcribed as described above. As for generation of the full-length antigenomic RNA transcript, pHDVd2 plasmid was digested with *BamH I* (NEB, Canada) and then transcribed in the presence of 1U/µl T3 RNAP (NEB, Canada). Prior to transcription the restriction enzymes were removed from the digestion reactions as follows. DNA templates were subjected to agarose gel electrophoresis in the presence of SYBR green I. The DNA

bands were visualized using Kodak Image Station 2000MM (Molecular Dynamics) and excised from the gel. The DNA templates were purified using QIAquick Gel Extraction Kit according to the manufacturer's protocol (Qiagen, Ontario, Canada)

All transcription products were subjected to 4 U/100 µl of DNase I (Promega, Ontario, Canada) digestion for 30 min at 37°C to digest the DNA templates. The RNAs were fractionated by 10% denaturing polyacrylamide gel electrophoresis (PAGE) containing 1× TBE buffer (100 mM Tris–borate, pH 8.3, 1 mM Ethylene Diamine Tetra-acetic Acid (EDTA)) and 7 M urea. RNA bands were visualized by UV shadowing, excised, the gel was cut to small pieces and RNA was eluted overnight at 4°C in elution buffer (500 mM ammonium acetate, 0.1% sodium dodecyl sulfate (SDS)). The eluted RNAs were washed and precipitated in 3X their volume of 100% ethanol in the presence of 1/10 volume of 3M sodium acetate, pH 5.2. Following precipitation, the pellets were washed with 70% ethanol. All the washing and precipitation steps were centrifuged at 10,000 xg for 10 min. The precipitated RNA pellets were resuspended in 100 µl H_2O and desalted by passing through Sephadex G-50 columns (Amersham Pharmacia Biotech, Canada). The desalted RNA templates were precipitated and washed in ethanol as described above. The pellets were air dried and then resuspended in 20 µl H_2O and stored at −20°C. RNA was quantified by measuring the absorbance at 260 nm by a spectrophotometer (Hermospectronic, Genesys 10 UV, Thermo Electron Corporation, Canada).

In the case of *in vitro* HeLa transcription, the purified RNAs were passed through RNA purification columns from the all-in-One Purification Kit (Norgen, Canada). The columns were washed three times using 1X washing buffer. The RNAs were eluted using RNA elution buffer according to the manufacturer's protocol. The samples were then

subjected to ethanol precipitation and washing as described earlier. All the prepared RNAs were resuspended in H_2O and stored at -20^{0}C for future use.

Radiolabeled RNA templates were generated by transcription reactions under the same conditions as described above with the substitution of 0.16 mM [α-^{32}P] UTP (GE Health Sciences, Canada) for UTP.

2.3 5' Radiolabeling of HDV RNAs

In order to obtain 5' radiolabeled RNAs, 10 pmol RNA of each sample were dephosphorylated for 30 min at 37^{0}C using 0.2 units of calf intestinal phosphatase (CIP) (NEB, Canada) in 0.1 M Tris-HCl, (pH 8). The final reaction volume was 10 µl. The enzyme was then removed using phenol and chloroform extraction. The reaction total volume was brought up to 100 µl with H_2O, 100 µl of phenol was added to the reaction mixture and centrifuged for 15 min at 10000 xg. The upper aqueous phase was removed and placed in a clean tube. An equal volume of chloroform was then added to remove the phenol. The chlorofom extractions were repeated twice by centrifugation at 10000 xg for 10 min. The RNA was precipitated from the aqueous phase by addition of ethanol as described above. Finally, samples were air dried at room temperature for 20 min and resuspended in 20 µl of H_2O. 10 µl of each sample was radiolabeled at the 5'-end using 20 units of T4 polynucleotide kinase (PNK) (NEB, Canada) and 2 µl of 5X PNK buffer (350 mM Tris-HCl, 50 mM $MgCl_2$, 25 mM Dithiothreitol, pH 7.6). The reaction was performed in the presence of 25 µCi/mmol of [γ-^{32}P] ATP (GE Health Sciences, Canada). After 15 min of incubation at 37°C, the labelled RNAs were purified by phenol/chloroform extraction and ethanol precipitation as described above. RNAs were dried and resuspended in 100 µl of

H_2O. The RNA was desalted by passing through Sephadex G- 50 columns. Radiolabled RNAs were ethanol washed and resuspended in H_2O then stored at -20^{0}C.

2.4 HeLa nuclear extracts preparation

HeLa nuclei obtained from Stratagene were used for HeLa nuclear extract (NE) preparation. The nuclei were first washed three times with 1X phosphate buffered saline (PBS: 137 mM NaCl, 12 mM sodium phosphate, 2.7 mM KCl, pH 7.4) and centrifuged at 2000 xg for 10 min. After each washing step, the nuclei were suspended in 10 ml of 1X PBS and transferred into new tubes. 400 ul of 25X protease inhibitors (complete *Protease Inhibitor* Cocktail tablets, Roche) were added to each 10 ml of all buffers. NE preparations were carried out at 4°C. After the nuclei had been washed, the nuclei were resuspended in 10 ml of buffer C (20 mM HEPES-KOH (pH 8.0), 0.6 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25 % (v/v) glycerol, 1 mM DTT). The suspended nuclei were sonicated five times on ice for 30 seconds with 1 minute intervals at power 7.0 (~40 watts) using an automatic sonicator and homogenates were transferred into high speed centrifuge-tubes, mixed gently by rocking, and centrifuged at 16,500 xg for 30 min (Avanti J-25 Centrifuge). The supernatant was transferred into double clipped dialysis bag (12,000-14,000) and dialyzed twice against buffer D (20 mM Hepes-KOH (pH 8.0), 100 mM KCL, 0.2 mM EDTA, 20 % (v/v) glycerol, 1 mM DTT) for two hours. The dialysis bag was placed in a 1- litre beaker on ice. The last dialysis step was done overnight at 4°C. The protein extract was collected from the bag using a Pasteur pipette. Glycerol was added to 40 % to all of the obtained protein samples. The obtained extract was aliquoted and stored at -80 ^{0}C for future use.

2.5 Co-immunoprecipitation of HDV RNAs with RNAP II and TBP antibody

The Protein-G Immunoprecipitation Kit (Sigma-Aldrich, Canada) was employed to co-immunprecipitate HDV RNAs in accordance with the manufacurer's protocol with some modifications. Briefly, 5 pmoles of each HDV RNA were 5' end radiolebled as described above. Radiolabeled RNA fragments were incubated with 260 μg of HeLa nuclear extract proteins at 30°C for 30 min in a 1X ribonucleoprotein immunoprecipitation (RIP) buffer (50 mM Tris–Cl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl). ~5μg of α-RNAP II mouse monoclonal antibody (8WG16; Upstate, USA) directed against the RPB1 CTD of RNAP II or α-TBP rabbit monoclonal antibody (ProteinOne, USA), were added to immunoprecipitation spin columns (Sigma-Aldrich, USA). A fifty fold molar excess of poly (A) RNA as an unrelated competitor and goat anti-mouse IgG antibody as an unrelated antibody were added to different spin columns. The total volume was brought to 60 μl using 1X RIP buffer and incubated at 4°C for 1 h. During the incubation period, 30 μl of the protein G-agarose beads were prepared for each sample by washing the beads three times with 700 μl of 1X RIP buffer. Fifty μl of the bead slurry aliquot was added to each reaction tube. The spin columns were mixed head-over-tail and incubated overnight at 4°C. Each column was then transferred into a 2 ml microcentrifuge tube provided with the kit. The reaction tubes (spin columns) were subjected to five washes using 1X IP buffer at 12,000 xg for 30 seconds at 4°C. The final wash was performed with 0.1X RIP buffer, and the effluent was discarded. Twenty μl of acrylamide loading dye (80% formamide, 10 mM EDTA, pH 8.0, 0.1% xylene cyanol. 0.1% bromophenol blue) was added. The columns were transferred and inserted into a new

microcentrifuge tube and heated at 90 °C for 10 min. The reaction mixtures were eluted at 13,000 xg for 1 min. The collected samples were then fractionated on denaturing 10% TBE 7 M urea PAGE (bisacrylamide: acrylamide, 1:20) in 1X TBE (100 mM Tris–borate, pH 8.3, 1 mM EDTA) buffer. The gel was then exposed to a phosphor screen and scanned using Kodak Image Station 2000MM and visualized using ImageQuant software (Molecular Dynamics, Canada).

2.6 Electrophoretic mobility shift assay (EMSA)

Radiolabeled RNA molecules were incubated with purified proteins. The RNAP II protein complex was purified from HeLa nuclear extract by conventional chromatography and was 60%-70% pure (ProteinOne, USA). RNAP II, TFIID, or Glutathione-S-transferase-TBP (GST-TBP) (ProteinOne, USA) were incubated in binding buffer (20mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 6.0 mM $MgCl_2$, 0.5 mM DTT and 20% glycerol). Unless otherwise indicated, 1 pmol of RNA, 50 ng of RNAP II, 50 ng of TFIID, or 64 ng of GST-TBP were used. After incubation at 30°C for 30 min, the complexes were resolved by 5% nondenaturing PAGE (bisacrylamide: acrylamide, 1:49) under native conditions at room temperature in 1X TBE. The gels were then exposed to a phosphor screen overnight, which was scanned using Kodak Image Station 2000MM and visualized using ImageQuant software (Molecular Dynamics, Canada).

2.7 Binding competition assays

EMSAs were used to investigate the binding specificity of RNAP II, TFIID and GST-TBP to radiolabeled R199G and R38G RNA templates. Proteins were pre-incubated

for 5 min with a 50-fold molar excess of non-radioactive P11.60 RNA, R199G RNA, or CMV DNA promoter. Thereafter, radiolabeled R199G and R38G were added as probes to the binding reactions in different experiments. Binding assays and image analysis were performed as described above.

2.8 HeLa Nuclear transcription assay

R199G (50 pmol) was incubated in 25 μl of transcription buffer (20mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 6.0 mM $MgCl_2$, 0.5 mM DTT and 20% glycerol) containing 0.16 mM [α-^{32}P] UTP (GE Health Sciences, Canada) with either 0.4 mM ATP or 0.4 mM NTPs. Thirteen μg of HeLa nuclear extract was then added and the reaction mixtures were incubated at 30°C for 60 min. The transcription reactions were stopped by adding 175 μl of HeLa extract stop solution (0.3M Tris-HCl pH 7.4, 0.3M sodium acetate, 0.5% SDS, 2mM EDTA and 3μg/ml tRNA). Reactions were then subjected to phenol-chloroform extraction, ethanol precipitation, and resuspended in 40 μl H_2O and 20 μl of acrylamide loading dye. Samples were then fractionated on a denaturing 10% PAGE. The gel was then exposed to a phosphor screen overnight and scanned using Kodak Image Station 2000MM and visualized using Image Quant software (Molecular Dynamics, Canada). In the inhibition study with α-amanitin, HeLa NE was pre-incubated with 30 μg/ml of α-amanitin for 1 hour on ice, and then used in the transcription assay as above.

2.9 Non-radioactive HeLa transcription assay

X-R199G template (50 pmol) was incubated in 25 µl of transcription buffer containing 0.4 mM NTPs and either HeLa NE or HeLa NE pre-incubated with 2.5 µg of α-RNAP II antibody (CTD4H8; Upstate, USA) or α-IgG antibody as unreleated antibody for 1 h on ice. Reactions were incubated at 30°C for 60 min. After that, the reactions were terminated by 175 µl of HeLa extract stop solution. The RNA products were subjected to phenol/chloroform extraction and ethanol precipitation. The pellets were dried at room temperature and resuspended in 40 µl H_2O and subjected to reverse transcription. One µl of 100 nM non-HDV primer (5'-GGGGGGAACAAGGGAC-3') was used to detect the RNA transcript. Reaction mixtures were prepared as follows: 4 µl of the RNA and 10.5 µl of Diethylpyrocarbonate (DEPC) treated water (Superscript™ II RT, Invitrogen, Canada) were heated at 70°C for 10 min to denature the RNA. After that they were chilled on ice to preserve the RNA denaturation. The following components were then added to the reaction: 2.5 µl 10X PCR buffer, 2.5 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP, and 2.5 µl of 0.1 M DTT (dithiothreitol). Final reaction volumes were brought up to 24 µl. The reaction mixtures were preheated at 42°C for 1 min. One µl of SuperScript™ II RT (Invitrogen, Canada) enzyme was added to the mixtures and samples were incubated at 42°C for 50 min. The enzyme was then inactivated by heating at 70°C for 15 min. In order to degrade the RNA template, 1 µl of RNase mix (RNase H and RNase A, SuperScript™ II RT, Invitrogen, Canada) were added and samples were incubated for 30 min. Following RNA degradation the synthesized cDNAs were purified using the cDNA purification kit (Invitrogen, Canada) as follows: 120 µl of the first strand binding solution (6 M NaCl) were added to the prepared cDNA. Samples were transferred to cDNA purification columns, which were centrifuged for

20 s at 13,000 xg. The columns were transferred to new tubes and washed three times using 0.4 ml of cold 1X washing buffer (SuperScript^{TM} II RT, Invitrogen, Canada). The final wash was performed with 0.4 ml of cold 70% ethanol and the cDNA was eluted from the column by H_2O preheated to 65°C. PCR was then performed on the cDNA using either of the reverse transcription primers B (5'-^{1631} TAAGAGTACT GAGGACT GCC ^{1612}-3') or primer A (5'-^{1651}CTCCTCTTTACAGAAAAGAG^{1632}-3'). PCR products were mixed with SYBR green I and resolved on a 2% agarose gel. The gel was scanned using Kodak Image Station 2000MM and visualized using Image Quant software (Molecular Dynamics, Canada).

To further confirm the location of the initiation site, an RNA template with a deletion of nucleotide 1629–1631 was constructed to remove the UUA sequence at the initiation site (i.e. X-R199GUUA). Following *in vitro* transcription and reverse transcription the gel was examined as above.

2.9.1 5' Rapid Amplification of cDNA Ends (5'RACE) of mRNA

In an attempt to identify the 5' end of the RNA transcript, 5'RACE version 2.0 was used (Invitrogen, Canada). Five µl of the purified cDNA obtained from the above RT experiment was subjected to terminal deoxynucleotidyl transferase (TdT) tailing as follows: the reaction was performed in cDNA tailing buffer (10mM Tris-HCl (pH 8.4), 25mM KCl and 1.5 mM $MgCl_2$) supplemented with 200µM of dCTP for 10 min at 37°C, the enzyme was heat inactivated at 65°C for 10 min. The poly(C) tailed cDNA was purified using a cDNA purification column according to the manufacturer's protocol (Invitrogen, Canada). The purified cDNA was then used as a template for the following PCR using the X-primer (5'-GGGGGGAACAAGGGAC-3') and 5'RACE Abridged Anchor Primer (AAP) (5'-

GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3'). A second round of PCR was done using nested gene-specific primers that contained the underlined T7 promoter sequence (5'-GAATTCTAATACGACTCACTATAGGGACTGCTCGATCTCTT-3') and Abridged Universal Amplification Primer (AUAP) (5'-GGCCACGCGTCGACTAGTAC-3'), which specifically binds and amplify the RACE product. The PCR products were sequenced directly by Bio Basic Inc. (Markham, Ontario, Canada). using a T7 primer inserted into the gene-specific primer.

2.10 RNA affinity column chromatography

2.10.1 Formation of RNAP II pre-initiation complex on R199G

RNA molecules (200 pmol) were oxidized in 20 mM Tris-HCl pH 7.5 and 10 mM Na-m-periodate in a total volume of 100 µl. The reaction mixtures were incubated for 1 h at 4°C in the dark, and the oxidized RNAs were precipitated in 3X their volume of 100% ethanol in the presence of 1/10 volume of 3M sodium acetate, pH 5.2. Pellets were washed with 70% ethanol in filter tubes obtained from the IP kit (Sigma-Aldrich, USA). The dried RNA pellets were then resuspended in 60 µl of 0.1 M sodium acetate (pH 5.0). Adipic acid dihydrazide agarose beads (400 µl of 50% slurry; Sigma-Aldrich, Canada) were prepared by washing four times with 0.1 M sodium acetate (pH 5). For direct coupling, the beads were resuspended in 600 µl of 0.1 M sodium acetate and loaded onto filter tubes (Sigma-Aldrich, USA) and incubated with the oxidized RNA overnight at 4°C. After washing the beads four times with 2M NaCl to remove unbound RNA, 15 µg of the HeLa NE were added and the mixtures were incubated at room temperature for 30 min. The columns were washed three times with wash solution (50 mM Tris–Cl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium

deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl) and proteins were eluted using SDS loading dye (50 mM Tris–HCl, pH 6.8, 2% SDS, 0.1% bromophenol blue (BPB) and 10% glycerol Figure 12 step A). The eluted proteins were separated by 10% SDS–polyacrylamide gel electrophoresis (PAGE). The proteins were then transferred onto nitrocellulose membranes (One-Step Western^{TM} Complete Kit Genscript, USA) at 100 V, constant voltage for 1 hour. The membranes were blocked using a 1:1 mixture of buffer A and B (One-Step Western^{TM} Complete Kit Genscript, USA) for 10 min. They were then washed three times using 10 ml of 1X washing buffer (One-Step Western^{TM} Complete Kit Genscript, USA) and probed with a specific antibody against each of the examined proteins for 40 min at room temperature. After that, they were incubated with the modified horseradish peroxidase conjugated antibody to generate a chemiluminescence signal. This signal was amplified by incubation in a 1:1 mixture of buffer A and B of the enhanced chemiluminescence (ECL) western blot detection reagents (One-Step Western^{TM} Complete Kit Genscript). Protein signals were visualized using X-Ray film. Antibodies targeting RNAP II CTD (CTD4H8; Upstate), TFIID (TBP, Upstate), TFIIA (p55, ProteinOne, USA), TFIIE (p56, ProteinOne USA), TFIIF (RAP74, ProteinOne, USA), SC35 (ProteinOne, USA), TFIIB (IIB8; ABR Affinity Bio Reagents, USA), TFIIS (TCEA3, Abcam, USA), and TFIIH (Cdk7, Abcam, USA) were used.

Two oligonucleotides were used to make a DNA molecule containing a CMV DNA promoter which was biotinylated at its 5' end (Invitrogen, Canada). To ensure that both DNA molecules form the DNA promoter, both strands were mixed, heated at 95^{0}C for 10 min and allowed to cool down at room temperature for 20 min. The DNA promoter containing biotin was then bound to streptavidin beads (Sigma-Aldrich, USA). The beads

were pre-washed three times with the 1X wash solution (One-Step Western™ Complete Kit Genscript, USA). The mixture was incubated for 1 h in binding buffer (20 mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT and 20% glycerol) and washed four times with the same solution. Uncoupled streptavidin beads were used as a control. HeLa nuclear extract proteins (15 μg) were added and the streptavidin columns were processed in the same manner as the RNA affinity columns.

2.10.2 Formation of RNAP II initiation complex on R199G

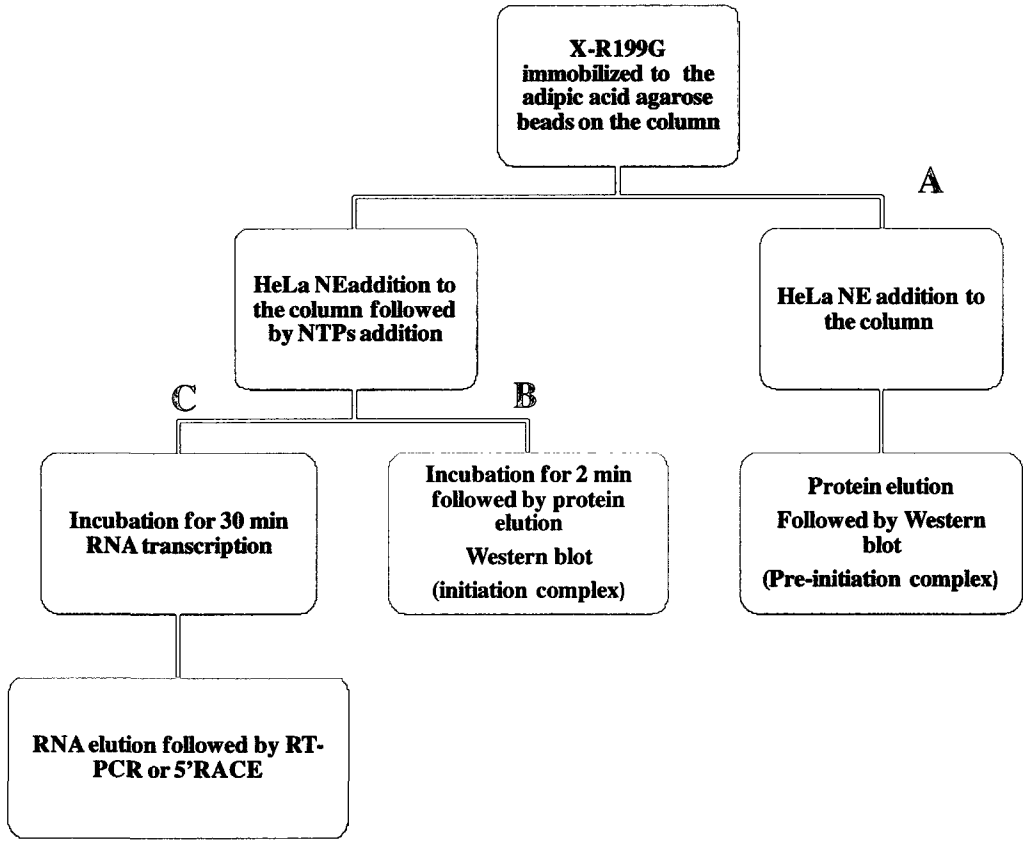
To examine the formation of an initiation complex on the R199G template, the RNA affinity chromatography experiment was repeated using two affinity columns as previously described. Following the washing of the unbound proteins, transcription buffer and 0.4 mM NTPs were added to one of the RNA–protein mixtures. The mixtures were incubated for 2 min at room temperature. After that, the columns were washed three times with wash solution and proteins were eluted using SDS loading dye and were separated by 10% SDS–PAGE. The separated proteins were then transferred onto nitrocellulose membranes (One-Step Western™ Complete Kit Genscript, USA) for 1 hour. The membranes were then blocked and washed three times. The proteins of interest were probed using several antibodies; RNAP II CTD (CTD4H8; Upstate, USA), TBP (Upstate, USA), TFIIA (p55, ProteinOne, USA), TFIIE (p56, ProteinOne, USA, TFIIF (RAP74, ProteinOne, USA), TFIIB (IIB8; ABR Affinity BioReagents, USA). Modified horseradish peroxidase conjugated were added to the antibodies to generate a chemiluminescence signal (One-Step Western™ Complete Kit Genscript, USA). The protein signals were developed and visualized using X-Ray film as described above (Figure 12 step B). These experiments were

all repeated three times. The images representing the results were scanned and processed for further analysis. Image processing was performed using Convas software. The intensity of the pre-initiation complex (-NTP) signals was quantified using the above mentioned software and normalized to 100% and compared to the initiation complex signal intensity (+NTP). Results were analyzed and presented as means ± standard deviation with paired student t- test with Microsoft Excel (version 2003).

2.10.3 RNA transcription experiments on column affinity chromatography

When required for subsequent RT–PCR, RNA affinity chromatography experiments (initiation complex) were processed as above with the exception that the incubation time was for 30 min in the presence or absence of α-RNAP II antibody (CTD4H8; Upstate, USA)) or α-IgG antibody as unrelated antibody. The RNA transcripts were collected by the addition of transcription buffer and centrifugation at 10,000 g for 1 min. The samples were subjected to phenol/chloroform extraction, and ethanol precipitation. The RNA was then resuspended in 40 µl H_2O and analyzed by RT–PCR as described above (Figure 12 step C). 5'RACE were performed as previously described.

Figure 12: Cartoon illustration of the RNA affinity chromatography method. X-R199G immobilized to the adipic acid agarose beads on the column, HeLa nuclear extract addition to the column **A)** Pre-initiation complex formation; **B)** Initiation complex formation upon the addition of NTPs followed by protein elution; and, **C)** RNA transcription on the column after the addition of the NTPs and RNA precipitation followed by RT-PCR and 5'RACE.


X-R199G immobilized to the adipic acid agarose beads on the column
A
HeLa NEaddition to the column followed by NTPs addition
HeLa NE addition to the column
C
B
Incubation for 30 min RNA transcription
Incubation for 2 min followed by protein elution
Western blot
(initiation complex)
Protein elution
Followed by Western blot
(Pre-initiation complex)
RNA elution followed by RT-PCR or 5'RACE

CHAPTER THREE

RESULTS

3.1 HDV RNA of genomic and antigenomic polarities interact with RNAP II

Many studies have suggested a role for RNAP II in HDV replication and transcription, but direct evidence for this role is still missing. As an initial step, a direct association of RNAP II with both genomic and antigenomic HDV RNA polarities was tested. HeLa NE was allowed to form RNA-protein complexes with either internally radiolabeled full-length genomic or antigenomic HDV RNAs. A monoclonal antibody (8WG16) specific for the CTD of RNAP II was added, and the RNA-RNAP II complex was co-immunoprepitated using protein-G beads. After washing away the unbound proteins, the RNA-RNAP II complexes were eluted using a loading dye and were resolved by a 10% denaturing PAGE. Bands corresponding to the genomic and antigenomic RNA were detected in the presence or absence of 50 fold excess of poly (A) RNA. Poly (A) RNA was used as an unrelated competitor to account for nonspecific binding to RNA templates (Figure 13). On the other hand, no radiolabled bands were detected when the RNAP II antibody was substituted with α-IgG (which was used as unrelated antibody). No band was also detected when HeLa nuclear extracts were omitted from the reactions or when fifty fold molar excess of the homologous unlabeled RNA was added to the reaction. In conclusion, the association of the RNAs with RNAP II is specific.

To establish the association of the RNAP II complex with the two HDV RNA extremities that have been shown to act as putative RNA promoters [51, 34], co-immunoprecipitation experiments were performed as described above, but with the substitution of the full length genomic and antigenomic HDV RNAs for R199G and

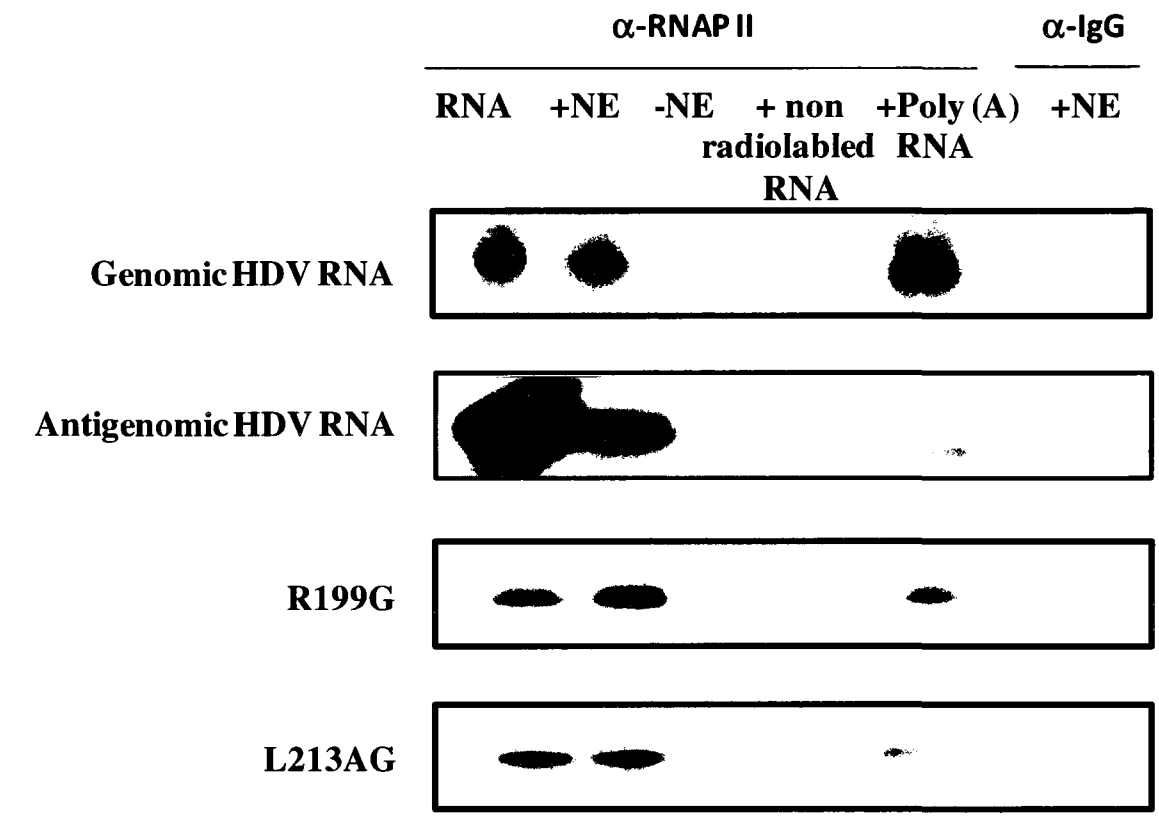
L213AG. Using this procedure radiolabeled bands corresponding to R199G and L213AG were detected in the presence of unrelated competitor poly (A) RNA (Figure 13). However, the radiolabeled RNAs were not co-immunoprecipitated in the absence of HeLa nuclear extracts, or when the RNAP II antibody was replaced by the anti-mouse IgG antibody. At the same time, RNA fragments that corresponds to the double-stranded central region of both polarities of HDV RNA (nts 681 to 54; nts 900 to 1540, which span the regions between the two HDV RNA extremities tested in this study) were found not to be associated with RNAP II [96]. This finding indicated that RNAP II or an RNAP II complex binds to the two HDV RNA extremities (R199G and L213AG).

3.2 Both genomic and antigenomic polarities of HDV RNA interact with TATA-binding protein (TBP)

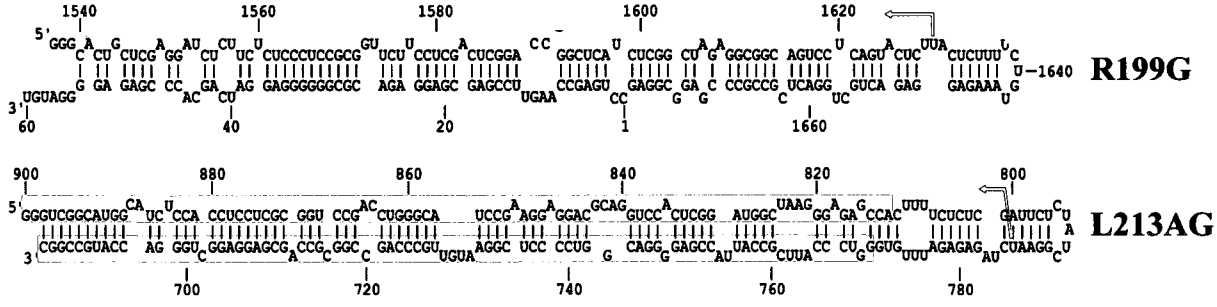
The accurate initiation of transcription by RNAP II on DNA promoters requires the presence of general transcription factors [97, 98]. Usually the first step in forming the pre-initiation complex on a DNA promoter is the binding of TFIID to the promoter [73]. In some cases, the first protein of the TFIID complex that binds to a DNA promoter is the TBP subunit, which then forms a base for RNAP II complex assembly. To investigate if the same mechanism is implicated in RNA promoter recognition, TBP interaction with four HDV derived RNA molecules was investigated. These RNAs were allowed to form complexes with NE proteins in the presence or absence of either 50-fold excess of non-radioactive homologous RNAs or 50-fold excess of poly (A) RNA as an unrelated competitor. RNA Protein complexes were co-immunoprecipitated using a monoclonal antibody against TBP. Mouse α-IgG was used as a control antibody. The formed RNA-protein complexes were

Figure 13: Denaturing 10% TBE-urea PAGE showing interaction of RNAP II with HDV-derived RNAs and secondary structure of the two extremities (R199G, L213AG). **(A)** Co-immunoprecipitation of radiolabeled full length genomic, antigenomic HDV RNAs, R199G and L213AG with RNAP II in HeLa nuclear extract. 50-fold molar excess of nonradioactive homologous RNAs and poly (A) RNA was used in competition experiments. Mouse α IgG was used as unrelated antibody. **(B)** Predicted secondary structure of both R199G and L213AG showing the folded rod like structure, the start site in both RNAs indicated with red arrow. The blue and the green shaded sequences are HDAg mRNA and delta ribozyme motifs.

A)


α-RNAP II
α-IgG
RNA
+NE
-NE
+ non radiolabled RNA
+Poly (A) RNA
+NE
Genomic HDV RNA
Antigenomic HDV RNA
R199G
L213AG

B)

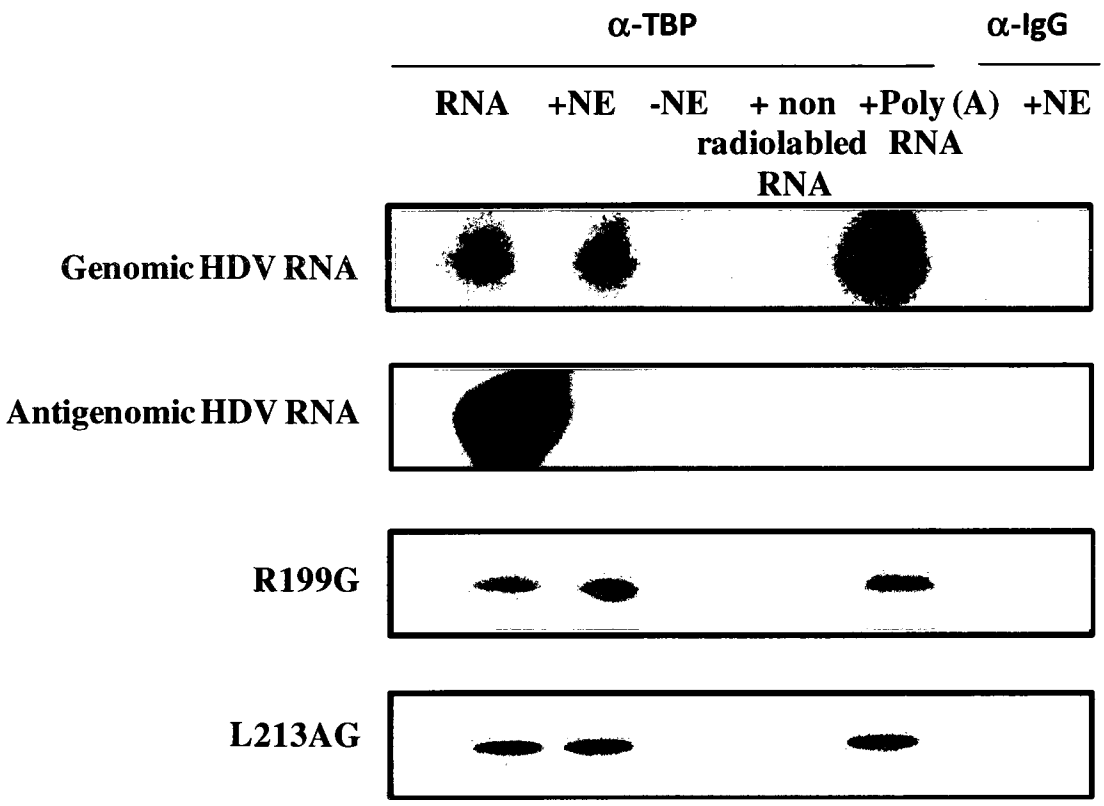

R199G
L213AG

eluted using loading dye and were then separated by denaturing 10% TBE-urea PAGE. Under these conditions, radiolabeled RNA corresponding to the full-length genomic and antigenomic molecules as well as the two extremities of HDV (R199G, L213AG) were co-immunoprecipitated from HeLa NE with the α-TBP (Figure 14). No bands were detected in the absence of HeLa NE or when the mouse α-IgG antibody was substituted for the α-TBP antibody. The ability of the TBP antibody to associate with the RNA-protein complex in the presence of the unrelated RNA competitors pointed to the specific association of TBP or a TBP-containing complex with the HDV RNAs. To conclude, my results clearly showed that this transcription factor associates with both HDV RNA extremities and demonstrated the association of RNAP II to HDV RNA either directly or through other proteins.

3.3 R199G acts as an RNA promoter for RNAP II

The previous experiments confirmed the association of RNAP II and TBP with the two RNA fragments corresponding to the segments located at the extremities of the rod like structure of HDV RNA. One of these segments is R199G. Based on the previous studies, R199G has been found to: (*i*) include the reported initiation site for HDAg mRNA transcription (i.e., position 1630; [95]); (*ii*) be probably used by RNAP II, because the HDAg mRNA is post-transcriptionally processed with a 5'-cap and a 3'-poly (A) tail [35, 95]; (*iii*) fold into an overall rod-like conformation important for HDV RNA synthesis [51]. To corroborate these findings with RNAP II and TBP association (Figures 13 and 14) and to investigate if this segment of the HDV RNA genome can act as an RNA promoter, a transcription experiment using HeLa nuclear extract and R199G RNA as a template was performed. In addition, since end-labeled products are often an artifact of *in vitro*

Figure 14: Denaturing 10% TBE-urea PAGE showing the interaction of TBP with HDV-derived RNAs. Co-immunoprecipitation of full-length genomic, antigenomic HDV-RNAs and the two extremities (R199G and L213AG) using α-TBP. The presence and absence of HeLa nuclear extract (NE), 50-fold molar excess of non-radiolabeled homologus HDV RNAs or 50-fold molar excess of poly (A) RNA were used in competition experiments in the presence of HeLa nuclear extract proteins. Mouse anti IgG antibody was used as unrelated antibody.

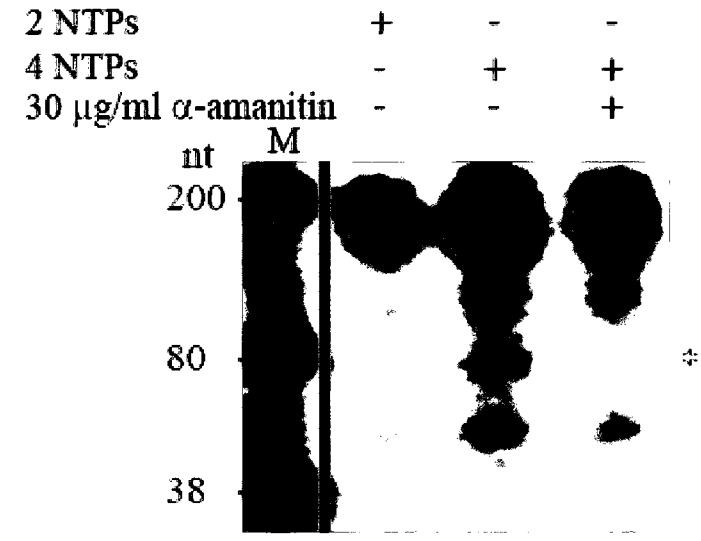

α-TBP
α-IgG
RNA
+NE
-NE
+ non radiolabled RNA
+Poly (A) RNA
+NE
Genomic HDV RNA
Antigenomic HDV RNA
R199G
L213AG

transcription assays on small RNA templates [99, 100], the assays were carried out in the presence of either two nucleotides (i.e. ATP and [α-^{32}P]UTP) or all four nucleotides (NTPs and [α-^{32}P]UTP). These conditions should differentiate between template extension (i.e. 3' end addition) and the synthesis of new RNA molecules (i.e. transcription). When the assays were performed with only two nucleotides, a band corresponding to template extension was detected primarily in the upper portion of the gel (Figure 15A, 2NTPs). Under transcription conditions (i.e. with the four NTPs present), an additional band corresponding to a short RNA product of about 80 nts was detected (Figure 15A, 4NTPs). To confirm the involvement of RNAP II, the transcription reaction was repeated in the presence of α-amanitin. Under our conditions, α-amanitin at 30 μg/ml strongly inhibits the production of the 80-nt transcription product (Figure 15A 4NTPs + α-amanitin). No effect was observed where less than 30 μg/ml of α-amanitin was used, which might put some doubt about RNAP II involvement in the transcription. In addition, the radiolabeled product did not migrate to the expected size. The expected transcription product should be around 90 nts. The smaller band size than expected could be due to improper gel migration caused by RNA secondary structure or due to the abortion of the transcription before it reaches the end of the template. Another possibility is that the transcription started a few nucleotides downstream of the proposed transcription start site.

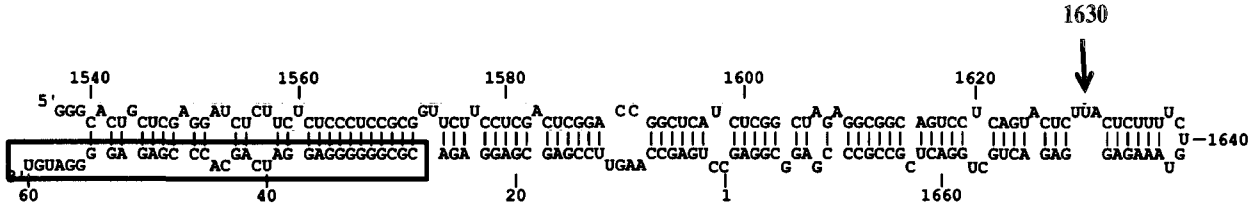
To corroborate that the observed ~80 nts radiolabeled band was an RNA corresponding to the R199G transcription product, RT-PCR reaction was performed following the transcription reaction using a primer that recognizes the 3'end of the expected transcription product. No PCR product corresponding to the cDNA of the transcription

Figure 15: Denaturing 10% TBE-urea PAGE (bisacrylamide: acrylamide, 1:20) showing *in vitro* RNA product using R199G as an RNA template, and predicted secondary structure of R199G. **A)** Transcription reactions performed on R199G using HeLa NE proteins in the presence of either two NTPs (ATP and [α-^{32}P]UTP), four NTPs (NTP and [α-^{32}P]UTP), or four NTPs with 30 μg/ml α-amanitin. The upper band corresponds to the labelling of the R199G template and an asterisk indicates the ~80 nts RNA product. A different exposure of the gel was used for the RNA markers.**B)** Unbranched rod-like structure of R199G showing the secondary structure and sequence. The red arrow represent a start site predicted.

A)



B)



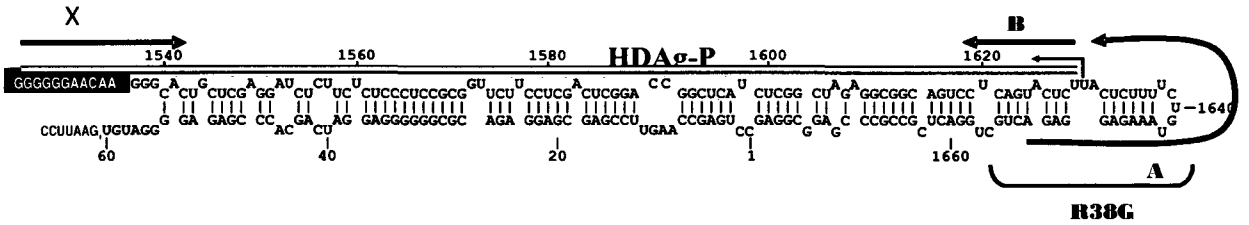
Similar sequence to the transcription product

product was detected. As the RNA templates were present in large molecular excess compared to the RNA product, and due to the high sequence similarity between them (see the green box in Figure 15B), it is likely that the RT-PCR was not able to discriminate between the RNA template and the RNA transcript. Therefore, a non-HDV sequence was added to the 5'end of R199G to generate an extended RNA template (i.e. X-R199G, Figure 16A). The transcription reaction was performed on X-R199G, and the resulting RNA product was subjected to RT using primer X (a primer complementary to the 3' extended extremity of the transcription product). The cDNA product was amplified using primer X and primers that corresponded to the sequences located either before (i.e. primer B) or after (i.e. primer A) the initiation site of HDAg mRNA (Figure 16A; primers A and B). Using this approach, a specific RT-PCR product that migrated between the 100 and 200 nts markers was detected when the transcription reaction was performed in the presence of X-R199G (Figure 16B). This indicated that the produced RNA is complementary to the upper strand of the RNA template (i.e. of antigenomic polarity). No product was detected when X-R199G was omitted during the transcription reactions, which indicated that the product observed is dependent on the RNA template. More importantly, the product was only detected when primer B was used during the PCR amplification, suggesting that the initiation of RNA synthesis occurred near the proposed initiation site of HDAg mRNA.

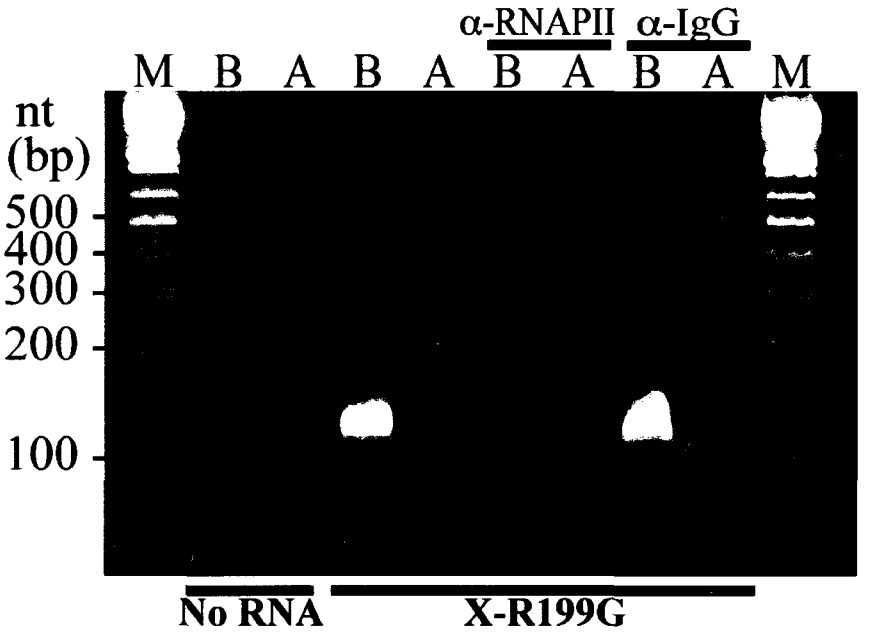
To authenticate the involvement of RNAP II in this RNA-dependent reaction, the transcription reaction was repeated in the presence of a monoclonal antibody specific for the CTD of RNAP II (CTD4H8). This antibody can recognize both phosphorylated and unphosphorylated forms of RNAP II CTD [101, 102]. As reported previously, by binding to

Figure 16: Predicted X-R199G secondary structure and agarose gel electrophoresis showing the RT-PCR product of HeLa NE *in vitro* transcription using X-R199G as template. **A)** Representation of the unbranched rod-like secondary structure of X-R199G. The initiation site of transcription as determined in this study by 5' RACE is marked with the red arrow. The blue box indicates the 5' extremity of the HDAg gene, the inverted characters (X) represent the nucleotides that were added to R199G to form X-R199G, and the square bracket represents the location of R38G. Locations of primers used in this study are indicated. The yellow box indicates the transcription product **B)** *In vitro* transcription using X-R199G as template with HeLa NE proteins followed by RT to amplify cDNA using primer X and PCR using primer X and either primers A and B. The transcription reaction is inhibited by a monoclonal antibody against the CTD of RNAP II. Lane M contains 1 Kb plus DNA ladder (Invitrogen).

A)


X
B
1540
1560
1580
HDAα-P
1600
1620
A
R38G

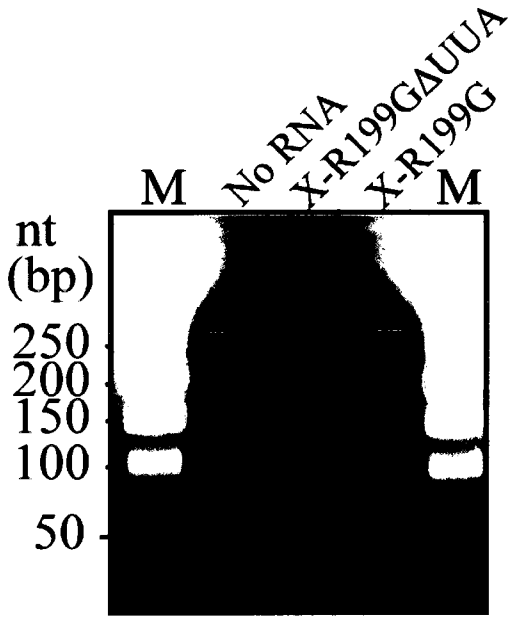
B)


α-RNAPII
α-IgG
M B A B A B A B A M
nt
(bp)
500
400
300
200
100
No RNA
X-R199G

the CTD domain, the antibody should inhibit the RNAP II activity. When the HeLa NE was pre-incubated with the CTD4H8 antibody, the transcription from X-R199G was strongly inhibited (Figure 16B), as indicated by the presence of only a very weak band corresponding to the RT-PCR product. As a control for the non-specific effects of the antibody, the RNAP II antibody was replaced with an α-IgG antibody, which showed no inhibitory effect on the transcription. Together, these data indicate that X-R199G, which corresponds to the right terminal stem-loop domain of the genomic polarity of HDV RNA contains an RNA promoter and that the CTD of RNAP II is required for transcription from this RNA template.

To precisely map the initiation site of transcription from X-R199G, the cDNA products were subjected to 5'RACE as follows. The transcription products were collected from the columns and reverse transcribed using reverse transcriptase III enzyme in the presence of primer X. The synthesized cDNAs were tailed with CTPs using terminal deoxynucleotidyl transferase (TdT). Abridged anchor primers from the 5'RACE kit version 2.0 (Invitrogen, Canada) were used to amplify the products, which were then subjected to a second round of PCR using the anchor bridged primer and a nested product specific primer that contains the T7 promoter sequence in its 5' end. The PCR products were then resolved on an agarose gel. A specific RT-PCR product that migrated between the 100 nts and 150 nts markers was detected when the RACE reaction was performed in the presence of X-R199G (Figure 17). The PCR product was sequenced directly by Bio Basic Inc (Markham, Ontario, Canada). The sequence indicated that transcription initiated near the proposed initiation sites for HDAg mRNA at position 1630nts [95], as shown by the red arrow in figure 16A. This finding is supported by a previous report, that used a similar HDV-derived

Figure 17: Agarose gel electrophoresis of cDNA resulting from 5' RACE reactions using X-R199G and X-R199GΔUUA RNA in *in vitro* transcription experiments. 5' RACE reactions performed on the transcription product with no RNA, X-R199G or X-R199GΔUUA as template. Lanes M contain the MiniSizer 50-bp DNA ladder (Norgen). The black blob is the bromophenol blue dye.


M
No RNA
X-R199GΔUUA
X-R199G
M
nt
(bp)
250
200
150
100
50

RNA molecule as a template in an *in vitro* HeLa nuclear extract transcription system, demonstrating that the initiation site of HDV mRNA synthesis occurs near this location [51]. To further support this initiation site, the transcription reaction was repeated with X-R199GΔUUA, an RNA template with nucleotides deletion at 1629(UUA)1631 nts. Following transcription, RT and 5'RACE reactions, no product was detected (Figure 17).

3.4 RNAP II binds directly to R199G

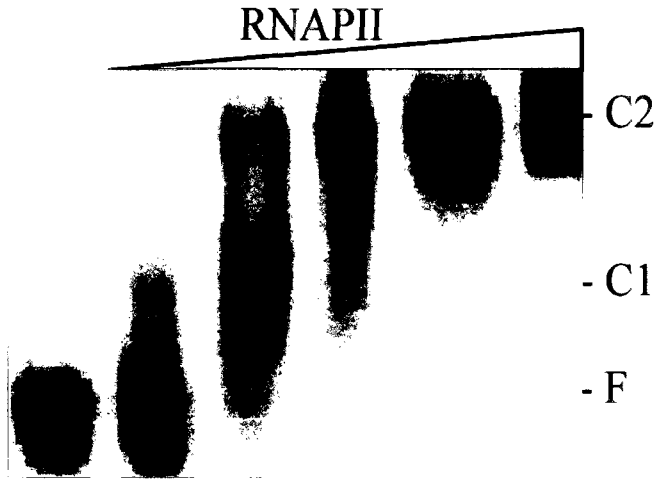
For initiation to occur, the RNAP II holoenzyme must bind to promoter sequences to form a pre-initiation complex. To assess the direct binding of RNAP II to R199G, I performed EMSAs using a native RNAP II complex that had been purified from the HeLa cell nuclear pellet through conventional chromatography (ProteinOne, USA). Radiolabeled R199G was allowed to bind to increasing concentrations of RNAP II and the mixture was subjected to non-denaturing PAGE. The RNAP II bound directly to R199G, as illustrated by a slower migration of the complex when compared to that of free R199G (Figure 18). Notably, two retarded species were detected, and their formation was dependent on the RNAP II concentration. A fast migrating species first formed at low concentrations of RNAP II holoenzyme (Figure 18A, C1), followed by a slower migrating species formed at higher protein concentrations (Figure 18A, C2). This result suggested the involvement of at least one protein in the formation of the first complex; when the concentration of RNAP II is increased, it might subsequently recruit the rest of the proteins, thus forming a larger RNAP II-R199G complex. Next, I investigated the specificity of the interaction between R199G and RNAP II by challenging the formation of the complex with various competitors. A small

RNA hairpin derived from the left terminal stem-loop domain of the peach latent mosaic viroid genome (P11.60) [97] was added as an unrelated RNA competitor. This small hairpin like RNA acts as an excellent RNA competitor in order to test for RNAP II non-specific binding to double-stranded regions and stem-loops. As shown in figure 18B, 50-fold molar excess of P11.60 did not displace RNAP II from R199G. However, inclusion of 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G) or a control DNA promoter fragment (i.e. CMV promoter) greatly inhibited the formation of the radiolabeled R199G-RNAP II complexes.

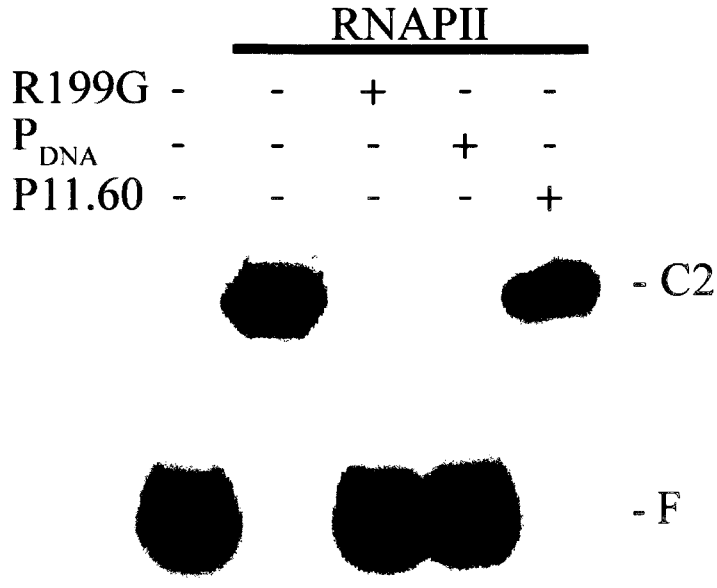
To provide evidence that RNAP II was binding to an internal site on R199G close to the initiation site, I performed EMSA with R38G, which corresponds to a 38 nts RNA located at the tip of the hairpin (Figure 19A). As observed for R199G, two retarded species were detected when radiolabeled R38G was allowed to bind to RNAP II (Figure 19B). To investigate the secondary structure and rod-like requirement for RNAP II binding, EMSA was performed with two R38G derivatives which were previously demonstrated by Pelchat's group to be deficient in their interaction with the largest subunit of RNAP II by co-immunoprecipitation [93]. R38GPU contains mutations that change the pyrimidine-rich sequence that is located upstream from the terminal loop to purines, thus destabilizing the rod-like secondary structure of the RNA molecule. R38GSW is a "flip mutant" of the region; the pyrimidine-rich sequence strand is replaced with a purine-rich sequence strand and vice-versa. For both mutants, the quantity of the two retarded species was reduced, compared with the wild type R38G (Figure 19B). The relative intensities of the fast migrating species were reduced to 65% and 31% for R38GPU and R38GSW, respectively

Figure 18: EMSAs showing the binding of RNAP II to R199G. **A)** Titration of a constant amount of radiolabeled R199G with increasing amounts of purified RNAP II holoenzyme (0, 6.25, 12.5, 25, 50, and 100 ng). **B)** EMSA showing the specific binding of R199G to the purified RNAP II holoenzyme in the presence of 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G), a control DNA promoter fragment (i.e. CMV promoter(P_{DNA})), or P11.60. In all binding assays the free RNA (F) and both complexes observed (C1 and C2) are indicated.

A)



B)



(Figure 19B, C1). The slower migrating species were reduced to 25% and 4% for R38GPU and R38GSW, respectively (Figure 19B, C2). These results correlate with a previous report from Dr. Pelchat's group that suggests that the maintenance of the rod-like conformation and polarization in the purine/pyrimidine content near the tip of the rod-like structure of both polarities of HDV RNA might be required for RNAP II interaction [96]. More importantly, these results provide direct evidence that the RNAP II holoenzyme interacts with an internal site within R199G.

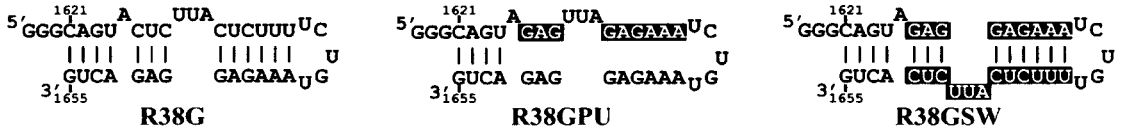
3.5 An active pre-initiation complex forms on R199G

If similar mechanisms exist for DNA- and RNA-directed transcription by RNAP II, the specific binding of the RNAP II to the HDV RNA promoter (i.e. R199G) should require the coordinated assembly of RNAP II and general transcription factors. To identify the transcription machinery recruited to form a pre-initiation complex on R199G, I used an RNA affinity purification procedure that involves cross-linking of R199G to adipic acid dehydrazide agarose beads. The RNA binds to the beads by oxidizing the RNA using Na-m-periodate [99]. Such an approach allowed me to monitor the entire transcription machinery assembly on R199G utilizing active HeLa nuclear extract proteins.

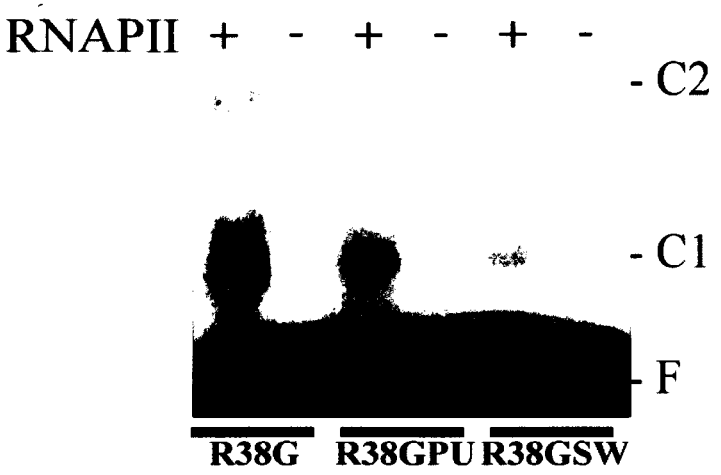
To compare the pre-initiation complex formation between R199G and a control DNA promoter, a DNA fragment containing the CMV promoter was biotinylated and immobilized on streptavidin agarose beads. As a negative RNA control, P11.60 was also immobilized on adipic acid dehydrazide agarose beads, and subjected to the same treatment as R199G. In addition, the complexes on R199G were challenged with a large molecular

Figure 19: Predicted secondary structure of R38G, R38GPU and R38GSW and EMSA revealing the binding of R38G and two RNA mutants to purified RNAP II holoenzyme. **A)** Representation of the predicted secondary structures of R38G, R38GPU and R38GSW. Inverted characters represent mutations from R38G. **B)** EMSA showing the binding of R38G and the two RNA mutants to the purified RNAP II holoenzyme in the presence of a 50-fold molar excess of P11.60. The free RNA (F) and both complexes observed (C1 and C2) are indicated.

A)



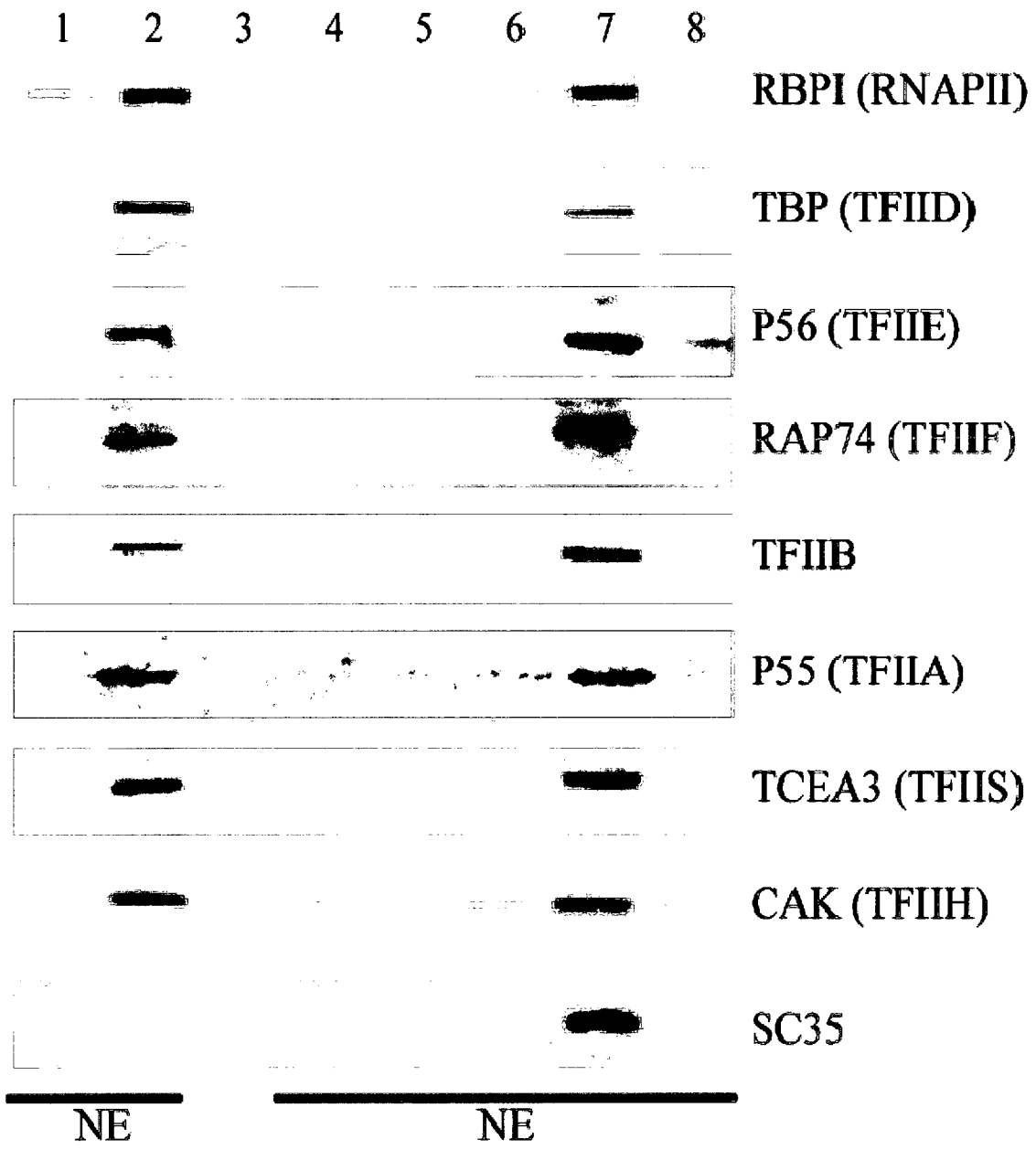
B)



excess of either the DNA promoter fragment or R199G to assess the specificity of pre-initiation complex formation. Finally, unbound adipic acid dehydrazide agarose and streptavidin agarose beads were tested for protein interaction to rule out a non-specific binding of nuclear extract proteins to the beads.

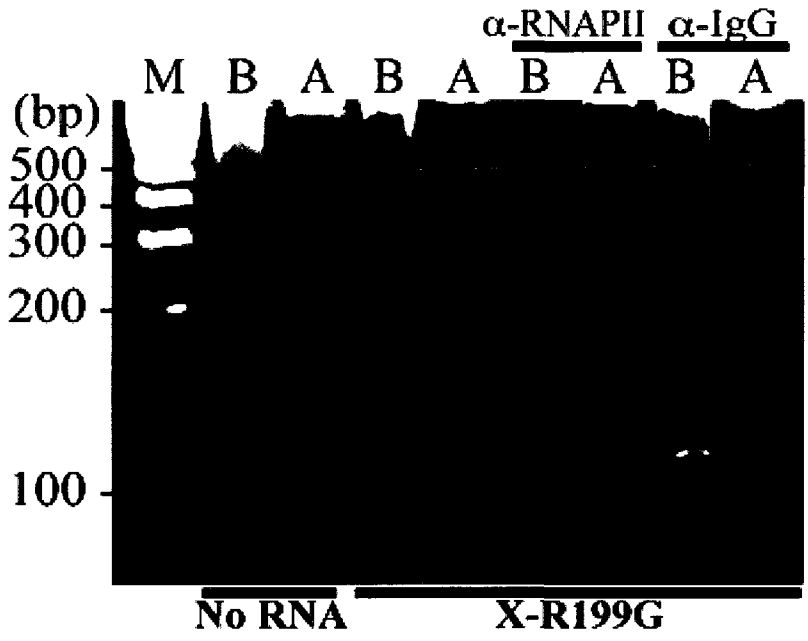
Nucleic acid-coupled beads were incubated with HeLa NE proteins and washed to remove unbound proteins. The ribonucleoprotein complexes were then eluted with SDS-loading dye and heated for 10 min at 90^0C. The eluted RNA proteins were separated by SDS-PAGE, and analyzed by Western blot using antibodies raised against various proteins which are involved in the formation of the pre-initiation of RNAP II for transcription from DNA promoters. Using this approach, the presence of RPB1 (RNAP II), p55 (TFIIA), TFIIB, TBP (TFIID), p56 (TFIIE), RAP74 (TFIIF), CAK (TFIIH), and TCEA3 (TFIIS) in association with both R199G and the DNA promoter were detected (Figure 20). As expected, no significant binding of the RNAP II transcriptional machinery to either P11.60 or to the beads alone was observed, which indicated that the interactions with R199G were specific. Furthermore, molecular excess of unlabeled DNA promoter and R199G was able to abolish the protein interactions with R199G. In addition, because HDV was previously suggested to be associated with SC35-containing nuclear speckles [104], I tested for the presence of SC35 which could serve as a control to differentiate between the proteins binding to R199G, but not to the DNA promoter. SC35 interacted with R199G but not with the DNA promoter (Figure 20). This experiment suggested the formation of a pre-initiation complex on the R199G RNA promoter that is similar to the pre-initiation formed on a CMV DNA promoter (P_{DNA}).

Figure 20: Western blot analysis illustrating the formation of an active pre-initiation complex on R199G. The samples were incubated with HeLa NE proteins in the presence or absence of either molecular excess of free DNA promoter or R199G. After elution, the proteins were detected using antibodies specific for RNAP II and its general transcription factors. Lane 1, streptavidin agarose beads; lane 2, the biotinylated control DNA promoter bound to streptavidin agarose beads; lanes 3 and 8, adipic acid agarose beads; lane 4, P11.60 bound to adipic acid agarose beads; lane 5, R199G bound to adipic acid agarose beads in the presence of molecular excess of free control DNA promoter; lane 6, R199G bound to adipic acid agarose beads in the presence of molecular excess of free R199G; lane 7, R199G bound to adipic acid agarose beads. With the exception of lane 3, all samples were incubated with HeLa NE.


1
2
3
4
5
6
7
8
RBPI (RNAPII)
TBP (TFIID)
P56 (TFIIE)
RAP74 (TFIIF)
TFIIB
P55 (TFIIA)
TCEA3 (TFIIS)
CAK (TFIIH)
SC35
NE
NE

Although R199G is an RNA promoter for RNAP II, it was possible that the observed interactions of R199G with RNAP II and its general transcription factors were not related to pre-initiation complex formation and transcription initiation. A modified affinity purification procedure was performed to test this possibility. After removing unbound proteins, the ribonucleoprotein complexes were equilibrated with a transcription buffer, and then free NTPs were added to allow the initiation of transcription to occur. After elution of the sample, I was unable to detect the presence of the 80 nts RNA product, probably due to the low efficiency of the RNA-directed transcription reaction (Data not shown). As it is a more sensitive assay, RT-PCR experiments were performed to confirm the transcription initiation site and the involvement of RNAP II in this transcription reaction. The transcription reactions were performed on X-R199G-coupled beads and RT-PCR was performed after the elution of the RNA transcripts from the columns using the transcription buffer. The specific RT-PCR product was detected when the transcription reaction was performed in the presence of X-R199G (Figure 21), which demonstrates the RNA synthesis by the immobilized RNAP II complex. As observed above, the transcription product from the bound RNA template was only detected when primer B was used during the PCR amplification, which indicates that the initiation of synthesis occurred near the proposed initiation site for HDAg mRNA. No product was detected in the transcription reaction when X-R199G was omitted, or when the NE was pre-incubated with anti-RNAP II. In addition α-IgG antibody had no inhibitory effects on the transcription (Figure 21). The exact location of the transcription initiation was confirmed by subjecting the cDNA product to 5' RACE, as above. Analysis of the sequence of the RACE product indicated that transcription initiated near the mRNA initiation site at position 1630 nts [95]. Taken together, my results indicated

Figure 21: Agarose gel electrophoresis of cDNA resulting from RT-PCR of the transcription product of X-R199G-coupled bead in *in vitro* transcription experiments. Transcription reactions were performed on the X-R199G-coupled beads followed by RNA extraction and RT-PCR. RT-PCR products were generated using primers complementary to sequences located either before (i.e. primer B) or after (i.e. primer A) the proposed HDAg mRNA 5'end. Lane M contains 1 Kb plus DNA ladder (Invitrogen). For some reactions, monoclonal antibodies against either the CTD of RNAP II or IgG were used.

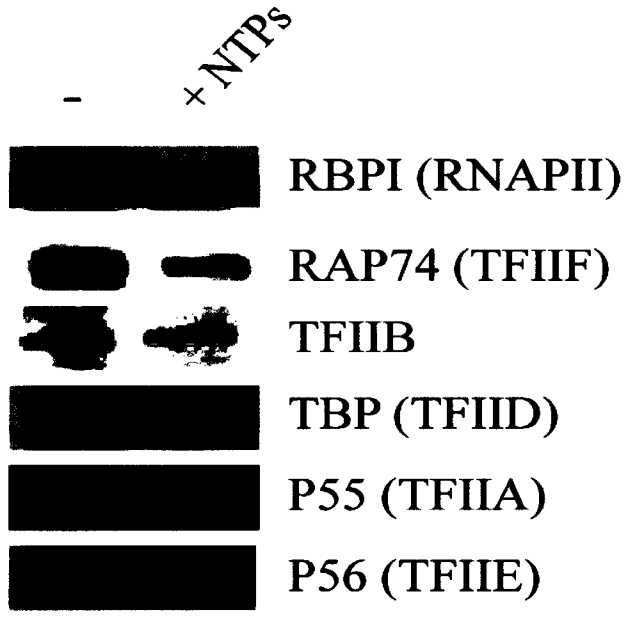

α-RNAPII
α-IgG
M B A B A B A B A
(bp)
500
400
300
200
100
No RNA
X-R199G

that an active RNAP II pre-initiation complex forms on the bound RNA template and that this complex is able to perform the transcription reaction.

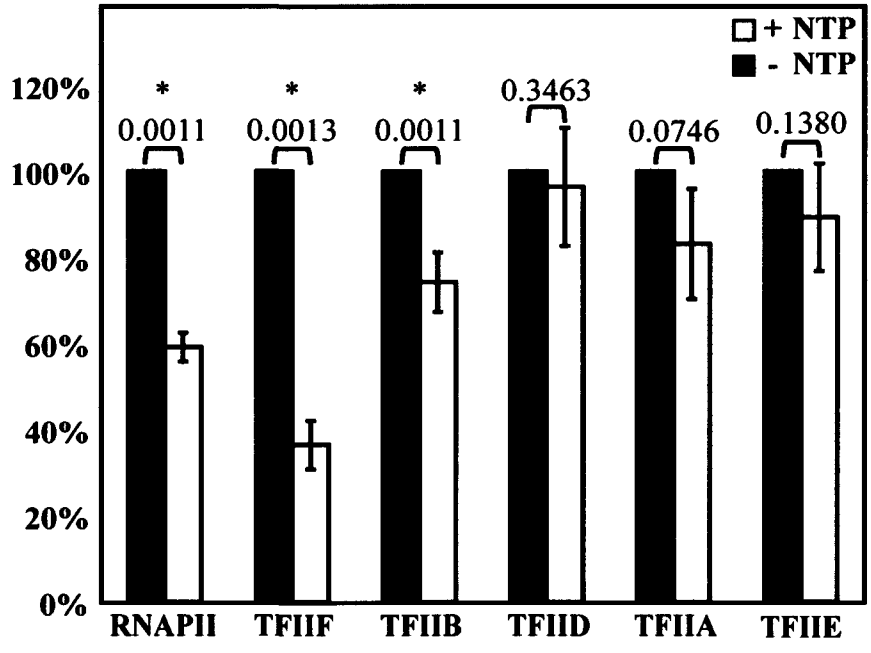
Using the same modified affinity purification procedure, I looked for the RNAP II subunits following transcription initiation on R199G coupled beads to investigate the formation of the initiation complex. The bound X-R199G was incubated with HeLa NE proteins and the beads were washed to remove unbound proteins. Following the addition of free NTPs, the ribonucleoprotein complexes were then eluted with SDS-loading dye for 10 min at 90^0C. The eluted proteins were separated by SDS-PAGE, and analyzed by Western blot using antibodies against some of the proteins involved in the formation of RNAP II transcription initiation complex on DNA promoters. The addition of free NTPs significantly reduced the levels of RPB1 (RNAP II), TFIIB and RAP74 (TFIIF), on X-R199G (Figure 22). The experiment was repeated three times and p-values (Paired Student's t-Test) were obtained, which indicate significant differences (p-value < 0.05). In contrast, there was no significant release of either p55 (TFIIA), TBP (TFIID), or P56 (TFIIE) (Figure 22). Such a phenomenon is in agreement with what was observed to occur on canonical DNA promoters [94]. Specifically, the core subunits of RNAP II and various transcription factors leave the DNA promoters after initiation, whereas a subset of the transcriptional machinery (i.e. TFIIA, TFIID, TFIIE, and TFIIH) remains to form a scaffold for the assembly of a second transcription complex [94]. Taken together, these findings lead me to conclude that an RNAP II PIC forms on X-R199G, and that this complex contains the core RNAP II subunit and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and TFIIS.

Figure 22: Western blot analysis displaying the formation of an active initiation complex on X-R199G and its statistical analysis **A)** Decrease in the levels of various pre-initiation complex components following initiation of transcription. **B)** The experiment was repeated three times and p-values (paired Student's t-Test) were obtained (indicated above the bars). Asterisks indicate significant differences (p-value < 0.05).

A)


-
+ NTPs
RBPI (RNAPII)
RAP74 (TFIIF)
TFIIB
TBP (TFIID)
P55 (TFIIA)
P56 (TFIIE)

B)


+ NTP
- NTP
120%
100%
80%
60%
40%
20%
0%
*
*
*
0.3463
0.0746
0.1380
0.0011
0.0013
0.0011
RNAPII
TFIIF
TFIIB
TFIID
TFIIA
TFIIE

My results also suggest that a re-initiation complex remains on the RNA promoter following the initiation of transcription, a process analogous to that observed on DNA promoters.

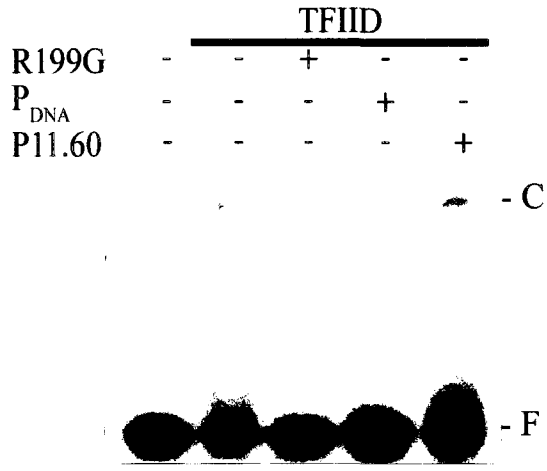
3.6 The TATA box-binding protein directly binds to R199G

The previous binding experiments confirmed the association of RNAP II with R199G and indicate that RNAP II might form a pre-initiation complex on R199G promoter similar to the one forming on a DNA promoter. Consequently RNAP II binding to R199G should be through TFIID. Purified TFIID (ProteinOne, Canada) was allowed to bind to radiolabeled R199G, and the samples were subjected to electrophoresis using non-denaturing PAGE. The result obtained confirmed direct binding of TFIID to R199G, as illustrated by the retardation of the migration of the complex when compared to free RNA (Figure 23A).

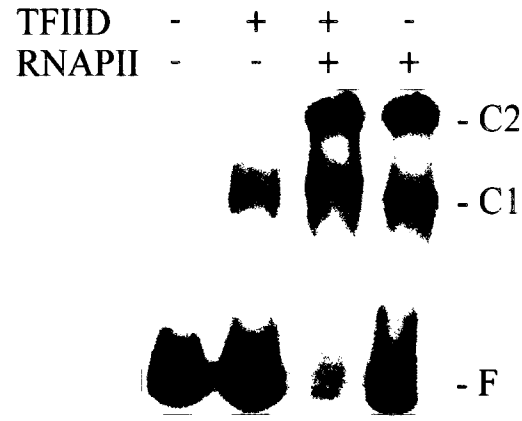
To test the specificity of this interaction, the binding of TFIID to R199G was challenged by various competitors as described earlier. Fifty fold molar excess of P11.60 was not able to compete for TFIID binding to R199G. However, either 50-fold molar excess of non-radiolabeled R199G or the control DNA promoter fragment greatly inhibited the formation of the R199G-TFIID complex (Figure 23A). More importantly, when the migration of the R199G-TFIID complex was compared to those formed between R199G and the RNAP II holoenzyme, it corresponded to the fast-migrating species (Figure 23B, C1), which suggests the formation of the R199G-TFIID prior to the assembly of the RNAP II complex (Figure 23B, C2). In addition, TFIID was able to bind to R38G, suggesting that TFIID was binding to R199G close to the initiation site, as observed above with RNAP II (Figure 23C).

Figure 23: EMSAs showing the specific binding of purified TFIID and RNAP II to HDV-derived RNAs. **A)** Binding of R199G in the presence of purified TFIID with 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G), a control DNA containing CMV promoter fragment (i.e.P_{DNA}), or P11.60. **B)** EMSA of R199G in the presence of purified TFIID and RNAP II. **C)** EMSA indicating the specific binding of R38G with purified TFIID in the presence of 50-fold molar excess of P11.60. In all EMSA figures, free RNA (F), RNA-protein complex one (C1) and RNA-protein complex two (C2) are shown.

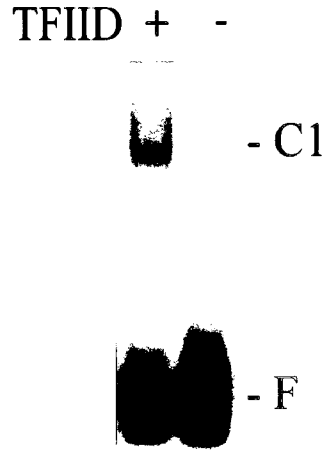
A)



B)



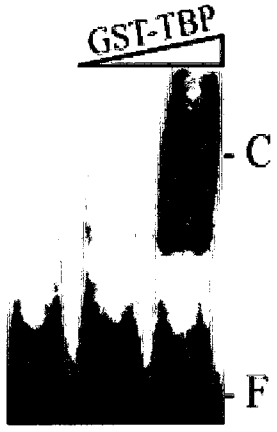
C)



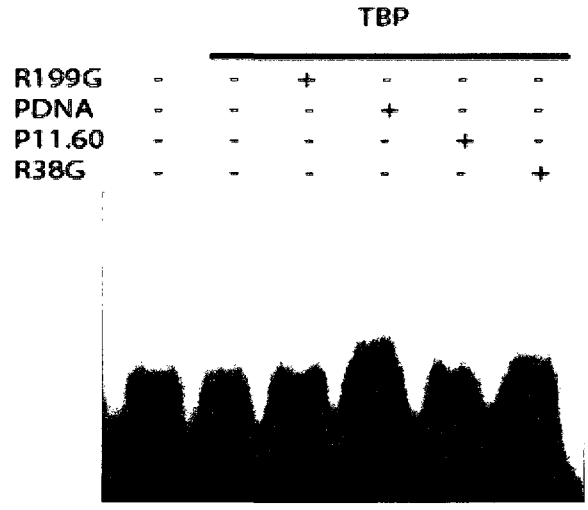
It is well established that TFIID interacts with DNA promoters through its TBP subunit; it is therefore possible that TBP serves a similar role with respect to R199G. To test this hypothesis, the formation of a complex between purified GST-tagged TBP and either R199G or R38G was analyzed by EMSA. A typical gel showing a titration of a constant amount of radiolabeled R199G with increasing amounts of purified GST-TBP is presented in Figure 24A. The observation that purified GST-TBP bound directly to R199G only when present at a high concentration (i.e. 64 μg/ml), suggests that it has a very low affinity for the substrate when it is not part of the TFIID complex. This low affinity was also reported between purified TBP and an adenovirus major late (AdML) promoter region [105]. The addition of non-radiolabeled R199G, R38G and the control DNA promoter fragment greatly inhibited the formation of the R199G-GST-TBP complex (Figure 24B). In correlation with the binding efficiency of both mutants to the RNAP II holoenzyme, the binding of GST–TBP to R38GPU and R38GSW were found to be reduced to 36% and 20%, respectively (Figure 24C). Taken together, these results provide direct evidence that TBP interacts directly with R199G and R38G, which contain the initiation site of mRNA transcription. Although I cannot exclude the possibility that other TFIID component(s) such as TAFII 150 and TAFII 250, which bind to the Inr element in DNA promoters might also bind to the Inr-like element in the RNA promoter, my results suggest that TFIID might be involved in the nucleation of the RNAP II pre-initiation on R199G, similar to what occurs on DNA promoters

Figure 24: EMSAs demonstrating the binding of purified GST-tagged TBP to HDV-derived RNAs. **A)** Titration of a constant amount of radiolabeled R199G with increasing amounts of purified GST-tagged TBP (0, 32, and 64 ng). **B)** R199G bound to purified GST-TBP in the presence of 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G), a control DNA promoter fragment (i.e. CMV promoter: P_{DNA}), P11.60 or R38G. **C)** Binding of R38G and two mutants with purified GST-TBP in the presence of 50-fold molar excess of P11.60.

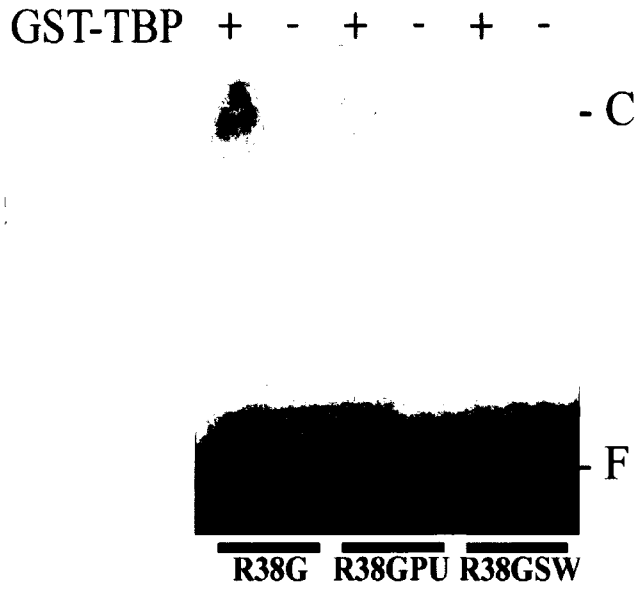
A)


GST-TBP
C
F

B)


TBP
R199G - - + - - -
PDNA - - - + - -
P11.60 - - - - + -
R38G - - - - - +

C)

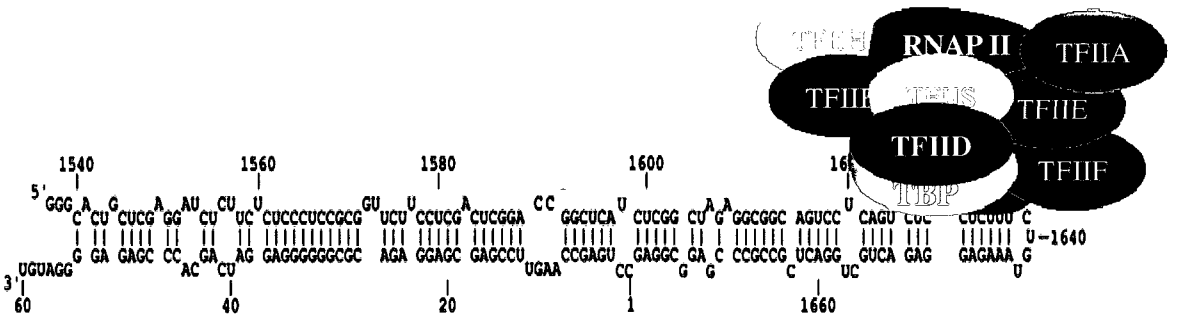

GST-TBP + - + - + -
C
F
R38G R38GPU R38GSW

CHAPTER FOUR

DISCUSSION

I investigated and confirmed the association of RNAP II and TBP with different HDV RNA templates. I then used an RNA template (R199G) and established that RNAP II binds directly to the RNA molecule and initiates transcription at the proposed mRNA start site. I have also confirmed that the transcription initiates at the proposed mRNA start site. The responsibility of RNAP II for the transcription from R199G was determined using RNAP II antibody specific against CTD of RNAP II. RNA affinity chromatography was then used to identify the components of RNAP II involved in the formation of pre-initiation complex on this unusual promoter, which led me to identify the TATA-binding protein (TBP) as an RNAP II subunit directly interacting with the RNA promoter. I also identified the general transcription factors associated with the RNA promoter (R199G), they include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and TFIIS. Formation of the transcription complex was shown to be similar to the complex formation on a DNA promoter. Overall, my results support a model in which an RNA promoter is recognized by RNAP II, in the same way a DNA promoter is recognized (Figure 25). Moreover, the mechanism of promoter recognition likely requires the involvement of the TBP-containing complex (TFIID) to nucleate the RNAP II complex on the HDV-derived RNA promoter. Further experiments are required to unveil the mechanism of RNAP II complex assembly on the RNA promoter.

Figure 25: Model showing the formation of an active RNAP II pre-initiation complex on an RNA promoter (R199G). The binding of the TBP, RNAP II and the rest of the transcription factors (TFIID, TFIIE, TFIIF, TFIIB, TFIIS and TFIIH) to the R199G promoter is illustrated.


RNAP II
TFIIA
TFIIE
TFIID
TFIIF
TBP
1540
1560
1580
1600
1640
1660
60
40
20
1

4.1 The role of DdRPs in RNA transcription from RNA templates

DdRPs catalyze the transcription of cellular RNA from DNA templates. In certain contexts, some of these DdRPs have been reported to use RNA as transcription templates. For example, T7 RNAP was shown to replicate a 64 nts RNA contaminant of unknown genetic origin in a commercial sample of the enzyme [106]. In addition, at high concentration, this enzyme was able to replicate a large variety of different RNA species when incubated for an extended period of time [107]. Moreover, *E. coli* RNAP was shown to amplify selected small, random RNA polymers [108, 109]. Recently, *E. coli* 6S RNA, which forms a stable complex with the σ^{70}-containing RNAP holoenzyme and inhibits σ^{70}-mediated transcription in the late stationary phase of *E. coli* [110], was found to serve as a template for the synthesis of short RNA products by *E. coli* RNAP [110]. In addition to these findings, DdRPs were also implicated in viroid replication. *E. coli* RNAP was reported to specifically initiate *in vitro* transcription from a stem-loop domain of the peach latent mosaic viroid RNA genome [100, 111].

Inspired by the previous studies that have shown DdRP to be involved in RNA-directed RNA polymerization of some viroids and many different RNAs [56, 31], and the similarity between viroids and HDV RNA secondary structure; HDV replication was comprehensively investigated [4, 47, 54, 56, 57, 51, 34, 112]. Most of the information obtained on HDV replication comes from early studies using cDNA transfection systems [33, 35, 28]. The use of such systems introduces artefacts into these studies since replication was initiated under the control of a foreign DNA promoter rather than RNA promoter, which forced the replication cycle to start with DNA-dependent RNA polymerase II. Another system that generated antigenomic RNA from genomic RNA was also used. This system

involved the continuous supply of both HDAg-S and genomic RNA. The continuous supply of HDAg-S and genomic RNA could overwhelm the transcription system creating an unnatural environment. Other groups used an inducible system to express HDAg-S under a tetracycline inducible promoter [57]. Even though this approach could reduce the amount of HDAg-S produced in comparison to the uncontrolled promoter, it provides a large supply of HDAg-S that might influence RNAP II activity. This activity will force the cell to carry out antigenomic transcription using any of the available RNAP, which could cast some doubt about the validity of this system in mimicking the real replication process. The use of an RNA transfection system partially solved this problem [54]. On the other hand, another artefact was introduced, which is the fact that transfection use more than unit length of RNA to start the replication; the extra length is required for the cleavage and circularization of HDV RNA. The higher number of HDV RNA copies in the cell might further force the transfected cells to activate other proteins which would involve different transcription machineries that are not part of the natural replication process.

4.1.1 Involvement of RNAP II in HDV RNA transcription

Even though these studies have answered a few questions regarding HDV RNA transcription, none of them have addressed the mechanism by which RNAP II recognizes HDV RNA or identified the RNAP II subunits involved in HDV RNA recognition.

In order for RNAP II to use RNA templates for transcription, it has to follow three steps similar to the steps taking place when acting on DNA promoters. The first step is the formation of the pre-initiation complex, the second is the initiation and elongation of transcription and the third is the termination of transcription. Due to the limitations in the

availability of an *in vivo* system to study these steps, I have used in vitro systems as follows. To accomplish the first step, the RNAP II complex has to make a close contact with the RNA template either directly or through one of its general transcription factors. To investigate this association, RNA-RNAP II containing complex was co-immunoprecipitated using an antibody against the CTD of RNAP II. The association of full-length genomic and antigenomic HDV RNAs with an RNAP II containing complex was shown to be specific. This finding provided a basis to investigate the association of the two HDV extremities with the RNAP II. The two HDV RNA molecules corresponding to the right terminal stem-loop domain of the genomic polarity (R199G) and the left terminal stem-loop domain of the antigenomic polarity (L213AG) were tested [34,51] for their association with RNAP II and were found to be specifically co-immunoprecipitated with RNAP II. To further investigate the possibility of the involvements of an important transcription factor, co-immunoprecipitation experiments using TBP or TFIID antibodies were used to unveil the association of TFIID with the HDV RNAs, particularly with both HDV RNA extremities. All the aforementioned HDV RNA templates were found to be associated with TBP or TFIID. Because R199G contains the proposed initiation site of HDAg mRNA and because this mRNA is post-transcriptionally processed with a 5'-cap and a 3'-poly (A) tail, R199G presented a perfect RNA template to further investigate the unconventional use of RNA promoters by RNAP II.

A HeLa nuclear extract transcription assay was used to examine the capability of R199G to act as an RNA promoter. An RNA transcript of approximately 80 nts was obtained when NTPs were included in the reaction. In order to examine the involvement of RNAP II in R199G transcription, HeLa NE proteins were pre-incubated with α-amanitin

prior to transcription. α-amanitin at concentrations lower than 30 μg/ml did not completely inhibit the synthesis of the 80 nts product from R199G. This would point to the involvement of a different RNAP or a different mechanism of RNAP II pre-initiation complex formation such as the association of TBP, TAF or other modulator proteins prior to RNAP II binding. Another possibility might be differences in cell lines response to α-amanitin. Moreover, the sensitivity of RNAP II, or any other host RNAP, to α-amanitin might be affected by the nature of the template (i.e. RNA versus DNA). Such a phenomenon was observed with tagetitoxin, a potent inhibitor of bacterial RNAP and nuclear DNA-dependent RNAP III [113]. Like α-amanitin, tagetitoxin can inhibit the elongation phase of RNA synthesis by interacting with the ternary complex that contains the RNAP, the template, and the nascent RNA. Even with a high concentration of tagetitoxin, only a partial inhibition of *E coli* RNAP transcription was observed when an RNA template was used, suggesting that the sensitivity to tagetitoxin depends on the nature of the template [113]. Another possibility is that the difference in an RNAP II response to α-amanitin is caused by the nucleotide composition of the template. The latter assumption is supported by the finding that wheat-germ RNAP II was able to synthesize RNA from a DNA template in the presence of 100 μg/ml of α-amanitin [114]. This high concentration is known to inhibit transcription by all three RNAPs. The resistance effect was due to the nucleotide composition of the DNA template. The DNA templates used were either single-stranded or double-stranded homopolymers of poly (dC) or poly (dC)-poly (dG). In contrast, the transcription from other DNA templates made of poly (dT), poly (dA)-poly (dT) or poly [d(A-T)] were sensitive to 1 μg/ml of α-amanitin [114]. This finding showed that the nucleotide content of the DNA

templates is crucial for α-amanitin inhibitory effect. Likewise, HDV RNA, which contains stretches of G-C (Figure 8), might mediate RNAP II resistance to α-amanitin. The difference in the RNAP II response to different RNA templates could be mistakenly interpreted as an involvement of RNAPs other than RNAP II in the transcription. This could account for the differences observed in the sensitivity of antigenomic RNA transcriptions to α-amanitin. Due to these complexities and differences in α-amanitin effect, these RNAP inhibitors (i.e. α-amanitin and tagetitoxin) must be used with caution when used as a means to investigate the involvement of RNAPs in RNA-dependent RNA transcription. Therefore using a specific monoclonal antibody to inhibit the transcription would be more conclusive.

4.1.1.1 RNAP II antibody inhibited the transcription from X-R199G in an *in vitro* Hela NE transcription system

Because it was required to purify the 80 nts faint RNA band for further investigation, such as RNase digestion, and due to the failure of RT-PCR to detect the RNA transcript in the presence of the R199G template, it was not possible to continue with the same transcription assay. To overcome this problem a small non HDV sequence was added to the 5'end of the template which was then used to reverse transcribe the transcription product. Using primer X, which hybridized only to the transcription product, my results showed that the transcription product is complementary to the template. The obtained HDV cDNA was then analyzed by PCR using primer X and either primer A or B. The amplification of the cDNA in the presence of the B primer confirmed that the transcription is specific and starts near the proposed transcription initiation site of HDAg mRNA (i.e. 1630, [33]). The

initiation site of transcription from X-R199G was determined by 5'RACE to be at position 1631. However, the technique used could not indicate if the RNA transcript is capped or not, which would have distinguished between the mRNA and antigenomic RNA transcription.

Pre-incubating RNAP II monoclonal antibody (CTD4H8) with HeLa NE prior to its use in transcription revealed the role of RNAP II for the transcription and strongly established the direct involvement of RNAP II in HDV transcription. Together, these data indicate that the right terminal stem-loop domain of the genomic polarity of HDV contains an RNA promoter, and provide strong evidence that this RNA promoter is recognized and used by RNAP II. Moreover, since the RNAP II antibody was targeted against the heptad repeats of the large subunit of RNAP II, my results indicate that the carboxy-terminal domain of RNAP II is needed for transcription from this RNA template. Its requirement is probably related to its role in complex assembly or CTD phosphorylation. My finding that the RNAP II antibody is able to inhibit transcription is supported by a study that showed a similar effect on DNA promoters. In that study pre-incubation of HeLa NE proteins with anti-RNAP II monoclonal antibody (3WG2) prior to transcription inhibited transcription from both the adenovirus-2 major late promoter, which contains a TATA box, and the murine dihydrofolate reductase (*dhfr*) promoter, which lacks the TATA box [91].

Although not tested and because it was not required in our transcription system, it is possible that the addition of HDAg-S might stimulate RNA synthesis from R199G. In support, a study using nuclear homogenates from an HDAg-S positive stable cell line (A3) and an HDAg-negative parental cell line (HepG2) revealed that the synthesis of complementary HDV RNA was detected in both extracts [50]. In some *ex vivo* (cell cultures; HepG2, COS7and Huh7) studies HDAg-S was proven to be required to cause HDV RNA

accumulation [5,50]. The differences in HDAg-S requirement between cell cultures and *in vitro* transcriptions might result from the malfunctioning (denaturation during extract preparation) of a repressor protein in the nuclear homogenates such as the negative elongation factor (NELF). HDAg-S is believed to be essential for HDV replication through the stimulation of the RNAP II transcription elongation machinery by displacing NELF and binding directly to RNAP II [52]. NELF has a negative regulatory effect on RNAP II transcription in mammalian cells [52]. This suggests that HDAg-S may enhance RNAP II elongation during both cellular mRNA synthesis and HDV RNA replication, which if it did so would interfere with the liver cell.

4.1.1.2 Direct binding of RNAP II and its general transcription factors to R199G

In humans, DNA dependent RNA transcription by RNAP II is carried out based on the availability of different sets of DNA promoter elements, which are mostly divided into two types. The first type is the basal elements that drive RNAP II to direct low levels of transcription. The second type contains the modulator promoter elements, which are recognized by modulator proteins and act to enhance or reduce the basal transcription levels [115]. RNAP II alone does not bind to its target promoters directly [116]. As a substitute for the RNAP II deficiency in binding, basal promoter elements are first bound by specific transcription factors and then recruit RNAP II and other required proteins such as general transcription factors and modulator proteins.

In DNA dependent transcription, the formation of a TBP/DNA promoter complex is the first step in the assembly of the RNAP II transcription initiation complex in mammalian cells [118]. Following this binding, RNAP II and other general transcription factors are

recruited to form a pre-initiation complex [119]. As the HDV RNA promoter has no known promoter elements similar to DNA promoter elements. RNAP II recognition of the HDV RNA might be mediated through its secondary structure which has been shown to have a large terminal stem loop with a bulge at the mRNA initiation site. This bulge is located within a thermodynamically unstable stem (i.e. low G-C pairing). Accordingly, the RNA promoter might mimic a pre-melted DNA promoter bubble which would give it a general affinity for RNAP II complex binding. A similar phenomenon was reported for DNA templates constructed with arbitrary non-promoter sequences and a small single stranded bulge [117]. In this case, core RNAP II affinity for DNA templates is thought to be mediated by electrostatic interactions with the phosphate backbone. These interactions could be preserved by an RNA template that has specific structural features.

Since the co-immunoprecipitation studies showed that the genomic HDV RNA segments are associated with RNAP II and TBP containing complexes, I tested the direct binding of R199G region with purified RNAP II, TFIID and TBP. By analyzing the formation of a complex between R199G and the RNAP II holoenzyme, it was observed that a fast-migrating species was formed at low concentrations of RNAP II, and a slower migrating species was formed at higher concentrations. When the migration of the R199G-TFIID complex was compared to that of complexes formed between R199G and the RNAP II holoenzyme, the migration of R199G-TFIID was found to correspond to the fast-migrating species. Accordingly, these results suggest that the formation of the R199G-TFIID complex is required to nucleate the RNAP II complex. As the purified GST-TBP binds directly and specifically to R199G, it is also tempting to suggest that RNA promoter recognition by RNAP II occurs primarily through TBP. This finding correlates with what is

known to occur on DNA promoters and supported my hypothesis of the existence of a similar mechanism for RNA and DNA promoter recognition.

Despite the clear association of TBP (or TFIID) with R199G, no consensus sequence similar to the TATA box was found in R199G. When acting on DNA promoters, the 38 KDa TBP recognizes and binds to the TATA-box through contact between its concave surface and the minor groove of DNA, which induces a 90 degree bending of the DNA [120, 121]. Proteins that bind the DNA double helix through the minor groove must operate mainly by indirect readout since the symmetric positioning of donor and acceptor groups in this groove makes it difficult to differentiate an A-T base pair from a T-A pair (and likewise for G-C and C-G base pairs [122]). It was proposed that the intrinsic structure of the DNA promoter region favors TBP binding [123]. Since TBP induces significant bending at the TATA box, sequences that are already bent, or are more flexible in a particular orientation, could be energetically favored.

TBP interaction with R199G could be mediated through its concave underside region, which is known to be essential for DNA binding. The fact that the CMV DNA promoter has a TATA box in its sequence and competed against the R199G binding site supports this hypothesis. It was also observed that TBP binding to the R199G promoter is less efficient than to the DNA promoter. However, TBP binding might be sufficient to cause the assembly of the rest of the general transcription factors on the R199G HDV RNA promoter to form the pre-initiation complex. This inefficiency of the binding of TBP to the R199G might account for the low level of RNA synthesis.

It is obvious that the initiation site is surrounded by an unusually high concentration of pyrimidine residues. In light of this data, I suggest that the observed purine/pyrimidine

polarity in this region, which is composed of mostly pyrimidines around the reported initiation site (Inr-like motif), is essential for transcription initiation. Moreover, several studies have confirmed that both Inr and DPE elements are bound by the transcription factor TFIID, where both motifs function together as a single core DNA promoter unit in the absence of a TATA box [94]. This transcription factor might, therefore, be involved in RNA promoter recognition by RNAP II during HDV transcription/replication. Since DNA Inr motifs are usually found on the coding strand, and the HDV Inr-like motif is located on the template strand, further investigations are needed to assess the importance of the Inr-like motif for RNA promoter recognition.

The R199G promoter characteristics were partially investigated using several mutations at the stem loop. The UUA bulge and the Inr-like motif were investigated using the R38GSW flip mutation and R38GPU in which the pyrimidine-rich region is mutated to purine. This disrupted the double stranded region generating a large bulge (Figure 19A). Binding experiments showed similar reductions in RNAP II and TBP affinity toward these two RNA mutations suggesting that they recognize RNA templates in a similar manner. This suggestion is also supported by the capability of R38G to compete against R199G for the TBP binding site (Figure 24B), indicating that TBP is exclusively binding to R38G. Further investigation is clearly required to determine the significance of the association of TBP (and probably TFIID) with RNA promoters and to establish the role of the Inr-like element in the formation of the pre-initiation complex, and their effects on transcription. Moreover, further experiments are required to investigate the capability of R38G to act as an RNA promoter. In this case, a transcription experiment using an RNA template containing the R38G region followed by non-HDV RNA sequences might be used.

4.1.1.3 Formation of an RNAP II pre-initiation and initiation complexes on X-R199G

The association of RNAP II and its general transcription factors with the R199G was investigated using RNA affinity chromatography. I demonstrated that RNAP II, along with its general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and TFIIS, bind to the RNA promoter and form a pre-initiation complex similar to those observed on canonical DNA promoters [94]. Moreover, this pre-initiation complex was active, since addition of free NTPs resulted in production of the transcript and the release of the large subunit of RNAP II, along with both the TFIIF and TFIIB subunits. Similar to what was reported for a DNA promoter, my results suggest that TFIIA, TFIID, and TFIIE might form a scaffold for re-initiation complex formation [94]. However, even though the level of RNAP II, TFIIF and TFIIB were reduced, they were not reduced to the extent reported for a DNA promoter [94]. This is consistent with my observation that *in vitro* transcription initiation is reduced on such an RNA template as compared to a DNA template (i.e. it is a very weak promoter). The inhibition of the transcription of R199G on the column by pre-incubating the HeLa nuclear extract with anti RNAP II antibody (Figure 21) further confirmed the role of RNAP II in the transcription.

Another protein that was identified by RNA affinity column chromatography was SC35, which is a splicing factor that was recently demonstrated to be involved in RNAP II transcription [70]. It was found to recruit P-TEFb, a factor that is required for RNAP II CTD phosphorylation at serine 2, to activate transcription by RNAP II [70]. Depletion of SC35 resulted in a remarkable decrease in RNA transcription by RNAP II but did not affect the transcription by RNAP I [70]. In the same study, SC35 was found to co-immunoprecipitate with RNAP II and the cyclin-dependent kinase 9 (CDK9), a component of transcriptional

elongation factor P-TEFb. The transactivation response (TAR) RNA stem loop of HIV is bound by Tat protein [124]. P-TEFb has also been shown to interact with Tat protein to restore its activity as a transcription elongation factor. Because the HDV RNA stem loop exhibited structural similarity with the TAR RNA stem loop that is recognized by Tat and stimulates HIV transcription elongation, a similar mechanism might exist for HDV transcription. In HDV, SC35 may act in a similar manner by recognizing the HDV RNA stem loop, which then recruits P-TEFb to activate transcription from HDV RNA by phosphorylation of RNAP II CTD. This could be utilized as a mechanism to compensate for the absence of HDAg-S in the *in vitro* transcription system. Even though it is highly speculative, a role for SC35 as a transcription elongation factor in HDV replication remains to be investigated.

4.2 Prospective

4.2.1 Thoughts

Even though my research answered some questions regarding HDV transcription and the involvement of RNAP II in its replication, other questions remain to be answered such as; can RNA cause infection humans similar to plant infection by viroids? Does it need the subgenomic mRNA to propagate? Are all the identified RNAP II general transcription factors required for RNA transcription? Are there any other proteins involved in the transcription process? Is there any RdRP activity other than RNAP II in the human cells that could replicate RNA? Speculative answers to these questions might be drawn from some proposed experiments and studies involving mouse models and the newly discovered RdRP activity.

In an early study, a mouse model using an integrated HDV cDNA was used to examine the HDV RNA replication in various organs [125]. The HDV RNAs were shown to accumulate in different organs including the liver, kidney and skeletal muscles. A ccumulation was 100 folds higher in skeletal muscle than in the liver, which indicated that HDV RNA replication is not restricted to the liver and that there might be other factors that allowed this high replication [125]. Another study used a mouse model artificially infected with genomic RNA and mRNA through the tail vein. After five days, these genomic RNAs were found to be replicating exclusively in the liver cells [126]. Since the human liver is the cleaning site in the body this might explain the restricted HDV RNA localization to the liver. These findings suggested that the original HDV RNA liver infection might be mediated by liver specific RNA receptor. Transfection of liver cells with HDV RNA in the absence of HBsAg and examining for the spread of the HDV RNA into the neighboring cells might answer the possibility that an RNA pathogen can cause infection with no protein requirements. It is also possible that HDV infection might have originated from an RNA pathogen that managed to infect human liver cells and then become coated with HBsAg. In today's infection, HDV use HBsAg to infect and propagate in the liver. During the course of infection, HDV is not able to cause systematic infection of other organs pointing to the specificity of the infection to the liver when coated with HBsAg. This could be related to the presence of a specific receptor for the envelope protein in the liver cell membrane.

4.2.2 Viroid analogy

Another suggestion is that this virus is not originally a human virus. It is a virus that has a liver specific coat obtained from its original host. After that, it adapted to its new life

by using RNAP II to replicate and becoming coated with HBsAg to propagate. This speculation is supported by the presence of some structural and functional similarities between viroids and HDV. Viroids are unencapsidated, highly base-paired, covalently closed circular, single-stranded RNA pathogens. They have very small genomes that range in length from approximately 200 to 450 nts. Although, they do not encode for any proteins, once inside the infected plant cells, viroids undergo robust RNA replication using cellular host enzymes [127, 128]. They are replicated either in the nucleus by RNAP II [129] or in the chloroplast by chloroplastic RNAP [130]. Purified RNAP II from healthy tomato and wheat germ tissues was able to synthesize full-length (-) PSTVd RNAs from (+) PSTVd viroid [131]. Citrus exocortis and chrysanthemum stunt viroids were less efficiently transcribed by the purified DdRP II [131].

HDV has an RNA genome three or four times longer than found in viroids, but all these RNAs adopt the rod-like structure. Both viroids and HDV replicate their RNA genomes by a rolling circle mechanism. Another similarity is the use of the *cis* acting ribozyme domains in the cleavage of the RNA strands by HDV and some viroids [29, 132]. The major between difference HDV and viroids is the presence of an open reading frame that encodes for HDAgs, which do not possess RNA dependent RNA polymerase activity [54].

New discoveries in Dr. Pelchat's laboratory have carried the HDV and viroid analogy to a higher level (personal communication, Dr. Pelchat). HDV RNA multimers were inoculated into the leaves of tomato plants, the inoculated HDV RNAs were found to replicate and spread in the plant. Cytopathic effects were also observed that include the inhibition of the growth of the tomato plant. In the same study, multimers of PSTVd RNAs

were transfected into animal cells and the accumulation of the new RNA was detected only in the presence of HDAg-S. Furthermore, when an engineered HDV RNA genome, which has been reduced to 342 nts was transfected into animal cells and also inoculated into tomato plant leaves, HDV RNAs were detected in both plant and animal cells. In the case of the infected animal cells the presence of HDAg-S was required to detect the HDV RNA accumulation. These findings supported my speculation that RNA pathogens other than HDV might be present in the human body without notice, especially in the presence of HBV infection similar to HDV. These RNA pathogens may require a similar protein to HDAg-S and acquire the capability to coat themselves with HBsAg or other similar proteins. RNA pathogens might be replicated either by RNAP II or previously unrecognized RdRP(s)

It is important to note that structural studies of RdRPs of positive-strand RNA viruses and dsRNA viruses showed structural similarities not only to each other, but also to DdRP-DNA polymerases and reverse transcriptases [133, 134]. These structures were proved to have a right hand palm with thumb and finger domains. The palm domain structure is particularly conserved and contains four sequence motifs preserved in all RNA and DNA polymerases [135]. As these features have been conserved throughout evolution, DdRP may have the potential to initiate transcription from both RNA and DNA templates. These structural similarities support a role for RNAP II in HDV replication and suggest the presence of RdRP activity in human cells.

A newly described RdRP activity was shown to use an RNA template for transcription. The human telomerase reverse transcriptase (TERT) catalytic subunit forms a complex with the RNA component of mitochondrial RNA processing endoribonuclease (RMRP) [136]. The RdRP activity utilizes a mitochondrial RNA as a template to generate

double-stranded RNA. These observations suggest the involvement of other unknown RdRP activity in human biology and raise the possibility of the existence of RNA promoters, which could lead to the discovery of RNA genomes and human RNA pathogens similar to viroids.

The presence of RNA pathogens in the human body could be investigated using systematic evolution of ligands by exponential enrichment SELEX with RNAP II. In this experiment, extracted RNAs from liver cells are passed through a column containing RNAP II, that is mounted to the column by RNAP II specific antibody.The bound RNAs are then eluted from the complex by competing with the CMV DNA promoter. The eluted RNAs are then amplified using degenerate primers that have the T7 promoter sequence added to it. The amplified cDNAs are then *in vitro* T7 transcribed and exposed to subsequent cycles of selection. The recovered RNAs are then cloned, sequenced and tested by EMSA for their specific binding to RNAP II and transfected into human cells to test for their replication capability, which might lead to the discovery of previously unrecognized RNA pathogens and/or RNAi silencing mechanisms.

The role of RNAP II subunits and general transcription factors in HDV RNA transcription could be tested using immunodepletion studies. All these transcription factors and RNA polymerase II subunits could be expressed and purified from bacterial culture. These proteins could be used to making the replication complexes containing different composition of these proteins or in immunodepletion studies. Each of these proteins can be immunodepleted from HeLa NE using a specific antibody. The immunodepleted HeLa NEs are then tested for their ability to support transcription from R199G. In cases where transcriptions levels were not supported by the immunodepleted HeLa NE, individual

purified proteins would be added back to test for restoration of the transcription capability. This approach would be sufficient to identify the basic transcription machinery involved in HDV RNA transcription.

4.2.3 Transcription of RNA templates by RNAP II and gene silencing analogy

In eukaryotes, RNAP II is the polymerase responsible for the transcription of all protein coding genes into pre-mRNAs, the transcribed pre-mRNAs are then processed by splicing into mature mRNA. These mature mRNAs are then transported to the cytoplasm, where their translation into proteins is carried out by ribosomes. Since RNAP II, in my findings, is capable of using RNA as a template, an involvement of this activity in gene silencing involving RNA interference (RNAi) might be expected. RNAi is a natural gene expression regulation process that causes the inhibition of gene expression in a sequence specific manner [137]. In some organisms such as plants and mammals the resulting RNA is used as a primer for RdRP polymerization to make the antisense strand, which serves as substrate for Dicer [138]. These RNAs are then cut into siRNAs, which in turn will either unwind and prime RdRP polymerization or mediate the cleavage of target mRNAs in a RISC dependent manner [138]. It is possible that RNAP II might act in RdRP polymerization to make antisense strands. In this case, RNAP II might generate double stranded RNAs. These RNAs might then be used by the RNAi RISC machinery to produce siRNA to block HDV RNA expression or interfere with cellular RNA translation. However, this RdRP activity has not yet been observed in human cells.

Recently, a study showed that the transcription of HDV RNA could generate 24 and 25 nts long RNA products. The transcription of these short RNAs start at the proposed

mRNA initiation site (1630 and 1631) and the other RNA was mapped to the other polarity of the same region (genomic hairpin RNA) [103]. This same study also found that *MOV10*, the human homolog of *Arabidopsis thaliana* RNA amplification factor *SDE3* and *Drosophila melanogaster* RISC-maturation factor gene *Armitage* (*armi*), interacts with HDAg proteins [103]. *MOV10* knockdown resulted in the inhibition of HDV RNA replication with no effect on HDAg mRNA translation. These findings suggest the involvement of *MOV10* in two different mechanisms, one involving HDV replication and another in gene silencing (HDAg mRNA). It is worth further investigation to explore the role of HDV RNA in RNA silencing, which if confirmed would further support the RNAP II RdRP activity.

REFERENCES

1. Diener, T.O. and W.B. Raymer, *Potato spindle tuber virus: a plant virus with properties of a free nucleic acid.* Science, 1967. **158**(799): p. 378-81.
2. Rizzetto, M., et al., *Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers.* Gut, 1977. **18**(12): p. 997-1003.
3. Lai, M.M., *The molecular biology of hepatitis delta virus.* Annu Rev Biochem, 1995. **64**: p. 259-86.
4. Macnaughton, T.B., et al., *Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases.* J Virol, 2002. **76**(8): p. 3920-7.
5. Kuo, M.Y., M. Chao, and J. Taylor, *Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen.* J Virol, 1989. **63**(5): p. 1945-50.
6. Li, Y.J., et al., *RNA-templated replication of hepatitis delta virus: genomic and antigenomic RNAs associate with different nuclear bodies.* J Virol, 2006. **80**(13): p. 6478-86.
7. Domingo, E., et al., *Emergence and selection of RNA virus variants: memory and extinction.* Virus Res, 2002. **82**(1-2): p. 39-44.
8. Chao, Y.C., et al., *Sequence conservation and divergence of hepatitis delta virus RNA.* Virology, 1990. **178**(2): p. 384-92.
9. Shakil, A.O., et al., *Geographic distribution and genetic variability of hepatitis delta virus genotype I.* Virology, 1997. **234**(1): p. 160-7.

10. Imazeki, F., M. Omata, and M. Ohto, *Complete nucleotide sequence of hepatitis delta virus RNA in Japan.* Nucleic Acids Res, 1991. **19**(19): p. 5439.

11. Ivaniushina, V., et al., *Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia.* J Gen Virol, 2001. **82**(Pt 11): p. 2709-18.

12. Lee, C.M., et al., *Characterization of a new genotype II hepatitis delta virus from Taiwan.* J Med Virol, 1996. **49**(2): p. 145-54.

13. Wu, J.C., T.Y. Chiang, and I.J. Sheen, *Characterization and phylogenetic analysis of a novel hepatitis D virus strain discovered by restriction fragment length polymorphism analysis.* J Gen Virol, 1998. **79 (Pt 5)**: p. 1105-13.

14. Radjef, N., et al., *Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades.* J Virol, 2004. **78**(5): p. 2537-44.

15. Domingo, E., et al., *[Quasispecies and molecular evolution of viruses].* Rev Sci Tech, 2000. **19**(1): p. 55-63.

16. Saracco, G., et al., *HDV infection from blood and blood products.* Prog Clin Biol Res, 1987. **234**: p. 361-5.

17. Rizzetto, M., R.H. Purcell, and J.L. Gerin, *Epidemiology of HBV-associated delta agent: geographical distribution of anti-delta and prevalence in polytransfused HBsAg carriers.* Lancet, 1980. **1**(8180): p. 1215-8.

18. Chai, N., et al., *Properties of subviral particles of hepatitis B virus.* J Virol, 2008. **82**(16): p. 7812-7.

19. Cupps, T.R., et al., *In vitro T cell immune responses to the preS2 antigen of the hepatitis B virus envelope protein in preS2 + S vaccine recipients. Absence of cross-*

reactivity of subtypes at a major T cell recognition site. J Immunol, 1993. **151**(6): p. 3353-60.

20. Ganem, D., *Virology. The X files--one step closer to closure.* Science, 2001. **294**(5550): p. 2299-300.

21. Sureau, C., B. Guerra, and R.E. Lanford, *Role of the large hepatitis B virus envelope protein in infectivity of the hepatitis delta virion.* J Virol, 1993. **67**(1): p. 366-72.

22. Farci, P., *Delta hepatitis: an update.* J Hepatol, 2003. **39 Suppl 1**: p. S212-9.

23. Rizzetto, M., *Hepatitis D: the comeback?* Liver Int, 2009. **29 Suppl 1**: p. 140-2.

24. Hadziyannis, S.J., *Review: hepatitis delta.* J Gastroenterol Hepatol, 1997. **12**(4): p. 289-98.

25. Wang, K.S., et al., *Structure, sequence and expression of the hepatitis delta (delta) viral genome.* Nature, 1986. **323**(6088): p. 508-14.

26. Kuo, M.Y., et al., *Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications.* J Virol, 1988. **62**(6): p. 1855-61.

27. Perrotta, A.T. and M.D. Been, *A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA.* Nature, 1991. **350**(6317): p. 434-6.

28. Hsieh, S.Y., et al., *Hepatitis delta virus genome replication: a polyadenylated mRNA for delta antigen.* J Virol, 1990. **64**(7): p. 3192-8.

29. Kuo, M.Y., et al., *Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus.* J Virol, 1988. **62**(12): p. 4439-44.

30. Reid, C.E. and D.W. Lazinski, *A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs.* Proc Natl Acad Sci U S A, 2000. **97**(1): p. 424-9.

31. Wang, C.C., et al., *Nucleic acid binding properties of the nucleic acid chaperone domain of hepatitis delta antigen.* Nucleic Acids Res, 2003. **31**(22): p. 6481-92.

32. Greco-Stewart, V.S., E. Schissel, and M. Pelchat, *The hepatitis delta virus RNA genome interacts with the human RNA polymerases I and III.* Virology, 2009. **386**(1): p. 12-5.

33. Gudima, S., et al., *Characterization of the 5' ends for polyadenylated RNAs synthesized during the replication of hepatitis delta virus.* J Virol, 1999. **73**(8): p. 6533-9.

34. Filipovska, J. and M.M. Konarska, *Specific HDV RNA-templated transcription by pol II in vitro.* RNA, 2000. **6**(1): p. 41-54.

35. Nie, X., J. Chang, and J.M. Taylor, *Alternative processing of hepatitis delta virus antigenomic RNA transcripts.* J Virol, 2004. **78**(9): p. 4517-24.

36. Casey, J.L., *RNA editing in hepatitis delta virus genotype III requires a branched double-hairpin RNA structure.* J Virol, 2002. **76**(15): p. 7385-97.

37. Linnstaedt, S.D., et al., *The fraction of RNA that folds into the correct branched secondary structure determines hepatitis delta virus type 3 RNA editing levels.* RNA, 2009. **15**(6): p. 1177-87.

38. Jayan, G.C. and J.L. Casey, *Increased RNA editing and inhibition of hepatitis delta virus replication by high-level expression of ADAR1 and ADAR2.* J Virol, 2002. **76**(8): p. 3819-27.

39. Jayan, G.C. and J.L. Casey, *Inhibition of hepatitis delta virus RNA editing by short inhibitory RNA-mediated knockdown of ADAR1 but not ADAR2 expression.* J Virol, 2002. **76**(23): p. 12399-404.

40. Sato, S., C. Cornillez-Ty, and D.W. Lazinski, *By inhibiting replication, the large hepatitis delta antigen can indirectly regulate amber/W editing and its own expression.* J Virol, 2004. **78**(15): p. 8120-34.

41. Chang, F.L., et al., *The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus.* Proc Natl Acad Sci U S A, 1991. **88**(19): p. 8490-4.

42. Tinland, B., et al., *The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7442-6.

43. Moede, T., et al., *Identification of a nuclear localization signal, RRMKWKK, in the homeodomain transcription factor PDX-1.* FEBS Lett, 1999. **461**(3): p. 229-34.

44. Alves, C., N. Freitas, and C. Cunha, *Characterization of the nuclear localization signal of the hepatitis delta virus antigen.* Virology, 2008. **370**(1): p. 12-21.

45. Lazinski, D.W. and J.M. Taylor, *Relating structure to function in the hepatitis delta virus antigen.* J Virol, 1993. **67**(5): p. 2672-80.

46. Lee, C.Z., et al., *RNA-binding activity of hepatitis delta antigen involves two arginine-rich motifs and is required for hepatitis delta virus RNA replication.* J Virol, 1993. **67**(4): p. 2221-7.

47. Chang, M.F., C.J. Chen, and S.C. Chang, *Mutational analysis of delta antigen: effect on assembly and replication of hepatitis delta virus.* J Virol, 1994. **68**(2): p. 646-53.

48. Chang, M.F., et al., *Functional motifs of delta antigen essential for RNA binding and replication of hepatitis delta virus.* J Virol, 1993. **67**(5): p. 2529-36.

49. Lee, C.Z., et al., *Isoprenylation of large hepatitis delta antigen is necessary but not sufficient for hepatitis delta virus assembly.* Virology, 1994. **199**(1): p. 169-75.

50. MacNaughton, T.B., et al., *Hepatitis delta antigen is necessary for access of hepatitis delta virus RNA to the cell transcriptional machinery but is not part of the transcriptional complex.* Virology, 1991. **184**(1): p. 387-90.

51. Beard, M.R., T.B. MacNaughton, and E.J. Gowans, *Identification and characterization of a hepatitis delta virus RNA transcriptional promoter.* J Virol, 1996. **70**(8): p. 4986-95.

52. Yamaguchi, Y., et al., *Stimulation of RNA polymerase II elongation by hepatitis delta antigen.* Science, 2001. **293**(5527): p. 124-7.

53. Bell, P., et al., *Hepatitis delta virus replication generates complexes of large hepatitis delta antigen and antigenomic RNA that affiliate with and alter nuclear domain 10.* J Virol, 2000. **74**(11): p. 5329-36.

54. Modahl, L.E., et al., *RNA-Dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases.* Mol Cell Biol, 2000. **20**(16): p. 6030-9.

55. Weinmann, R., H.J. Raskas, and R.G. Roeder, *Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection.* Proc Natl Acad Sci U S A, 1974. **71**(9): p. 3426-39.

56. Fu, T.B. and J. Taylor, *The RNAs of hepatitis delta virus are copied by RNA polymerase II in nuclear homogenates.* J Virol, 1993. **67**(12): p. 6965-72.

57. Chang, J., et al., *Transcription of hepatitis delta virus RNA by RNA polymerase II.* J Virol, 2008. **82**(3): p. 1118-27.

58. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

59. Kettenberger, H., K.J. Armache, and P. Cramer, *Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage.* Cell, 2003. **114**(3): p. 347-57.

60. Lehmann, E., F. Brueckner, and P. Cramer, *Molecular basis of RNA-dependent RNA polymerase II activity.* Nature, 2007. **450**(7168): p. 445-9.

61. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution.* Science, 2001. **292**(5523): p. 1863-76.

62. Corden, J., et al., *Promoter sequences of eukaryotic protein-coding genes.* Science, 1980. **209**(4463): p. 1406-14.

63. Allison, L.A., et al., *The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function.* Mol Cell Biol, 1988. **8**(1): p. 321-9.

64. Lu, H., et al., *The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10004-8.

65. Cadena, D.L. and M.E. Dahmus, *Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II.* J Biol Chem, 1987. **262**(26): p. 12468-74.

66. Liu, Y., et al., *Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex.* Mol Cell Biol, 2004. **24**(4): p. 1721-35.

67. Coppola, J.A., A.S. Field, and D.S. Luse, *Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped.* Proc Natl Acad Sci U S A, 1983. **80**(5): p. 1251-5.

68. Jove, R. and J.L. Manley, *In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites.* J Biol Chem, 1984. **259**(13): p. 8513-21.

69. Dantonel, J.C., et al., *Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA.* Nature, 1997. **389**(6649): p. 399-402.

70. Lin, S., et al., *The splicing factor SC35 has an active role in transcriptional elongation.* Nat Struct Mol Biol, 2008. **15**(8): p. 819-26.

71. Orphanides, G., T. Lagrange, and D. Reinberg, *The general transcription factors of RNA polymerase II.* Genes Dev, 1996. **10**(21): p. 2657-83.

72. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription.* Cell, 2002. **108**(4): p. 501-12.

73. Nakajima, N., M. Horikoshi, and R.G. Roeder, *Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID.* Mol Cell Biol, 1988. **8**(10): p. 4028-40.

74. Albright, S.R. and R. Tjian, *TAFs revisited: more data reveal new twists and confirm old ideas.* Gene, 2000. **242**(1-2): p. 1-13.

75. Singer, V.L., C.R. Wobbe, and K. Struhl, *A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation.* Genes Dev, 1990. **4**(4): p. 636-45.

76. Suzuki, Y., et al., *Identification and characterization of the potential promoter regions of 1031 kinds of human genes.* Genome Res, 2001. **11**(5): p. 677-84.

77. Lo, K. and S.T. Smale, *Generality of a functional initiator consensus sequence.* Gene, 1996. **182**(1-2): p. 13-22.

78. Martinez, E., et al., *TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter.* EMBO J, 1994. **13**(13): p. 3115-26.

79. Kaufmann, J. and S.T. Smale, *Direct recognition of initiator elements by a component of the transcription factor IID complex.* Genes Dev, 1994. **8**(7): p. 821-9.

80. Burke, T.W. and J.T. Kadonaga, *Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters.* Genes Dev, 1996. **10**(6): p. 711-24.

81. Tora, L., *A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription.* Genes Dev, 2002. **16**(6): p. 673-5.

82. DeJong, J., R. Bernstein, and R.G. Roeder, *Human general transcription factor TFIIA: characterization of a cDNA encoding the small subunit and requirement for basal and activated transcription.* Proc Natl Acad Sci U S A, 1995. **92**(8): p. 3313-7.

83. DeJong, J. and R.G. Roeder, *A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA.* Genes Dev, 1993. **7**(11): p. 2220-34.

84. Tan, S., et al., *Crystal structure of a yeast TFIIA/TBP/DNA complex.* Nature, 1996. **381**(6578): p. 127-51.

85. Lagrange, T., et al., *New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB.* Genes Dev, 1998. **12**(1): p. 34-44.

86. Nikolov, D.B., et al., *Crystal structure of a TFIIB-TBP-TATA-element ternary complex.* Nature, 1995. **377**(6545): p. 119-28.

87. Tsai, F.T. and P.B. Sigler, *Structural basis of preinitiation complex assembly on human pol II promoters.* EMBO J, 2000. **19**(1): p. 25-36.

88. Ha, I., et al., *Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II.* Genes Dev, 1993. **7**(6): p. 1021-32.

89. Peterson, M.G., et al., *Structure and functional properties of human general transcription factor IIE.* Nature, 1991. **354**(6352): p. 369-73.

90. Rossignol, M., I. Kolb-Cheynel, and J.M. Egly, *Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH.* EMBO J, 1997. **16**(7): p. 1628-37.

91. Drapkin, R., et al., *Human cyclin-dependent kinase-activating kinase exists in three distinct complexes.* Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6488-93.

92. Shiekhattar, R., et al., *Cdk-activating kinase complex is a component of human transcription factor TFIIH.* Nature, 1995. **374**(6519): p. 283-7.

93. Eichner, J., et al., *Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex.* EMBO J, 2009.

94. Yudkovsky, N., J.A. Ranish, and S. Hahn, *A transcription reinitiation intermediate that is stabilized by activator.* Nature, 2000. **408**(6809): p. 225-9.

95. Gudima, S., et al., *Origin of hepatitis delta virus mRNA.* J Virol, 2000. **74**(16): p. 7204-10.

96. Greco-Stewart, V.S., et al., *The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome.* Virology, 2007. **357**(1): p. 68-78.

97. Sims, R.J., 3rd, S.S. Mandal, and D. Reinberg, *Recent highlights of RNA-polymerase-II-mediated transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 263-71.

98. Thompson, N.E., et al., *Inhibition of in vivo and in vitro transcription by monoclonal antibodies prepared against wheat germ RNA polymerase II that react with the heptapeptide repeat of eukaryotic RNA polymerase II.* J Biol Chem, 1989. **264**(19): p. 11511-20.

99. Zhao, X. and W. Herr, *A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition.* Cell, 2002. **108**(5): p. 615-27.

100. Pelchat, M. and J.P. Perreault, *Binding site of Escherichia coli RNA polymerase to an RNA promoter.* Biochem Biophys Res Commun, 2004. **319**(2): p. 636-42.

101. Brodsky, A.S., et al., *Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells.* Genome Biol, 2005. **6**(8): p. R64.

102. Reid, J. and J.Q. Svejstrup, *An assay for studying ubiquitylation of RNA polymerase II and other proteins in crude yeast extracts.* Methods Enzymol, 2006. **408**: p. 264-73.

103. Haussecker, D., et al., *Capped small RNAs and MOV10 in human hepatitis delta virus replication.* Nat Struct Mol Biol, 2008. **15**(7): p. 714-21.

104. Chang, J., et al., *Development of a novel system to study hepatitis delta virus genome replication.* J Virol, 2005. **79**(13): p. 8182-8.

105. Zhao, X. and W. Herr, *Role of the inhibitory DNA-binding surface of human TATA-binding protein in recruitment of human TFIIB family members.* Mol Cell Biol, 2003. **23**(22): p. 8152-60.

106. Konarska, M.M. and P.A. Sharp, *Replication of RNA by the DNA-dependent RNA polymerase of phage T7.* Cell, 1989. **57**(3): p. 423-31.

107. Biebricher, C.K. and R. Luce, *Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase.* EMBO J, 1996. **15**(13): p. 3458-65.

108. Biebricher, C.K. and L.E. Orgel, *An RNA that multiplies indefinitely with DNA-dependent RNA polymerase: selection from a random copolymer.* Proc Natl Acad Sci U S A, 1973. **70**(3): p. 934-8.

109. Wettich, A. and C.K. Biebricher, *RNA species that replicate with DNA-dependent RNA polymerase from Escherichia coli.* Biochemistry, 2001. **40**(11): p. 3308-15.

110. Barrick, J.E., et al., *6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter.* RNA, 2005. **11**(5): p. 774-84.

111. Pelchat, M., C. Grenier, and J.P. Perreault, *Characterization of a viroid-derived RNA promoter for the DNA-dependent RNA polymerase from Escherichia coli.* Biochemistry, 2002. **41**(20): p. 6561-71.

112. Macnaughton, T.B. and M.M. Lai, *Genomic but not antigenomic hepatitis delta virus RNA is preferentially exported from the nucleus immediately after synthesis and processing.* J Virol, 2002. **76**(8): p. 3928-35.

113. Pelchat, M., F. Cote, and J.P. Perreault, *Study of the polymerization step of the rolling circle replication of peach latent mosaic viroid.* Arch Virol, 2001. **146**(9): p. 1753-63.

114. Job, C., et al., *A DNA-dependent RNA synthesis by wheat-germ RNA polymerase II insensitive to the fungal toxin alpha-amanitin.* Biochem J, 1992. **285 (Pt 1)**: p. 85-90.

115. Hernandez, C. and R. Flores, *Plus and minus RNAs of peach latent mosaic viroid self-cleave in vitro via hammerhead structures.* Proc Natl Acad Sci U S A, 1992. **89**(9): p. 3711-5.

116. Nikolov, D.B. and S.K. Burley, *RNA polymerase II transcription initiation: a structural view.* Proc Natl Acad Sci U S A, 1997. **94**(1): p. 15-22.

117. Helmann, J.D. and P.L. deHaseth, *Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners.* Biochemistry, 1999. **38**(19): p. 5959-67.

118. Kim, J.L., D.B. Nikolov, and S.K. Burley, *Co-crystal structure of TBP recognizing the minor groove of a TATA element.* Nature, 1993. **365**(6446): p. 520-7.

119. Verrijzer, C.P. and R. Tjian, *TAFs mediate transcriptional activation and promoter selectivity.* Trends Biochem Sci, 1996. **21**(9): p. 338-42.

120. Nikolov, D.B., et al., *Crystal structure of TFIID TATA-box binding protein.* Nature, 1992. **360**(6399): p. 40-6.

121. Seeman, N.C., J.M. Rosenberg, and A. Rich, *Sequence-specific recognition of double helical nucleic acids by proteins.* Proc Natl Acad Sci U S A, 1976. **73**(3): p. 804-8.

122. Khrapunov, S. and M. Brenowitz, *Comparison of the effect of water release on the interaction of the Saccharomyces cerevisiae TATA binding protein (TBP) with "TATA*

Box" sequences composed of adenosine or inosine. Biophys J, 2004. **86**(1 Pt 1): p. 371-83.

123. Smale, S.T. and D. Baltimore, *The "initiator" as a transcription control element.* Cell, 1989. **57**(1): p. 103-13.

124. Harrich, D., C.W. Hooker, and E. Parry, *The human immunodeficiency virus type 1 TAR RNA upper stem-loop plays distinct roles in reverse transcription and RNA packaging.* J Virol, 2000. **74**(12): p. 5639-46.

125. Polo, J.M., et al., *Transgenic mice support replication of hepatitis delta virus RNA in multiple tissues, particularly in skeletal muscle.* J Virol, 1995. **69**(8): p. 4880-7.

126. Chang, J., et al., *Replication of the human hepatitis delta virus genome Is initiated in mouse hepatocytes following intravenous injection of naked DNA or RNA sequences.* J Virol, 2001. **75**(7): p. 3469-73.

127. Flores, R., *A naked plant-specific RNA ten-fold smaller than the smallest known viral RNA: the viroid.* C R Acad Sci III, 2001. **324**(10): p. 943-52.

128. Flores, R., et al., *Viroids and viroid-host interactions.* Annu Rev Phytopathol, 2005. **43**: p. 117-39.

129. Kolonko, N., et al., *Transcription of potato spindle tuber viroid by RNA polymerase II starts in the left terminal loop.* Virology, 2006. **347**(2): p. 392-404.

130. Navarro, J.A., A. Vera, and R. Flores, *A chloroplastic RNA polymerase resistant to tagetitoxin is involved in replication of avocado sunblotch viroid.* Virology, 2000. **268**(1): p. 218-25.

131. Rackwitz, H.R., W. Rohde, and H.L. Sanger, *DNA-dependent RNA polymerase II of plant origin transcribes viroid RNA into full-length copies.* Nature, 1981. **291**(5813): p. 297-301.

132. Hutchins, C.J., et al., *Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid.* Nucleic Acids Res, 1986. **14**(9): p. 3627-40.

133. O'Reilly, E.K. and C.C. Kao, *Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure.* Virology, 1998. **252**(2): p. 287-303.

134. Poch, O., et al., *Identification of four conserved motifs among the RNA-dependent polymerase encoding elements.* EMBO J, 1989. **8**(12): p. 3867-74.

135. Xiong, Y. and T.H. Eickbush, *Origin and evolution of retroelements based upon their reverse transcriptase sequences.* EMBO J, 1990. **9**(10): p. 3353-62.

136. Maida, Y., et al., *An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA.* Nature, 2009. **461**(7261): p. 230-5.

137. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.

138. Martens, H., et al., *RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase.* Mol Biol Cell, 2002. **13**(2): p. 445-53.

Appendix I: Oligonucleotides used in the PCR reactions

R199G sense	5'- GGAATTCTAATACGACTCACTATAGGGACTGCTCGAG GATCTCTTCTCTCCC-3'
R199G antisense	5'- GGAATTCACATCCCCTCTCGGGTGC-3'
L233AG sense	5'GGAATTCTAATACGACTCACTATAGGGTCGGCATGGCAT CTCCACCTCCTCGCG-3'
L233AG antisense	5'- GCCGGCATGGTCCCAGCCTCCTCGCTGGCG -3'
X-R199G sense	5'- GGAATTCTAATACGACTCACTATAGGGGGGAACAAG GGACTGCTCGATCTCTT-3'
(X-199GΔUUA) sense	5'-CTGCTCGAGGATCTCTTCTCTCCTCCGCGGTTCTTCCTC GACTCGGACCGGCTCATCTCGGCTAGTGGCGGCAGTCCTC AGTACTCCTCTTTTCTGT-3'
(X-199GΔUUA) antisense	5'- GGAATTCACATCCCCTCTCGGGTAC-3'
CMV sense	5'CCAAAATCAACGGGACTTTCCAAAATGTCGTAACAAGT GGGGCCCATTGA-3'
CMV antisense	(5'-TATATAGTCCTCCCACCGTACACGCCTACCGCCCATTT GCCTAAATGGGCCCCAC-3')
R38G sense	5'-GGAATTCTAATACGACTCACTATAGGGCAGTACTCTTA CUCTTTTCUGTAAAGAGG-3'
R38G antisense	5'-CAGTCTCCTCTTTACAGAAAAGAGTAAGAGGTACTG-3'
R38GPU sense	5'-GGGAATTCTAATACGACTCACTATAGGGCAGTAGAGTT AGAGAAATCTGTAAAGAGG-3'
R38GPU antisense	5'-CAGTCTCCTCTTTACAGAATTTCTCATTCTCGTACTG-3'
R38GSW sense	5'-GGGAATTCTAATACGACTCACTATAGGGCAGTAGAGGA GAAATCTGTTTTCTCATTCTC-3'
R38GSW antisense	5'CAGTGAGGAGAAAACAGATTTCTCCTCTACTGCCC3'
P11.60 sense	5'-GGAATTCTAATACGACTCACTATAGGGTCTCTGAAAT GAGACGAACTCA-3'
P11.60 antisense	(5'-GGGCCGCTGAACTTTTGATGATATGAGTTTCGTGTCAT TTCAGAGA-'3
Nested gene-specific primers	(5'-GAATTCTAATACGACTCACTATAGGGACTGCTCGATCT CTT-3')
AAP	5'-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3'
AUAP	5'-GGCCACGCGTCGACTAGTAC-3'

Under lined nucleotide sequence indicates T7 promoter

Appendix II: Greco-Stewert VS, Miron P, **Abrahem A**, Pelchat M. The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. Virology. 5; 357(1):68-78. (2007)

Available online at www.sciencedirect.com
ScienceDirect
VIROLOGY
ELSEVIER
Virology 357 (2007) 68–78
www.elsevier.com/locate/yviro

The human RNA polymerase II interacts with the terminal stem–loop regions of the hepatitis *delta* virus RNA genome

Valerie S. Greco-Stewart, Paul Miron, Abrahem Abrahem, Martin Pelchat *

Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5

Received 11 May 2006; returned to author for revision 2 June 2006; accepted 7 August 2006
Available online 7 September 2006

Abstract

The hepatitis *delta* virus (HDV) is an RNA virus that depends on DNA-dependent RNA polymerase (RNAP) for its transcription and replication. While it is generally accepted that RNAP II is involved in HDV replication, its interaction with HDV RNA requires confirmation. A monoclonal antibody specific to the carboxy terminal domain of the largest subunit of RNAP II was used to establish the association of RNAP II with both polarities of HDV RNA in HeLa cells. Co-immunoprecipitations using HeLa nuclear extract revealed that RNAP II interacts with HDV-derived RNAs at sites located within the terminal stem–loop domains of both polarities of HDV RNA. Analysis of these regions revealed a strong selection to maintain a rod-like conformation and demonstrated several conserved features. These results provide the first direct evidence of an association between human RNAP II and HDV RNA and suggest two transcription start sites on both polarities of HDV RNA.

Keywords: Hepatitis *delta* virus; RNA virus; RNA promoter; RNA polymerase II; Co-immunoprecipitation

Introduction

The hepatitis *delta* virus (HDV) is a subviral satellite virus of the hepatitis B virus (HBV) and is composed of a single-stranded, circular RNA molecule of approximately 1700 nucleotides (Chen et al., 1986). Its genome forms an unbranched, rod-shaped structure with a high proportion of canonical base pairing, contains two complementary ribozyme motifs (i.e. *delta* ribozymes), and has a single open reading frame producing two viral proteins (HDAgs; Fig. 1; Taylor, 2006). These two proteins are identical in sequence, except that the large HDAg (HDAg-L) contains 19 additional amino acids at its C-terminus resulting from RNA editing of the termination codon of the small HDAg (HDAg-S) gene (Casey et al., 1992; Wang et al., 1992; Weiner et al., 1988; Xia et al., 1990). Each protein has a distinct function: HDAg-S (195 amino acids) is essential for HDV accumulation, while the HDAg-L (214 amino acids) is required for virion assembly and is reported to be a dominant negative inhibitor of replication (Chang et al., 1991; Kuo et al., 1989; Ryu et al., 1992).

Although HDV requires HBV for encapsidation and transmission, it relies entirely on its host for replication. HDV replicates through a symmetrical rolling circle mechanism that involves only RNA intermediates (Chen et al., 1986; Kuo et al., 1988a,b; Lai, 2005; Taylor, 2006). Replication of the genomic circular monomer produces linear, multimeric strands which are subsequently self-cleaved and ligated, resulting in antigenomic polarity circular monomers. This process is repeated to generate the genomic RNA, which is found at a greater intracellular abundance than the antigenomic species (Chen et al., 1986). Because HDV does not encode its own replicase, host DNA-dependent RNA polymerases (RNAPs) are considered to be involved in the replication and transcription of HDV RNA (Lai, 2005).

Many studies using cultured cells and cell extracts have suggested that RNAP II might be responsible for HDV replication based on the sensitivity of the accumulation of HDV mRNA and processed unit-length genomic HDV RNAs to low levels of α-amanitin, a mycotoxin that inhibits DNA-dependent RNA synthesis by RNAP II (de Mercoyrol et al., 1989; Filipovska and Konarska, 2000; Fu and Taylor, 1993; Macnaughton et al., 1991). This hypothesis was substantiated by experiments using cells containing an α-amanitin-resistant

* Corresponding author. Fax: +1 613 562 5452.
E-mail address: mpelchat@uottawa.ca (M. Pelchat).

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.08.010

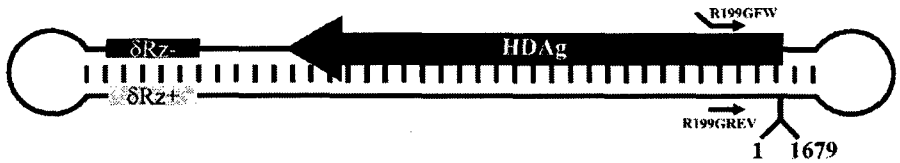


Fig. 1. Schematic representation of the HDV RNA genome. The structure illustrates the superimposed features of both the genomic and antigenomic HDV polarities. The genome is depicted in its circular, unbranched rod-like conformation. The location of the genomic and antigenomic ribozymes and the open reading frame for HDAg are indicated. The position of the primers used to generate the R199G RNA molecule is indicated. The numbering is in accordance with Kuo et al. (1988a).

allele of the largest subunit of RNAP II, which partially relieved transcription inhibition by α-aminitin (Modahl et al., 2000). RNAP II is furthermore speculated to be involved in the transcription of the HDAg mRNA because, *in vivo*, this mRNA was shown to be post-transcriptionally processed with a 5′-cap and a 3′-poly(A) tail, which are typical features of transcripts generated by RNAP II (Gudima et al., 2000; Nie et al., 2004). Additionally, in *in vitro* transcription assays using nuclear extract (NE) from HeLa cells and RNA derived from the left terminal stem–loop domain of HDV antigenomic RNA, synthesis of the complementary strand was possible (Filipovska and Konarska, 2000). Accumulation of this RNA product was highly sensitive to α-amanitin. This sensitivity was partially abrogated in experiments conducted in NE from cells containing an α-amanitin-resistant allele of the largest subunit of RNAP II, suggesting the involvement of RNAP II in this reaction. Interestingly, the transcription did not proceed by de novo initiation, but rather by cleavage of the RNA template followed by extension of the new 3′-end, generating a chimeric template/transcript product.

Conversely, the accumulation of the antigenomic RNA species has been shown to be resistant to a higher dose of α-amanitin, suggesting the involvement of another yet unknown RNAP in the life cycle of HDV (Macnaughton et al., 2002; Modahl et al., 2000). It has been hypothesized that RNAP I or an RNAP I-like polymerase might be involved in HDV replication because antigenomic HDV RNA was shown to accumulate in the nucleolus and its synthesis is associated with the RNAP I-specific transcription factor SL1 (Li et al., 2006).

Because general host DNA-directed RNAP II transcription is inhibited at relatively low doses of α-amanitin (de Mercoyrol et al., 1989), the observed inhibition of HDV RNA accumulation in the presence of this transcription inhibitor might be an indirect rather than a direct effect. It is difficult to differentiate between specific inhibition of HDV replication and non-specific toxic effects on host cells or the need for some additional host DNA-directed RNAP II transcripts. In addition, because the enzymes involved in HDV transcription/replication are host DNA-dependent RNAPs coerced to use an RNA template, sensitivity of RNAP II, or any other host RNAP(s), to α-amanitin might be affected by the nature of the template (i.e. RNA versus DNA). Such a phenomenon was reported for actinomycin D (Meienhofer and Atherton, 1973; Pelchat and Perreault, 2004) and was recently observed with tagetitoxin (Pelchat et al., 2001), a potent inhibitor of bacterial RNAP and nuclear DNA-dependent RNAP III (Mathews and Durbin, 1994). Like α-amanitin, tagetitoxin is believed to inhibit the elongation phase of RNA synthesis by interacting with the ternary complex that contains the RNAP, the template, and the nascent RNA (Mathews and Durbin, 1994; Gong et al., 2004). RNAP inhibitors must thus be used with caution when utilized as a means of identifying an RNAP involved in RNA-dependent transcription.

Although HDV replication has been shown to be sensitive to low concentrations of α-amanitin, suggesting RNAP II activity, no direct interaction between RNAP II and HDV RNA has been reported. The present study was undertaken (i) to confirm the association of RNAP II with HDV RNA, and (ii) to identify the regions on the HDV RNA genome responsible for this interaction. To this end, we took advantage of an antibody raised against the carboxy terminal domain (CTD) of RNAP II to clarify the interaction of RNAP II with HDV RNA by directly examining their association *in vivo* within HeLa cells. We identified the regions involved in this interaction by co-immunoprecipitating various HDV-derived RNAs with RNAP II using HeLa nuclear extract (NE) and the anti-RNAP II antibody. Then, we used covariation analysis on natural variants to analyze the conserved nucleotides and structural motifs of the interaction sites of RNAP II on HDV RNA. Our analyses support a role for the human RNAP II in the life cycle of HDV by demonstrating its association with HDV RNA *in vivo* and with the terminal stem–loop domains of both polarities of HDV RNA *in vitro*.

Results

RNAP II interacts with HDV RNA within HeLa cells

To establish the direct interaction between RNAP II and HDV RNA *in vivo*, we immunoprecipitated RNAP II from HeLa cells containing both polarities of HDV RNA and tested for the presence of viral RNA in the immunoprecipitated samples. To this end, we used the monoclonal antibody (mAb) 8WG16 which was raised against wheat germ RNAP II and found to recognize epitopes conserved among the CTDs of RNAP IIs derived from a variety of different eukaryotes (Thompson et al., 1989). When this commonly used antibody is tested against HeLa NE proteins, only one band corresponding to the largest subunit of RNAP II is revealed by Western blot. To obtain cells harboring both polarities of HDV RNA, HeLa cells were transiently transfected with a dimeric HDV transcript of genomic polarity and a eukaryotic expression vector overexpressing HDAg-S, which has been found to be essential for HDV RNA accumulation (Kuo et al., 1989).

Four days post-transfection, the cells were harvested and the RNA–protein complexes cross-linked with formaldehyde *in vivo* using the ribonucleoprotein immunoprecipitation (RIP) assay developed by Niranjanakumari et al. (2002). Following immunoprecipitation and high-stringency washes to ensure the specificity of the binding, the cross-links were reversed by heating the samples. Recovered nucleic acids were then extracted and subjected to RT-PCR. For this step, we used primers for a cDNA fragment corresponding to what we refer to as the right-terminal stem–loop domain of HDV (Fig. 1, R199GFW and R199GREV). In addition to the HDV sequence (nucleotides 1541 to 61), the 232 bp cDNA fragment generated after RT-PCR included a promoter for the T7 RNAP and recognition sites for restriction enzymes. When we used the 8WG16 mAb to immunoprecipitate the sample, cDNA fragments corresponding to both polarities of R199 were detected (Fig. 2, 8WG16 lanes). Cloning and subsequent sequence analysis confirmed the identity of the HDV-derived cDNA fragments. These cDNA fragments were detected neither in the control reactions lacking antibody nor in the goat anti-mouse IgG mAb control reactions (Fig. 2), indicating that cDNAs obtained from this assay were specific to the samples immunoprecipitated with the 8WG16 mAb. Moreover, these fragments were also absent from samples that were not subjected to the cross-linking reaction prior to co-immunoprecipitation (Fig. 2, No-XL 8WG16 lanes). To ascertain that the association with the RNAP II complex is specific for HDV RNA, RT-PCR was also performed on an abundant cellular RNA (β-actin mRNA). As shown in Fig. 2B, although the β-actin cDNA fragment was easily amplified from crude extract from HeLa cells, this fragment was not detected in the various immunoprecipitation reactions performed above. Overall, the presence of HDV RNA fragments in the cross-linked, immunoprecipitated samples provides evidence that formaldehyde covalently cross-links the largest subunit of RNAP II or a complex containing RNAP II to HDV RNA in HeLa cells. To corroborate this finding, we used a pair of primers corresponding to the other terminal stem–loop region of the HDV RNA genome (i.e. L213, nucleotides 687 to 900); identical results were obtained (data not shown).

The primers used in the RT reactions were designed to differentiate between both polarities of HDV RNA. Specifically, primers corresponding to sequences of the genomic (i.e. R199GFW) and antigenomic (i.e. R199GREV) strand of HDV RNA were only able to reverse transcribe from either the antigenomic or the genomic polarity of HDV RNA, respectively (Fig. 2, HDVG and HDVAG RNA lanes). Since both polarities of HDV RNA were recovered when the cross-linked sample was immunoprecipitated with the 8WG16 mAb, we conclude that both polarities of HDV RNA associate with RNAP II within HeLa cells, which directly establishes the association of the human RNAP II with HDV RNA and strongly suggests its involvement in the synthesis of both strands of HDV RNA.

Although the RT-PCR reactions were not quantitative, the intensity of the bands corresponding to both polarities of HDV RNA co-immunoprecipitated using the 8WG16 mAb was similar. Because it has been established that genomic HDV RNA accumulates more than antigenomic HDV RNA in infected cells (Chen et al., 1986), our results suggest a more efficient binding of RNAP II to the antigenomic polarity of HDV RNA. To confirm the preferential interaction of RNAP II with this polarity, semi-quantitative RT-PCR was performed on RNA samples taken either before or after co-immunoprecipitation with the 8WG16 mAb. Because no human RNA is expected to co-immunoprecipitate with RNAP II, it was not possible to normalize the different samples using an internal RNA control. To circumvent this difficulty, amplification curves were performed and used to determine the number of

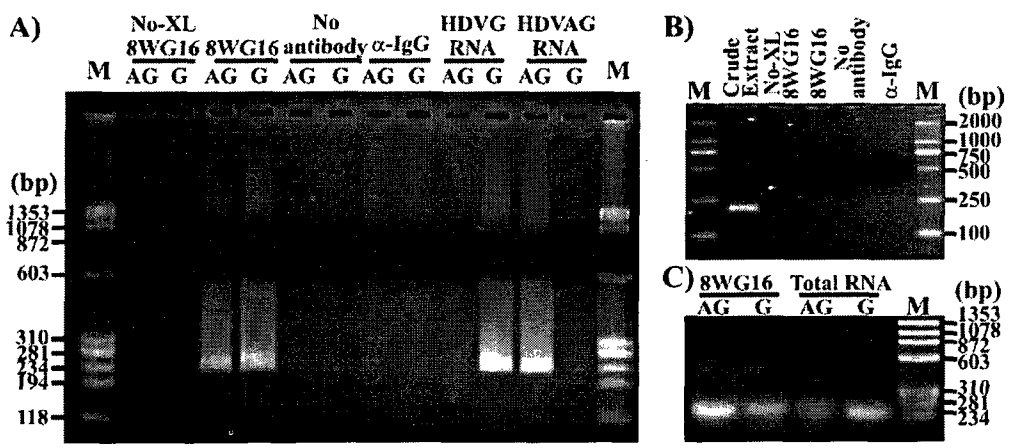


Fig. 2. Association of RNAP II with HDV RNA within HeLa cells. HeLa cells transfected with HDV RNA were used for the preparation of cross-linked lysates for immunoprecipitation using the 8WG16 mAb, no antibody, or the control goat anti-mouse IgG mAb. Non-cross-linked lysate immunoprecipitated using the 8WG16 mAb is also presented. (A) Association of RNA with both polarities of HDV RNA. HDVG and HDVAG RNA lanes are RT-PCR product controls using *in vitro* transcribed HDV RNA of genomic and antigenomic polarity, respectively. AG and G lanes contain RT reactions performed using primers complementary to either antigenomic or genomic HDV RNA corresponding to the R199 region. (B) Control RT-PCR using the β-actin cDNA to demonstrate the specificity of the co-immunoprecipitations for HDV RNAs. The co-immunoprecipitated samples are as described above. (C) Semi-quantitative RT-PCR amplifications of both polarities of HDV RNA before and after co-immunoprecipitation with the 8WG16 mAb. Lanes M contain DNA ladders: panels A and C, PhiX174 digested with *Hae*III; panels B, GelPilot Mid Range DNA ladder (Qiagen).

PCR cycles at which all the samples were within their linear phase of amplification. Then, the intensities of the cDNA bands corresponding to both polarities of HDV RNA were compared within each sample. Prior to co-immunoprecipitation, we found a genomic:antigenomic HDV RNA ratio of 1.99:1. Although less than reported using infected liver cells, this value is in conformity with a greater intracellular abundance of genomic HDV RNA when compared to antigenomic HDV RNA (Chen et al., 1986). By performing a similar calculation for the sample co-immunoprecipitated with the 8WG16 mAb, we found that the genomic:antigenomic HDV RNA ratio shifted to 1:1.56. As a result, under our conditions, antigenomic HDV RNA was enriched 3.1-fold after co-immunoprecipitation with the 8WG16 mAb. This enrichment indicates a preferential interaction of RNAP II with antigenomic HDV RNA.

RNAP II interacts with the terminal stem–loop domains of both polarities of HDV RNA in HeLa NE

The previous experiments provide evidence of the association of RNAP II with both the genomic and antigenomic HDV RNAs within HeLa cells. To determine the regions of the HDV RNA genome responsible for this interaction, we investigated the association of RNAP II with various HDV-derived RNAs *in vitro* using HeLa NE. As an initial step, both polarities of the HDV RNA genome were arbitrarily divided in three domains: 213 nt RNAs corresponding to what we refer to as the left terminal stem–loop (L213; nt 687 to 900), RNAs corresponding to the double-stranded central region (CENTR; nt 681 to 54 and nt 900 to 1540), and 199 nt RNAs corresponding to the right terminal stem–loop (R199; nt 1540 to 60). To guarantee that both RNA molecules forming the CENTR RNAs annealed, both strands were mixed, heated, slowly cooled, and tested for double-strand formation. Electrophoresis on a native acrylamide gel confirmed that most of the RNA molecules forming the CENTR RNAs annealed under our conditions (data not shown). The various radiolabeled HDV-derived RNAs were allowed to form RNA–protein complexes with HeLa NE proteins. The 8WG16 mAb was then added, the samples were immunoprecipitated using protein-G agarose beads, and the co-immunoprecipitated complexes were then subjected to denaturing gel electophoresis.

A typical gel using a radiolabeled RNA corresponding to the right terminal stem–loop domain of the genomic polarity of HDV RNA (i.e. R199G) is shown in Fig. 3A. Under our conditions, a band corresponding to the radiolabeled RNA was detected when the sample was immunoprecipitated using the 8WG16 mAb in the presence of HeLa NE. This band was not detected in the absence of NE or when the 8WG16 mAb was replaced with the anti-mouse IgG mAb, indicating that the signal observed was not the result of non-specific binding of R199G to the protein-G agarose beads or to the 8WG16 mAb. To ensure the specificity of the interaction, various RNA competitors were tested. When an excess of unlabeled homologous competitor was used, the intensity of the band corresponding to the radiolabeled R199G was considerably diminished. When poly-A RNA was used as an unrelated RNA competitor, no significant decrease in the co-immunoprecipitation of the radiolabeled R199G was observed. When acting on a DNA promoter, RNAP II binds to double-stranded nucleic acid. To verify that the observed association between RNAP II and R199G was not caused by non-specific affinity for double-stranded RNA, a small RNA hairpin (Fig. 3C, P11.60) derived from the peach latent mosaic viroid (PLMVd) that acts as an RNA promoter for the *Escherichia coli* RNAP (Pelchat et al., 2002) was used as a competitor. The interaction between RNAP II and R199G was not significantly decreased by addition of an excess amount of P11.60. Taken together, these observations demonstrate that RNAP II interacts specifically with R199G within HeLa NE.

The interaction of RNAP II with the various regions of both genomic and antigenomic HDV RNA iS summarized in Fig. 3B. Because several RNA molecules were tested, only the lanes corresponding to the amount of RNA used and co-immunoprecipitation reactions in the presence of the non-specific RNA competitor P11.60 with either the absence (−NE) or the presence (+NE) of HeLa NE are presented. Under our conditions, both polarities of the complete HDV RNA genome co-immunoprecipitated specifically using the 8WG16 mAb (Fig. 3B, HDVG and HDVAG). When the band intensities of the co-immunoprecipitated RNAs were normalized using the amount of radiolabeled RNA used, the antigenomic HDV RNA (HDVAG) was found to be co-immunoprecipitated 1.9-fold more efficiently than the genomic HDV RNA (HDVG). These results are in agreement with our previous observations using cross-linked lysates and indicate that RNAP II interacts more efficiently with the antigenomic polarity of HDV RNA.

Next, the various HDV-derived RNAs corresponding to the three arbitrary domains of both polarities were tested to determine the regions on the HDV RNA genome responsible for RNAP II interaction. Radioactive bands corresponding to the stem–loop domains of both polarities of HDV RNA (i.e. R199G, R199AG, L213G, and L213AG) were detected, whereas bands corresponding to the central regions of both polarities of HDV RNA (i.e. CENTRG and CENTRAG) were not. These results indicate that RNAP II interacting sites are located predominantly between nt 1540 to 60 and nt 687 to 900 on both polarities of the HDV RNA genome.

To further refine the regions interacting with RNAP II, smaller HDV-derived transcripts were synthesized and were used in the co-immunoprecipitation reactions. Specifically, R38 and L48, which correspond to 38 and 48 nt RNAs located at the tip of the right and the left hairpin domains, respectively, were tested for RNAP II interaction. Using the same co-immunoprecipitation conditions as above, we found that both polarities of R38 and L48 were able to co-immunoprecipitate using the 8WG16 mAb (Fig. 3C), indicating that the terminal stem–loop domains of both polarities of HDV RNA can be reduced to smaller hairpins without significantly affecting their interaction with RNAP II. As an additional negative control, P11.60 was tested directly in the co-immunoprecipitation reactions. The small PLMVd-derived RNA hairpin did not co-immunoprecipitate with RNAP II when using the 8WG16 mAb (Fig. 3C), which further supports the specificity of the co-immunoprecipitation experiments. Because RNAP II complexes are

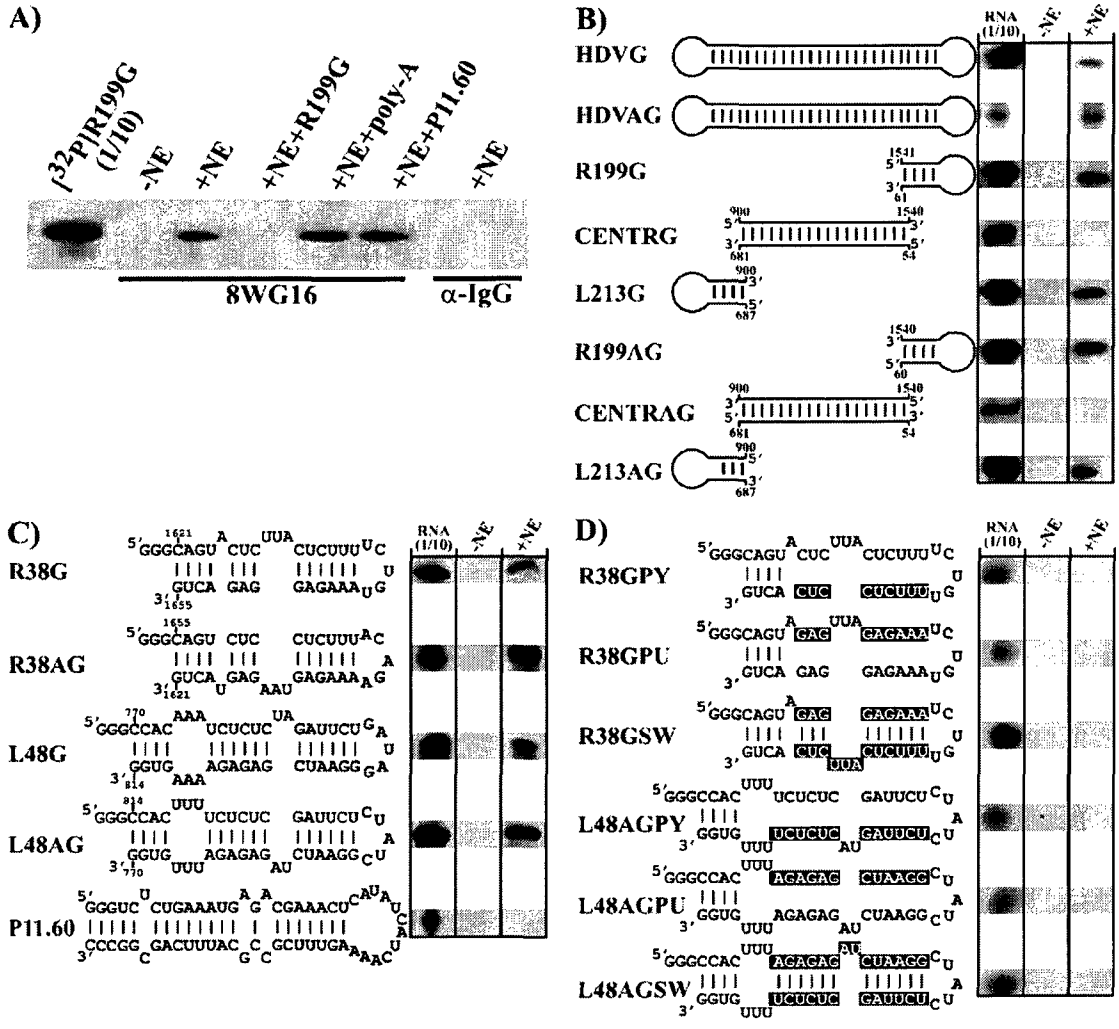


Fig. 3. Interaction of RNAP II with HDV-derived RNAs. (A) Co-immunoprecipitation of radiolabeled R199G with RNAP II in HeLa NE. 20-fold excess non-radioactive R199G, poly-A RNA, or P11.60 (Pelchat et al., 2002) was used in competition experiments. (B) Interaction of RNAP II with various HDV-derived RNAs. (C) Interaction of RNAP II with the extremities of the rod-like HDV secondary structure of both polarities. (D) Interaction of RNAP II with mutants affecting the secondary structure and the purine/pyrimidine polarization. The numbering is in accordance with Kuo et al. (1988a). One tenth of the RNAs used for each co-immunoprecipitation are used as markers. The co-immunoprecipitations presented in panels B – D were performed in the presence of a 20-fold excess P11.60, as a non-specific RNA competitor.

composed of many subunits and other related factors (Cramer, 2002), we cannot conclude that the binding of the RNAP II core enzyme to the various RNA molecules is direct. Also, we cannot exclude the possibility that other RNAPs might be able to recognize these RNA fragments. However, these results provide direct evidence that a complex containing the CTD of the largest subunit of RNAP II interacts specifically with the terminal stem–loop domains of both polarities of HDV RNA in HeLa NE.

Sequence and secondary structure analysis of the terminal stem–loop domains of both polarities of HDV RNA

To interact with HDV RNA, RNAP II must recognize specific features (i.e. sequences and/or secondary structures) on both polarities of the terminal stem–loop domains. We have taken advantage of the large number of known nucleotide sequences of *in vivo* selected variants to identify the features of terminal stem–loop domains of HDV RNA by studying positions of covariation and nucleotide conservation.

We collected the sequences corresponding to R38 and L48 from a total of 81 variants representing the various HDV genotypes indexed in the Subviral RNA Database (Rocheleau and Pelchat, 2006). First, we aligned all the RNA sequences using the Clustal software (Thompson et al., 1994), which generated a tree showing the sequence differences between HDV genotypes. Due to the imperfect alignment of the analyzed regions among the different HDV sequences, we applied alignments to smaller data sets according to HDV genotypes because of their lower degree of heterogeneity. For each of these genotypes, the consensus features of the RNA secondary structures generated by analysis of base-pair covariation were determined. Finally, we combined alignments by insertion of gaps so that homologous motifs (i.e. secondary structures) were

superimposed. Using the consensus sequence derived from the alignment, the most thermodynamically stable secondary structure was predicted and re-adjusted based on the base-pair covariation data. Sequence conservation and the resulting consensus RNA secondary structures for both polarities of R38 and L48 are summarized in Fig. 4 together with the fully conserved nucleotides (red nucleotides) and the location of the base-pairs predicted by the covariation analysis (blue circles).

Despite heterogeneity among the different sequences and genotypes, base-pair covariation analysis indicates a strong selective pressure to maintain the rod-like conformation of these regions in which most of the base-pairs can be predicted. A few nucleotides within this region are conserved across all the HDV sequences examined (Fig. 4, red nucleotides). Considering the degree of nucleotide divergence in the adjacent nucleotides, the observed conservation suggests a functional role for the conserved nucleotides. Strikingly, a short double-stranded stem (CUC/GAG) was found to be exceptionally conserved (Fig. 4B, yellow box). In fact, most of the nucleotides comprising this short CUC/GAG stem are completely conserved in all the analyzed variants and the other positions exhibit only minor variations. Notably, we observed a strong polarization in the purine/pyrimidine content near the tip of the rod up to the end of the CUC/GAG stem in all the sequences analyzed. A region of 12 nt upstream of the terminal loop of R38G and R38AG exhibits an average of 91.0±1.0% pyrimidine content which is matched by a region on the opposite strand containing almost exclusively purines (Fig. 4B). In the case of L48G and L48AG, the sequence upstream of the loop contains 86.7±5.9% pyrimidines which is matched by a region on the opposite strand having 82.0±5.8% purines (Fig. 4B).

The secondary structure and the polarization in the purine/pyrimidine content near the tip of the rod are required for RNAP II interaction

Sequence analysis of natural HDV variants suggests that the maintenance of the rod-like conformation and the polarization in the purine/pyrimidine content near the tip of the rod of both polarities of HDV RNA might be required for RNAP II interaction. We therefore introduced substitutions to R38G and L48AG aimed at destabilizing either of these conserved features. First, the involvement of the purine-rich segments located downstream of the terminal loops was investigated. R38GPY and L48AGPY are derivative RNAs in which the purine-rich segments were mutated to the pyrimidine-rich sequences located on their respective upper strands. Upon co-immunoprecipitation using the 8WG16 mAb, no significant association of RNAP II with either R38GPY or L48AGPY was detected (Fig. 3D). Similar results were obtained when R38GPU and L48AGPU, two derivative RNAs in which the pyrimidine-rich segments located upstream of the terminal loops were mutated to purine-rich sequences,

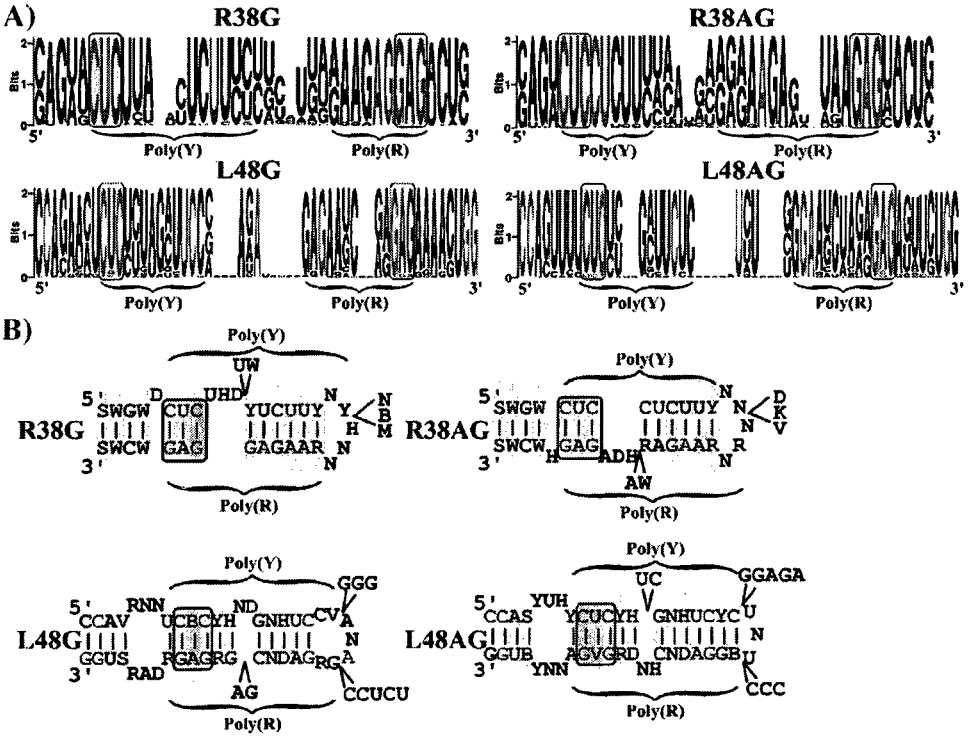


Fig. 4. Primary sequence conservation and proposed secondary structure of R38G/AG and L48G/AG. The nucleotide motif (A) and the consensus secondary structure (B) derived from the analysis of the alignment of 81 HDV natural variants are presented. Completely conserved nucleotides (red), observed base-pair covariation (blue ovals), and the conserved CUC/GAG motifs (yellow) are indicated. IUPAC 1-letter code abbreviations are used for the identity of the nucleotides on the secondary structures showing less than 95% conservation.

were used. Specifically, no co-immunoprecipitation of either R3G8PU or L48AGPU with RNAP II was detected (Fig. 3D).

Because the mutations found in these four mutants also result in the disruption of their predicted rod-like structures, the absence of RNAP II interaction with these RNAs could be the result of the loss of their rod-like conformation. R38GSW and L48AGSW, which contain compensatory mutations that restore the predicted secondary structure of these hairpins, were therefore synthesized and tested for association with RNAP II. No complex formation was detected when using either of these "flip" mutants (Fig. 3D). Taken together, these results are in accordance with the sequence analysis of natural HDV variants and indicate that the association of RNAP II with these regions is dependent on the conformation and on the polarization in the purine/pyrimidine content near the tip of the rod.

Discussion

The mechanism of RNA synthesis is highly conserved and can be roughly divided into three distinct phases: initiation, elongation, and termination. For initiation to occur, RNAPs search for and bind to promoter sequences to form a "closed complex"; thus, association of RNAP II would indicate sites of transcription (Smale and Kadonaga, 2003). Here, we have used an antibody raised against the CTD of RNAP II to clarify RNAP II interaction with HDV RNA by studying the first step of the transcription process (i.e. promoter recognition). This approach was used to directly demonstrate specific RNAP II interaction with HDV RNA, both *in vivo* within HeLa cells and *in vitro* using HeLa NE. Using various HDV-derived RNAs with HeLa NE proteins, we have shown that specific RNAP II interaction occurs *in vitro* predominantly via small domains located at the extremities of both polarities of the rod-like secondary structure of the HDV RNA genome, and that RNAP II might bind more efficiently to the antigenomic polarity of HDV RNA.

It is possible that the observed interaction of RNAP II with these regions might not lead to transcription initiation. When binding to a DNA promoter, the RNAP II found within an initiation site is hypophosphorylated at its C-terminal domain (CTD) and is distinct from the elongating or terminating forms of the enzyme that are hyperphosphorylated (Hahn, 2004; Sims et al., 2004). In this study, we have used the 8WG16 mAb, which specifically recognizes the unphosphorylated RNAP II CTD (Patturajan et al., 1998); the extremities of both polarities of HDV RNA are thus likely to be sites of pre-initiation complex formation (i.e. active promoters; Thompson et al., 1989; Smale and Kadonaga, 2003). Moreover, most of these interaction sites are in agreement with previous reports.

The 5′-end of HDAg mRNA corresponds to an initiation site located within R38G (Gudima et al., 2000). It is thus likely that the right terminal stem–loop domain contains an RNA promoter recognized by RNAP II. RNA fragments similar to both R199G and L213AG were reported to be able to initiate polymerization *in vitro* using HDV-derived RNAs and HeLa NE (Beard et al., 1996; Filipovska and Konarska, 2000). Interestingly, transcription from the RNA template similar to L213AG did not proceed by de novo initiation, but rather by cleavage of the RNA template, followed by extension of the new 3′-end, generating a chimeric template/transcript product (Filipovska and Konarska, 2000). An RNA fragment similar to L213G was also able to generate a specific product in this system, despite the authors' claim that the signal detected did not correspond to their typical chimeric template/transcript product (Filipovska and Konarska, 2000). Here, we have shown that the right terminal stem–loop region of antigenomic HDV RNA (i.e. R199AG) can also interact with RNAP II and we have demonstrated that these four RNA regions can be reduced significantly in length without affecting RNAP II interaction. In summary, our results suggest the involvement of RNAP II in the synthesis of both genomic and antigenomic HDV RNAs. In addition, because RNAP II interacts with the extremities of both polarities of the HDV RNA genome, our data suggest the existence of two transcription/replication start sites for both genomic and antigenomic HDV RNAs. Moreover, it is possible that other RNAP(s) are involved in HDV transcription/replication, given that only the 8WG16 mAb was tested.

The RNA fragments interacting with RNAP II were further investigated by taking advantage of a large number of natural HDV variants. Analysis of base-pair covariation indicated that although changes in the primary sequence of these segments might be tolerated, their respective secondary structures are most likely critical for HDV replication/accumulation. The conserved features of the two extremities of the rod-shaped HDV RNA are very similar. Our analysis revealed a highly ordered RNA secondary structure that is comprised of regions of double-stranded RNA with a conserved CUC/GAG triple base-pair motif and one strand of pyrimidines upstream of a 3–8 nt terminal loop followed by a complementary strand of purines. Because these features are found near all the terminal loops of HDV RNA which were shown to interact with RNAP II, it is tempting to suggest the involvement of these features in RNAP II binding. Accordingly, mutagenesis of R38G and L48AG, affecting the conformation and the polarization in the purine/pyrimidine content near the tip of the rod, indicated that the interaction of RNAP II with these regions is dependent on these features.

Previous studies have reported extensive mutagenesis on the extremities of the rod-like structure of HDV RNA and have observed their effects on HDV RNA accumulation *in vivo* (Beard et al., 1996; Gudima et al., 1999, 2006; Wang et al., 1997; Wu et al., 1997). Fig. 5 presents a summary of these published results. For the right terminal extremities, variation of the terminal loop sequence and size had no effect on HDV RNA accumulation (Gudima et al., 1999; Wu et al., 1997). In contrast, disruption of the overall rod-like conformation of this region or even small changes in the terminal stem had major effects on accumulation of HDV RNA (Beard et al., 1996; Gudima et al., 1999; Wu et al., 1997). In addition, HDV RNA accumulation was found to be sensitive to the size of the 3 nt external bulge, which includes the putative mRNA initiation site (Gudima et al., 1999). Our analysis indicates that this bulge can be on either side of the stem–loop structures and RNAP II interaction still occurs (compare R38R/R38AG). However, correct positioning of this

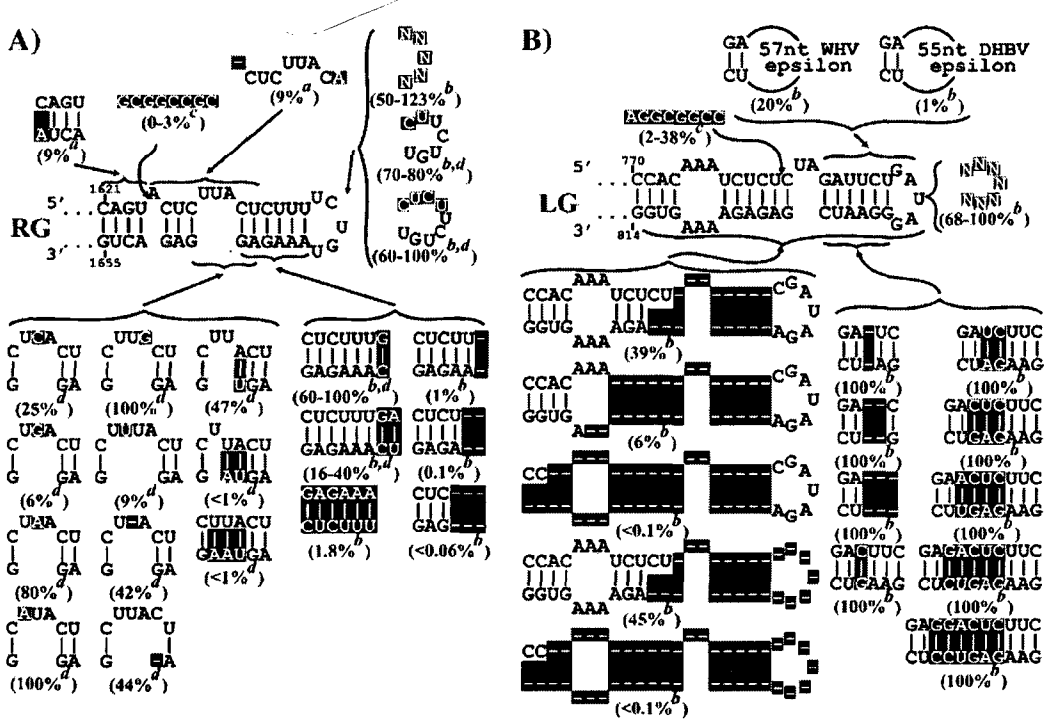


Fig. 5. Summary of published mutagenesis results on the extremities of the rod-like structure of HDV RNA. The mutations and the resulting consequences on HDV RNA accumulation were taken from (a) Beard et al., 1996, (b) Wu et al., 1997, (c) Wang et al., 1997, and (d) Gudima et al., 1999. Reported mutagenesis on either the right- or the left-terminal extremities of the rod-like secondary structure of the HDV RNA are presented in panels A or B, respectively.

non-paired region might be critical for the initiation stage of HDAg mRNA synthesis. Noteworthy, a "flip" mutant in the terminal stem was found to greatly affect HDV RNA accumulation (Wu et al., 1997) in complete agreement with our results. The association of RNAP II with this region and initiation of transcription are thus dependent on the conformation and on the polarization in the purine/pyrimidine content near the tip of the rod.

Conversely, the left-terminal region was found to be highly tolerant to nucleotide changes within the loop and to insertions as well as deletions in the adjacent stem region (Wu et al., 1997). Only the removal of large segments or the insertion of various non-HDV sequences in the conserved CUC/GAG motifs resulted in loss of accumulation of HDV RNA (Gudima et al., 2006; Wang et al., 1997). Interestingly, most genomes containing large inserts lost their inserted sequences suggesting the involvement of this region in HDV RNA replication (Gudima et al., 2006). The divergence of these reports with our results might be explained by the likely involvement of another RNAP, such as RNAP I. Moreover, because RNAP II was found to be associated with the stem–loop region of both polarities of HDV RNA, suggesting at least two replication start sites for each polarity of HDV RNA, complementation of replication impaired mutants of the left-terminal stem–loop structure might occur. Since the right extremity contains sequences required for HDAg mRNA transcription, such complementation might not occur, causing the right terminal stem–loop region to be more sensitive to mutation.

We cannot exclude the possibility that the conserved features of the HDV extremities might also be linked to activities unrelated to RNAP II binding, such as RNA stabilization, RNA transportation, or RNA encapsidation. For example, these polypyrimidine/polypurine tracts could serve as binding sites for other RNA-binding proteins. Typical representatives of these types of proteins are the polypyrimidine tract-binding (PTB) proteins and the heterogeneous nuclear ribonucleoproteins (hnRNPs), which, in uninfected cells, are involved in a variety of processes, such as pre-mRNA splicing in the nucleus (Garcia-Blanco et al., 1989; Swanson and Dreyfuss, 1988). However, we recently reported the finding that the polypyrimidine tract-binding protein-associated splicing factor (PSF) is also able to bind specifically to both polarities of HDV through the terminal stem–loop domains of the rod-like structure (Greco-Stewart et al., in press). Because PSF is known to bind to the CTD of RNAP II (Emili et al., 2002) and PSF binding sites correlate with the locations on the HDV RNA genome found to be associated with RNAP II, it is tempting to suggest that by interacting with both an HDV RNA promoter and RNAP II, PSF or another polypyrimidine/polypurine tract-binding protein might provide docking sites or assist RNAP(s) in the formation of pre-initiation complexes. The co-immunoprecipitation assay we used in this study will be a perfect tool in determining the content of the RNAP II complex interacting with HDV RNA and the contribution of additional *trans*-acting host factors in RNA promoter recognition by RNAP II. For instance, although HDAg-S was demonstrated to be essential for HDV accumulation (Kuo et al., 1989), our results indicate

that this protein is not required for the interaction of RNAP II with HDV RNAs in HeLa NE since it was absent from all co-immunoprecipitation assays. This observation correlates with previous results suggesting that HDAg-S could regulate the elongation phase of transcription by binding to RNAP II directly and by displacing the negative elongation factor (Yamaguchi et al., 2001).

Interestingly, the features and the locations of RNAP II interaction sites on HDV RNA share structural similarities with DNA-dependent RNAP binding and/or initiation sites reported for other related subviral RNAs. *E. coli* RNAP can bind to and transcribe from a stem–loop domain derived from the PLMVd genome (Pelchat et al., 2002; Pelchat and Perreault, 2004). Any structural alterations that disrupt the overall rod-like conformation of the RNA template were shown to interfere with *de novo* transcription by *E. coli* RNAP (Pelchat et al., 2002; Pelchat and Perreault, 2004). Replication of the avocado sunblotch viroid genome is proposed to initiate within large terminal loops (Navarro and Flores, 2000). Wheat germ RNAP II was shown to bind to the large terminal loops of potato spindle tuber viroid genome (Goodman et al., 1984), and it was recently reported that potato NE was able to initiate transcription of this viroid within a terminal stem–loop region of the RNA genome (Kolonko et al., 2006).

It is now known that the observed structural similarities between bacterial RNAP and eukaryotic RNAP II involve not only similarity in overall structural organization, but also detailed similarity in folding topologies of subunits (Cramer, 2002). In view of the similarities between the RNAPs and the similarities between the various subviral RNA domains reported to bind DNA-dependent RNAP, it is tempting to suggest a common mechanism of RNAP recognition acting on subviral RNA species. It seems clear that a defined hairpin domain is a common feature of RNA promoters for these DNA-dependent RNAPs. Such terminal stem–loop secondary structures with RNA features specific for each RNAP could thus represent "universal" RNA binding sites and/or promoter elements for these DNA-dependent RNAPs.

Materials and methods

Synthesis of HDV-derived RNAs

Polymerase chain reaction (PCR) with Vent polymerase (New England Biolabs; NEB) was used to amplify DNA fragments for the synthesis of both genomic and antigenomic cDNA fragments from a derivative of pBluescriptKS$^+$ (Stratagene) containing dimeric HDV genomic cDNA (Kuo et al., 1988a) flanked by T7 and T3 promoter recognition sites. P11.60 was synthetized as previously described (Pelchat and Perreault, 2004). All R38 and L48 derivatives were transcribed from double-stranded DNA oligonucleotides possessing a T7 RNA promoter. DNA products were used for *in vitro* run off transcription using T7 RNA polymerase (NEB). HDV dimeric RNA genomes were similarly synthesized using linearized plasmid containing a dimer of HDV cDNA as a template for transcription. After incubation of the transcription reactions at 37 °C for 2 h, the reactions were treated with DNase I (Promega) for an additional 30 min. The RNA products were electrophoresed on denaturing 5% or 10% polyacrylamide gels in 1× TBE buffer (100 mM Tris–borate, pH 8.3, 1 mM EDTA) and 7 M urea. Following electrophoresis, UV shadowing was used to detect the RNA bands, which were subsequently excised, eluted overnight at 4 °C in elution buffer (500 mM ammonium acetate, 0.1% SDS, 20 mM EDTA), precipitated with ethanol, and resuspended in H_2O. Debris and residual urea were removed by passage through Sephadex G-50 columns (Amersham Pharmacia Biotech) and the RNAs were precipitated with ethanol, resuspended in H_2O, quantified by spectrophotometry at 260 nm, and stored at −20 °C. CENTR molecules were generated using two RNA strands folded together by heat denaturation at 94 °C and slow renaturation at room temperature in 40 mM Tris–HCl, pH 8.0/10 mM $MgCl_2$.

Immunoprecipitation of HDV-derived RNAs

The Protein-G Immunoprecipitation Kit (Sigma-Aldrich) was employed to co-immunoprecipitate HDV RNAs in accordance with the manufacturer's protocol. Briefly, 5 pmol of RNA was 5′-end radiolabeled with [γ-^{32}P]ATP as previously described (Pelchat et al., 2002) and incubated with 26 μg of HeLaScribe Nuclear Extract (Promega) in RIP buffer (50 mM Tris–Cl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl; Niranjanakumari et al., 2002) at 37 °C for 30 min. A 20-fold excess of poly-A RNA or P11.60 was added to some reactions as non-specific competitors. A 20-fold excess of cold homologous RNA was used as a specific competitor. Approximately 5 μg of either mouse monoclonal antibody (8WG16; Upstate) directed against the Rpb1 C-terminal domain (CTD) of RNAP II or goat anti-mouse IgG (Sigma-Aldrich) was then added to the reaction mixtures and they were incubated for 2 h at 4 °C on a rotator. Pre-washed protein-G beads were then added to the mixtures and incubated overnight as above. Following incubation, reactions were centrifuged, washed 5 times in RIP buffer, once with 0.1× RIP buffer, and eluted according to manufacturer's protocols in acrylamide loading buffer (Sambrook et al., 1989). Samples were electrophoresed on denaturing 10% polyacrylamide gels and the bands were detected by autoradiography or phosphorimager scanning (Molecular Dynamics).

Ribonucleoprotein immunoprecipitation (RIP) assay

The RIP assay was performed as previously described (Niranjanakumari et al., 2002) with some modification. HeLa cells were transfected with 6 μg of pcDNA3-AgS and 10 μg of dimeric HDV RNA genome of positive polarity using Lipofectamine TM 2000 Reagent (Invitrogen) in accordance with manufacturer's protocol. pcDNA3-AgS is a derivative of pcDNA3 (Stratagene) encoding the small *delta* antigen under the control of a CMV promoter (Lazinski and Taylor, 1993). Formaldehyde (1%) was used to cross-link nucleic acids to cellular proteins for 5 min at 37 °C. About 10^6 cells were harvested by centrifugation, washed twice with ice-cold PBS,

the supernatant decanted, and pellets stored at −80 °C. Cells were resuspended in 2 ml RIP buffer, homogenized by sonication, and centrifuged at 16,000×*g* for 30 min at 4 °C. Immunoprecipitation was performed using the Protein-G Immunoprecipitation Kit as above using 500 µl of cell lysate. Samples were washed 5 times with high-stringency RIP wash buffer containing 1 M urea (Niranjanakumari et al., 2002), heated to 70 °C for 45 min in a solution of 50 mM Tris–HCl, pH 7.0, 5 mM EDTA, 10 mM dithiotriol (DTT), and 1% SDS to reverse the cross-links, and eluted by centrifugation. The nucleic acids were extracted by phenol/chloroform, ethanol precipitated, resuspended in H_2O, and treated with DNAse I (Promega) for 45 min at 37 °C. After a second RNA extraction with phenol/chloroform and precipitation with ethanol, the RNA was resuspended in H_2O. The RNA was heated for 5 min at 65 °C with 1 pmol of one of the primers used to generate the R199G cDNA (R199GFW: 5′-GGAATTCTAATACGACTCACTATAGGG^{1541}ACTGCTCGAGGATCTCTTCTCTCCC^{1565}-3′ or R199GREV: 5′-GGAATTC^{60}ACATCCCCTCTCGGGTGC^{43}-3′; underlined nucleotides indicate the T7 promoter) in 5 µl of annealing buffer (50 mM Tris–HCl, pH 8.3, 10 mM $MgCl_2$, 80 mM KCl solution), and then the mixture quickly snap frozen by immersion in −80 °C ethanol/dry ice for 2 min. Synthesis was performed by adding 4 units avian myeloblastosis virus reverse transcriptase (Amersham Pharmacia Biotech) in 5 µl of ice cold annealing buffer containing 8 mM DTT and 4 mM dNTP to the primer annealed templates, and then incubating at 42 °C for 30 min. The cDNA was extracted by phenol/chloroform, ethanol precipitated, and resuspended in H_2O. PCR amplification using the same primers was composed of 35 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min, and extension at 76 °C for 1°min. The primer B-ActinREV (5′-GGGGTACTTCAGGGTGAGGATGCCTCTCTT-3′) for the RT reaction and both B-ActinREV and B-ActinFW (5′-ATGGATGATGATATCGCCGCGCTCGTCGTC-3′) were used to perform RT-PCR to generate a small 200 nt cDNA fragment corresponding to the 5′-end of the β-actin mRNA. Bands were resolved by 2% agarose gel electrophoresis and visualized under UV light in the presence of ethidium bromide. Semi-quantitative RT-PCR was performed as above by ending the PCR amplification during the linear phase of amplification. Amplification curves were carried out by real-time RT-PCR of HDV RNA performed with an ABI Prism 7300 DNA analyzer (Applied Biosystems) in the presence of SYBR Green (Invitrogen).

Computer analysis of putative HDV promoter sequences.

All the HDV sequences were taken from the Subviral RNA Database (http://subviral.med.uottawa.ca/; Rocheleau and Pelchat, 2006). Multiple alignments of HDV sequences were obtained using ClustalW (Thompson et al., 1994). The alignments were corrected manually to maximize sequence identities and secondary structural features. Covariation analyses were performed using BioEdit v7.0.5 (Brown, 1991). Thermodynamic parameters of RNA promoters were established for each secondary structure using the RNAeval Software from the Vienna RNA Structure Software Suite (Hofacker, 2003). Logos were generated using WebLogo (Crooks et al., 2004).

Acknowledgments

Technical assistance in cell maintenance and transfection was provided by E.G. Brown and M. Gallant. V.S.G.-S. is supported by a Canadian Graduate Scholarship awarded by the Natural Science and Engineering Research Council of Canada. A.A. is supported by an award from the Libyan government. This work was funded by a grant from The Canadian Institutes of Health Research awarded to M.P.

References

Beard, M.R., MacNaughton, T.B., Gowans, E.J., 1996. Identification and characterization of a hepatitis *delta* virus RNA transcriptional promoter. J. Virol. 70, 4986–4995.

Brown, J.W., 1991. Phylogenetic comparative analysis on Macintosh computers. Comput. Appl. Biosci. 7, 391–393.

Casey, J.L., Bergmann, K.F., Brown, T.L., Gerin, J.L., 1992. Structural requirements for RNA editing in hepatitis *delta* virus: evidence for a uridine-to-cytidine editing mechanism. Proc. Natl. Acad. Sci. U. S. A. 89, 7149–7153.

Chang, F.L., Chen, P.J., Tu, S.J., Wang, C.J., Chen, D.S., 1991. The large form of hepatitis δ antigen is crucial for the assembly of hepatitis δ virus. Proc. Natl. Acad. Sci. 88, 8490–8494.

Chen, P.J., Kalpana, G., Goldberg, J., Mason, W., Werner, B., Gerin, J., Taylor, J., 1986. Structure and replication of the genome of hepatitis δ virus. Proc. Natl. Acad. Sci. 83, 8774–8778.

Cramer, P., 2002. Multisubunit RNA polymerases. Curr. Opin. Struct. Biol. 12, 89–97.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

de Mercoyrol, L., Job, C., Job, D., 1989. Studies on the inhibition by {alpha}-amanitin of single-step addition reactions and productive RNA synthesis catalysed by wheat-germ RNA polymerase II. Biochem. J. 258, 165–169.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J., Ingles, C.J., 2002. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA 8, 1102–1111.

Filipovska, J., Konarska, M., 2000. Specific HDV RNA-templated transcription by pol II *in vitro*. RNA 6, 41–54.

Fu, T., Taylor, J., 1993. The RNAs of hepatitis *delta* virus are copied by RNA polymerase II in nuclear homogenates. J. Virol. 67, 6965–6972.

Garcia-Blanco, M.A., Jamison, S.F., Sharp, P.A., 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. Genes Dev. 3, 1874–1886.

Gong, X.Q., Nedialkov, Y.A., Burton, Z.F., 2004. Alpha-amanitin blocks translocation by human RNA polymerase II. J. Biol. Chem. 279, 27422–27427.

Goodman, T.C., Nagel, L., Rappold, W., Klotz, G., Riesner, D., 1984. Viroid replication: equilibrium association constant and comparative activity measurements for the viroid polymerase interaction. Nucleic Acids Res. 12, 6231–6246.

Greco-Stewart, V.S., St-Laurent Thibault, C., Pelchat, M. in press, Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis *delta* virus RNA, Virology. doi:10.1016/j.virol.2006.06.040.

Gudima, S., Dingle, K., Wu, T.T., Moraleda, G., Taylor, J., 1999. Characterization of the 5′ends for polyadenylated RNAs synthesized during the replication of hepatitis *delta* virus. J. Virol. 73, 6533–6539.

Gudima, S., Wu, S.Y., Chiang, C.M., Moraleda, G., Taylor, J., 2000. Origin of hepatitis *delta* virus mRNA. J. Virol. 74, 7204–7210.

Gudima, S.O., Chang, J., Taylor, J.M., 2006. Restoration *in vivo* of defective hepatitis delta virus RNA genomes. RNA 12, 1061–1073.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. 11, 394–403.

Hofacker, I.L., 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429–3431.

Kolonko, N., Bannach, O., Aschermann, K., Hu, K.H., Moors, M., Schmitz, M., Steger, G., Riesner, D., 2006. Transcription of potato spindle tuber viroid by RNA polymerase II starts in the left terminal loop. Virology 347, 392–404.

Kuo, M.Y., Goldberg, J., Coates, L., Mason, W., Gerin, J., Taylor, J.M., 1988a. Molecular cloning of hepatitis *delta* virus RNA from an infected woodchuck liver: sequence, structure, and applications. J. Virol. 62, 1855–1861.

Kuo, M.Y., Sharmeen, L., Dinter-Gottlieb, G., Taylor, J., 1988b. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis *delta* virus. J. Virol. 62, 4439–4444.

Kuo, M.Y., Chao, M., Taylor, J., 1989. Initiation of replication of the human hepatitis *delta* virus genome from cloned DNA: role of the *delta* antigen. J. Virol. 63, 1945–1950.

Lai, M.M., 2005. RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis *delta* virus. J. Virol. 79, 7951–7958.

Lazinski, D.W., Taylor, J.M., 1993. Relating structure to function in the hepatitis *delta* virus antigen. J. Virol. 67, 2672–2680.

Li, Y.J., Macnaughton, T., Gao, L., Lai, M.M., 2006. RNA-templated replication of hepatitis *delta* virus: genomic and antigenomic RNAs associate with different nuclear bodies. J. Virol. 80, 6478–6486.

Macnaughton, T.B., Gowans, E.J., MacNamara, S.P., Burrell, C.J., 1991. Hepatitis δ antigen is necessary for access for hepatitis δ virus RNA to the cell transcriptional machinery but is not part of the transcriptional complex. Virology 184, 387–390.

Macnaughton, T.B., Shi, S.T., Modahl, L.E., Lai, M.M., 2002. Rolling circle replication of hepatitis *delta* virus RNA is carried out by two different cellular RNA polymerases. J. Virol. 76, 3920–3927.

Mathews, D.B., Durbin, R.D., 1994. Mechanistic aspects of tagetitoxin inhibition of RNA polymerase from *Escherichia coli*. Biochemistry 33, 11987–11992.

Meienhofer, J., Atherton, E., 1973. Structure–activity relationships in the actinomycins. Adv. Appl. Microbiol. 16, 203–300.

Modahl, L.E., Macnaughton, T.B., Zhu, N., Johnson, D.L., Lai, M.M., 2000. RNA-dependent replication and transcription of hepatitis *delta* virus RNA involve distinct cellular RNA polymerases. Mol. Cell. Biol. 20, 6030–6039.

Navarro, J.A., Flores, R., 2000. Characterization of the initiation sites of both polarity strands of a viroid RNA reveals a motif conserved in sequence and structure. EMBO J. 19, 2662–2670.

Nie, X., Chang, J., Taylor, J.M., 2004. Alternative processing of hepatitis *delta* virus antigenomic RNA transcripts. J. Virol. 78, 4517–4524.

Niranjanakumari, S., Lasda, E., Brazas, E., Garcia-Blanco, M.A., 2002. Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions *in vivo*. Methods 26, 182–190.

Patturajan, M., Schulte, R.J., Sefton, B.M., Berezney, R., Vincent, M., Bensaude, O., Warren, S.L., Corden, J.L., 1998. Growth-related changes in phosphorylation of yeast RNA polymerase II. J. Biol. Chem. 273, 4689–4694.

Pelchat, M., Perreault, J.P., 2004. Binding site of *Escherichia coli* RNA polymerase to an RNA promoter. Biochem. Biophys. Res. Commun. 319, 636–642.

Pelchat, M., Côté, F., Perreault, J.P., 2001. Study of the polymerization step of the rolling circle replication of peach latent mosaic viroid. Arch. Virol. 146, 1753–1763.

Pelchat, M., Grenier, C., Perreault, J.P., 2002. Characterization of a viroid-derived RNA promoter for the DNA-dependent RNA polymerase from *Escherichia coli*. Biochemistry 41, 6561–6571.

Rocheleau, L., Pelchat, M., 2006. The Subviral RNA Database: a toolbox for viroids, the hepatitis *delta* virus and satellite RNAs research. BMC Microbiol. 6, 24.

Ryu, W.S., Bayer, M., Taylor, J., 1992. Assembly of hepatitis *delta* virus particles. J. Virol. 66, 2310–2315.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Press, New York.

Sims III, R.J., Mandal, S.S., Reinberg, D., 2004. Recent highlights of RNA-polymerase-II-mediated transcription. Curr. Opin. Cell Biol. 16, 263–271.

Smale, S.T., Kadonaga, J.T., 2003. The RNA polymerase II core promoter. Annu. Rev. Biochem. 72, 449–479.

Swanson, M.S., Dreyfuss, G., 1988. RNA binding specificity of hnRNP proteins: a subset bind to the 3′ end of introns. EMBO J. 7, 3519–3529.

Taylor, J.M., 2006. Hepatitis *delta* virus. Virology 344, 71–76.

Thompson, N.E., Steinberg, T.H., Aronson, D.B., Burgess, R.R., 1989. Inhibition of *in vivo* and *in vitro* transcription by monoclonal antibodies prepared against wheat germ RNA polymerase II that react with the heptapeptide repeat of eukaryotic RNA polymerase II. J. Biol. Chem. 264, 11511–11520.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Wang, J.G., Cullen, J., Lemon, S.M., 1992. Immunoblot analysis demonstrates that the large and small forms of hepatitis *delta* virus antigen have different C-terminal amino acid sequences. J. Gen. Virol. 73, 183–188.

Wang, H.W., Wu, H.L., Chen, D.S., Chen, P.J., 1997. Identification of the functional regions required for hepatitis D virus replication and transcription by linker-scanning mutagenesis of viral genome. Virology 239, 119–131.

Weiner, A.J., Choo, Q.L., Wang, K.S., Govindarajan, S., Redeker, A.G., Gerin, J.L., Houghton, M., 1988. A single antigenomic open reading frame of the hepatitis *delta* virus encodes the epitope(s) of both hepatitis delta antigen polypeptides p24 *delta* and p27 *delta*. J. Virol. 62, 594–599.

Wu, T.T., Netter, H.J., Lazinski, D.W., Taylor, J.M., 1997. Effects of nucleotide changes on the ability of hepatitis *delta* virus to transcribe, process, and accumulate unit-length, circular RNA. J. Virol. 71, 5408–5414.

Xia, Y.P., Chang, M.F., Wei, D., Govindarajan, S., Lai, M.M., 1990. Heterogeneity of hepatitis *delta* antigen. Virology 178, 331–336.

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M.M., Handa, H., 2001. Stimulation of RNA polymerase II elongation by hepatitis *delta* antigen. Science 293, 124–217.

Appendix III: **Abrahem A.**, Pelchat M. Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis delta virus RNA genome. Nucleic Acids Res., Advance Access published on August 5, 2008

Published online 5 August 2008

Nucleic Acids Research, 2008, Vol. 36, No. 16 ***5201–5211***
doi:10.1093/nar/gkn501

Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis *delta* virus RNA genome

Abrahem Abrahem and Martin Pelchat*

Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada

Received March 28, 2008; Revised June 19, 2008; Accepted July 18, 2008

ABSTRACT

Although RNA polymerases (RNAPs) are able to use RNA as template, it is unknown how they recognize RNA promoters. In this study, we used an RNA fragment derived from the hepatitis *delta* virus (HDV) genome as a model to investigate the recognition of RNA promoters by RNAP II. Inhibition of the transcription reaction using an antibody specific to the largest subunit of RNAP II and the direct binding of purified RNAP II to the RNA promoter confirmed the involvement of RNAP II in the reaction. RNA affinity chromatography established that an active RNAP II preinitiation complex forms on the RNA promoter and indicated that this complex contains the core RNAP II subunit and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TFIIS. Binding assays demonstrated the direct binding of the TATA-binding protein and suggested that this protein is required to nucleate the RNAP II complex on the RNA promoter. Our findings provide a better understanding of the events leading to RNA promoter recognition by RNAP II.

INTRODUCTION

Cellular DNA-dependent RNA polymerases (RNAPs) are able to use RNA as template. T7 RNAP was shown to replicate a 64-nt RNA contaminant of unknown genetic origin in a commercial sample of the enzyme, and at high concentrations, is able to transcribe from a large variety of different RNA species (1,2). Another RNAP frequently associated with RNA-dependent RNA polymerization is the *Escherichia coli* RNA polymerase. This RNAP was shown to be capable of amplifying selected small, random, RNA polymers (3,4), to specifically initiate transcription from RNA templates derived from a stem-loop domain of the peach latent mosaic viroid RNA genome (5,6), and was recently reported to synthesize short RNA products from endogenous bacterial 6S RNA (7). Similarly, the small noncoding RNA genomes of plant viroids are replicated either in the nucleus by DNA-dependent RNAP II (8) or in the chloroplast by chloroplastic RNAP (9). Among the known mammalian RNA pathogens, the human hepatitis *delta* virus (HDV) is the only one known to utilize this potential of RNAP II to replicate itself (10).

HDV is the smallest known animal virus. Its genome consists of a small (~1700 nt) single-stranded, circular RNA molecule, and is thought to fold into an unbranched, rod-like structure (11). HDV contains two self-cleaving motifs (i.e. *delta* ribozymes) and encodes a single open reading frame (ORF). There are two viral proteins (HDAg) encoded by this ORF (i.e. HDAg-S and HDAg-L; 12,13). The large HDAg (HDAg-L) contains an additional 19 amino acids at its C-terminus resulting from RNA editing of the termination codon of the small HDAg (HDAg-S) gene (14). Although they are mostly identical in sequences, each protein has distinct functions. HDAg-S (195 amino acids) is essential for HDV replication (15), while the HDAg-L (214 amino acids) is necessary for virion assembly (16).

Replication of HDV is considered to take place in the nucleus by a symmetrical, rolling circle mechanism. During this replication, infecting HDV genomic circular monomer is replicated into linear multimeric minus strands which are then cleaved and ligated, yielding antigenomic circular monomers (17). Using the latter RNA as template, the same three steps are then repeated to produce the genomic progeny. During this process, RNAP II is believed to be involved in the transcription of the HDAg mRNA because the mRNA is posttranscriptionally processed with a 5′-cap and a 3′-poly(A) tail (18,19). Furthermore, studies using cultured cells and nuclear extracts (NE) have reported that low levels of α-amanitin, a known inhibitor of RNAP II, inhibits the accumulation of HDV mRNA and genomic HDV RNAs (20–23). Recently, we used a monoclonal antibody specific to the carboxy terminal domain (CTD) of the largest subunit of

*To whom correspondence should be addressed. Tel: +613 562 5800 (ext. 8630); Fax: +613 562 5452; Email: mpelchat@uottawa.ca

RNAP II, to establish the association of RNAP II with both polarities of HDV RNA in HeLa cells (24). This analysis revealed that RNAP II associates with the terminal stem-loop domains of both polarities of HDV RNA (24). In addition, RNAP I or an RNAP I-like polymerase might be involved in HDV replication (25–27). The accumulation of the antigenomic species is resistant to higher doses of α-amanitin and synthesis of HDV RNA was affected by an anti-SL1 antibody (27).

RNAP II is a multisubunit enzyme that is known to catalyze the synthesis of mRNAs from DNA templates (28,29). The two large subunits of human RNAP II [i.e. RPB1 (~220 kDa) and RPB2 (~140 kDa)] form the catalytic domain through which the DNA–RNA assembly takes place. Initiation of DNA-templated transcription by RNAP II involves multiple events, including decondensation of the locus, nucleosome remodeling, histone modification, binding of the activator and coactivator to the promoter elements, and recruitment of the general transcription factors to the promoter (30). The specific binding of the polymerase to the promoter requires the coordinated assembly of RNAP II and the six general transcription factors [i.e. TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH; (31–33)]. In addition, sequence-specific transcriptional activators play an important role in the recruitment of the RNAP II holoenzyme complex to the promoter region. In particular, TFIID and TFIIB were shown to play critical roles in the recognition of DNA promoter motifs by directly interacting with the TATA box motif (34), which is located at a fixed distance upstream of the transcription start site. There is considerable evidence that TFIID also binds to the initiator element (Inr), encompassing the transcription start site, in a sequence-specific manner and to the downstream core promoter element (DPE) (35). In addition, TFIIB is able to bind directly to the TFIIB recognition element (BRE), located immediately upstream of some TATA boxes (36).

Although many studies have investigated DNA-dependent RNA polymerization and DNA promoter recognition by RNAP II, little is known regarding how this enzyme recognizes an RNA template, and what the composition of the RNAP II preinitiation complex (PIC) on such RNA promoter might be. Thus, HDV offers a perfect model to study this unconventional use of RNAP II and the molecular mechanism underlying RNA template recognition by RNAP II. The present study was undertaken to (i) investigate the direct interaction of the RNAP II holoenzyme with an HDV-derived RNA template and (ii) to define the RNAP II subunit(s) involved in the formation of the PIC on the RNA promoter. For this purpose, we used an RNA fragment derived from the HDV RNA genome that was previously reported to include the initiation site for HDAg mRNA transcription (18) and confirmed that RNAP II binds directly to the RNA molecule and initiates transcription. RNA affinity chromatography was then used to identify the components of the RNAP II PIC forming on this unusual promoter, which led us to identify the TATA-binding protein (TBP) as an RNAP II subunit directly interacting with the RNA promoter. Together, our results support a model in which an RNA promoter is recognized by RNAP II in the same way as a DNA promoter. Moreover, the mechanism of promoter recognition likely requires the involvement of the TBP-containing complex TFIID to nucleate the RNAP II complex on the HDV-derived RNA promoter.

MATERIALS AND METHODS

Synthesis of RNAs

RNA molecules were synthesized by *in vitro* run-off transcription using T7 RNA polymerase (New England Biolabs; Pickering, Ontario, Canada NEB), as previously described (24). Sense (5′-GGAATTCTAATACGACTCACTATAGGG^{1541}ACTGCTCGATCTCTT^{1558}-3′; underlined nucleotide sequence indicates T7 promoter) and antisense (5′-GGAATTC^{60}ACATCCCCTCTCGGGTAC^{43}-3′) oligonucleotides were used during PCR amplification to generate R199G cDNA templates from a derivative of pBluescriptKS + (Stratagene, La Jolla, CA, USA) containing a dimeric EcoRI-flanked HDV cDNA insert [pHDVd2; (37)]. To generate X-R199G cDNA templates, 5′-GGAATTCTAATACGACTCACTATAGGGGGGA ACAAGGG^{1541}ACTGCTCGATCTCTT^{1558}-3′ was used as the sense oligonucleotide. X-R199GΔUUA cDNA was derived from X-R199G cDNA by PCR amplification using several oligonucleotides, in order to delete nucleotide 1629–1631 to remove the UUA sequence at the initiation site. R38G, R38GPU and R38GSW were produced from double-stranded DNA oligonucleotides possessing a T7 RNA promoter (24). P11.60 RNA was generated as previously described (6). Following the transcription reactions, the products were digested with DNase I (Promega, Madison, Wisconsin, USA) for 30 min at 37°C, fractionated by denaturing polyacrylamide gel electrophoresis (PAGE) in 1 × TBE buffer (100 mM Tris–borate, pH 8.3, 1 mM EDTA) and 7 M urea, visualized by UV shadowing, excised, and eluted overnight at 4°C in 500 mM ammonium acetate, 0.1% SDS. The RNAs were then precipitated in ethanol, resuspended in H_2O, desalted by Sephadex G-50 columns (GE Healthcare, Baie d'Urfé, Québec, Canada), and precipitated in ethanol. The purified RNAs were then resuspended in H_2O, quantified spectrophotometrically at 260 nm and stored at −20°C for further use.

Transcription assay

RNA templates (50 pmol) were resuspended in 25 µl of transcription buffer (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 6.0 mM $MgCl_2$, 0.5 mM DTT and 20% glycerol) containing 0.4 mM NTPs. Either HeLa NE (Cedarlane, Burlington, Ontario, Canada) or HeLa NE preincubated with 2.5 µg of anti-RNAP II antibody (CTD4H8; Upstate Lake Placid, NY, USA) for 1 h on ice, was added and the reaction was incubated at 30°C for 60 min. The transcription reactions were stopped by adding 175 µl of HeLa extract stop solution (0.3 M Tris–HCl pH 7.4, 0.3 M sodium acetate, 0.5% SDS, 2 mM EDTA and 3 µg/ml tRNA), subjected to phenol–chloroform extraction, ethanol precipitation and resuspended in 40 µl H_2O. RNA products were subjected to reverse transcription, as described previously (24),

using primer X (5′-GGGGGGAACAAGGGAC-3′). PCR was performed using primer X and either primer B (5′-^{1631}TAAGAGTACTGAGGACTGCC^{1612}-3′) or primer A (5′-^{1651}CTCCTCTTTACAGAAAAGAG^{1632}-3′), the DNA molecules resolved on a 2% agarose gel and the bands visualized using SYBR green. The 5′-end of the RNA product was determined using the '5′ RACE System for Rapid Amplification of cDNA Ends, Version 2.0' (Invitrogen, Burlington, Ontario, Canada) and primer X, and the PCR products were sequenced directly by Bio Basic Inc. (Markham, Ontario, Canada).

Electrophoresis mobility shift assay

RNA molecules were 5′-end radiolabeled with [γ-^{32}P]ATP (GE Health Sciences) by dephosphorylation using calf intestinal phosphatase and phosphorylation using T4 polynucleotide kinase according to the manufacturer's recommended protocol (NEB, Pickering, Ontario, Canada). The 5′-end radiolabeled RNA molecules were incubated with either purified RNAP II, TFIID or GST–TBP (ProteinOne) in the presence or absence of either 50-fold molecular excess of P11.60, unlabeled R199G or a double-stranded DNA fragment containing the CMV promoter (5′-CCAAAATCAACGGGACTT TCCAAAATGTCGTAACAAGTGGGGCCCATTGA GGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAG-3′) in 20 µl of binding buffer (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT and 20% glycerol). Unless otherwise indicated, 0.1 pmol of 5′-end radiolabeled RNA, 50 ng of RNAP II, 50 ng of TFIID or 64 ng of GST–TBP were used. After incubation at 30°C for 30 min the complexes were resolved on 5% PAGE (*bis*-acrylamid:acrylamide, 1:49) under *native conditions at room temperature in* 1 × TBE. The gels were then exposed to a phosphor screen overnight and scanned using ImageQuant software (Molecular Dynamics, Sunnyvale, CA, USA).

RNA affinity column preparation

RNA molecules (200 pmol) were oxidized in 20 mM Tris–HCl pH 7.5 and 10 mM Na-m-periodate in a total volume of 100 µl. The reaction mixtures were incubated for 1 h at 4°C in the dark, ethanol precipitated and resuspended in 60 µl of 0.1 M sodium acetate (pH 5.0). Adipic acid dihydrazide agarose beads (400 µl of 50% slurry; Sigma-Aldrich, Oakville, Ontario, Canada) were prepared by washing four times in 0.1 M sodium acetate. For direct coupling, the beads were resuspended in 600 µl of 0.1 M sodium acetate and loaded onto filter tubes and incubated with the oxidized RNA overnight at 4°C. After washing four times with 2 M NaCl to remove unbound RNA, 15 µg of the HeLa NE (Cedarlane) was added to the beads, and the mixture was incubated at room temperature for 30 min. The columns were washed three times with wash solution [50 mM Tris–Cl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl], eluted using SDS loading dye (50 mM Tris–HCl, pH 6.8, 2% SDS, 0.1% BPB and 10% glycerol), and the eluted proteins were separated by 10% SDS–PAGE (38). After transfer onto a nitrocellulose membrane, the proteins were visualized using X-ray films using the One-Step Western™ Complete Kit (Genscript, Piscataway, NJ, USA) according to the manufacturer's protocol. The following antibodies targeted against RNAP II CTD (CTD4H8; Upstate), TBP (Upstate), p55 (ProteinOne, Bethesda, MD, USA), p56 (ProteinOne), RAP74 (ProteinOne), SC35 (ProteinOne), TFIIB (IIB8; ABR—Affinity BioReagents, Golden, CO, USA), TCEA3 (Abcam, Cambridge, MA, USA), and Cdk7 (Abcam) were used. To perform the transcription initiation reactions, transcription buffer and 0.4 mM NTPs were added to the RNA–protein mixture following the last wash. The mixtures were incubated for 2 min at room temperature, the columns were washed and the samples were eluted and analyzed as above. When required for subsequent RT–PCR, the samples were eluted by the addition of transcription buffer, subjected to phenol–chloroform extraction, and ethanol precipitated. The RNA was then resuspended in 40 µl H_2O and analyzed by RT–PCR as above. For the experiments involving a control DNA promoter, two oligonucleotides were used to make the CMV DNA promoter, one of which was biotinylated at its 5′-end (Invitrogen). To guarantee that both DNA molecules forming the DNA promoter annealed, both strands were mixed, heated at 95°C for 10 min and incubated at room temperature for 20 min. The DNA promoter was then added to streptavidin beads (Sigma-Aldrich) that were prewashed three times with the wash solution. The mixture was incubated for 1 h in the presence of the binding buffer and washed four times with the same solution. Uncoupled streptavidin beads were used as control. HeLa NE proteins (15 µg) were added and the streptavidin columns were processed in the same manner as the RNA affinity columns.

RESULTS

The right terminal stem-loop domain of the genomic polarity of HDV contains an RNA promoter for RNAP II

To gain further insight into the mechanism by which RNAP II recognizes RNA promoters, we required the identification of a suitable RNA promoter for RNAP II. Previous works have shown that a region we refer to as the right terminal stem-loop domain of the genomic polarity of HDV (i) includes the reported initiation site for HDAg mRNA transcription (i.e. position 1630; 18), (ii) was speculated to be used by RNAP II because the HDAg mRNA is posttranscriptionally processed with a 5′-cap and a 3′-poly(A) tail (18,19), (iii) mutation analysis of this region revealed that the overall rod-like conformation was important for the accumulation of HDV RNA *in vivo* and for both RNAP II binding and transcription initiation *in vitro* (24,39).

To corroborate that this segment of the HDV RNA genome acts as a promoter, we used an RNA fragment containing 199 nt corresponding to nucleotide 1541 to 60 of the genomic polarity of HDV RNA (Figure 1A; R199G), and performed transcription assays using HeLa NE proteins. To detect the product, RT–PCR was performed following the transcription reaction.

However, the RT reaction was not able to discriminate between the RNA product and the large molecular excess of its RNA template because they have very similar sequences (data not shown). Therefore, a non-HDV sequence was added to the 5′-end of R199G to generate an extended RNA template (i.e. X-R199G; inverted sequence in Figure 1A). Transcription reaction was performed on X-R199G, and the resulting RNA was subjected to reverse transcription using a primer complementary to the 3′ extended extremity of the product (i.e. primer X; Figure 1A). The cDNA product was subjected to rapid amplification of cDNA ends (RACE) and resolved on an agarose gel. A specific RT–PCR product migrating between the 100- and 150-nt markers was detected when the RACE reaction was performed in the presence of X-R199G (Figure 2A), indicating the production of an RNA species complementary to the upper strand of the RNA template (i.e. of antigenomic polarity). To locate the initiation site of the transcription, the RACE product was sent for sequencing. Analysis of the sequence indicated that transcription initiated near or at the proposed initiation site for HDAg mRNA (Figure 1A, the asterisks next to position 1630; 18). This result is in accordance with a previous report using a similar HDV-derived RNA molecule as template for *in vitro* transcription using HeLa NE proteins and showing that the initiation of synthesis occurred near this location (39). To further confirm the location of the initiation site, an RNA template with a deletion of nucleotide 1629–1631 was constructed to remove the UUA sequence at the initiation site

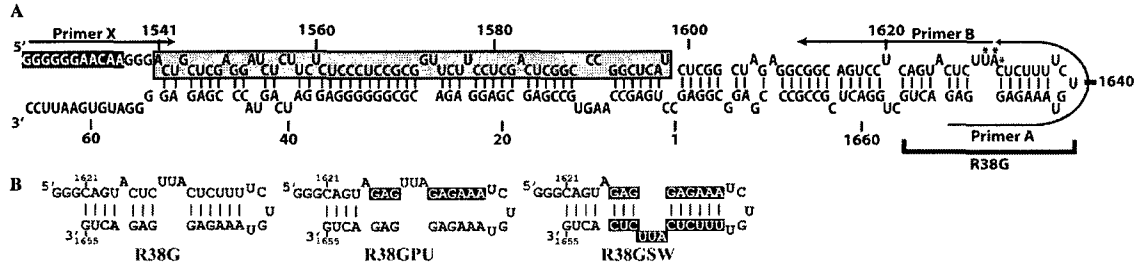


Figure 1. Representation of the predicted secondary structures of the RNA molecules used in this study. (**A**) Unbranched rod-like secondary structure of R199G and X-R199G. The initiation sites of transcription as determined in this study by 5′ RACE are marked with asterisks. The gray box indicates the 5′ extremity of the HDAg gene, the inverted characters represent the nucleotides that were added to R199G to form X-R199G, and the square bracket represents the location of R38G. Location of the primers used in this study is indicated. (**B**) Predicted secondary structures of R38G, R38GPU and R38GSW. Inverted characters represent mutations from R38G.

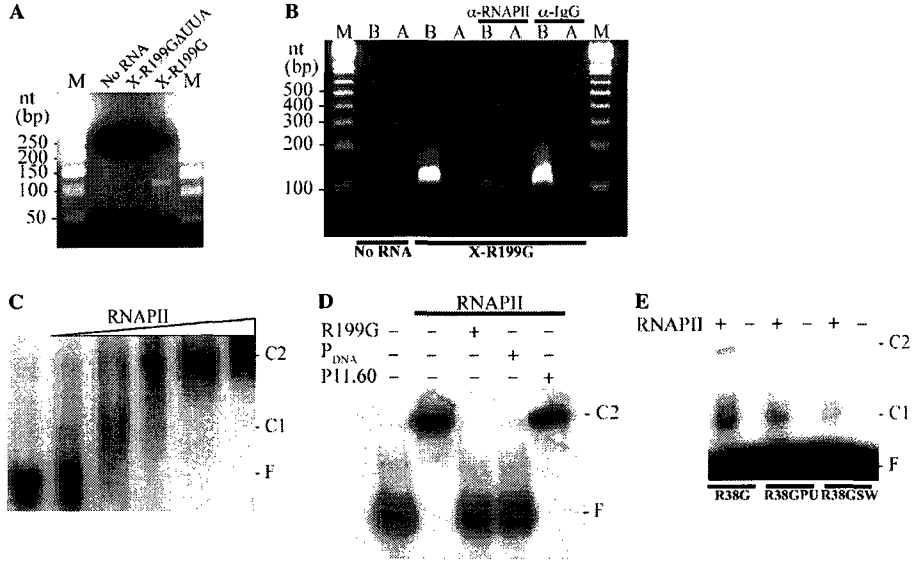


Figure 2. RNAP II interacts with and transcribes from the right terminal stem-loop extremity of the genomic polarity of HDV RNA. (**A**) 5′ RACE reactions performed on the transcription product using either no RNA, X-R199G or X-R199GΔUUA. Lanes M contain MiniSizer 50-bp DNA ladder (Norgen, Thorold, Ontario, Canada). (**B**) Inhibition of the transcription reaction by a monoclonal antibody against the CTD of RNAP II. RT-PCR products were generated following transcription reactions performed on X-R199G using primer X with either primer B or primer A. Lanes M contain 1 Kb Plus DNA ladder (Invitrogen). (**C**) Titration of a constant amount of radiolabeled R199G with increasing amounts of purified RNAP II holoenzyme (0, 6.25, 12.5, 25, 50 and 100 ng). (**D**) EMSA of R199G bound to purified RNAP II holoenzyme in the presence of 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G), a control DNA promoter fragment (i.e. CMV promoter) or P11.60. (**E**) Binding of R38G and two RNA mutants to purified RNAP II holoenzyme in the presence of a 50-fold molar excess of P11.60. The free RNA (F) and both complexes observed (C1 and C2) are indicated.

(i.e. X-R199GΔUUA). Following *in vitro* transcription and RACE, no HDV-dependent product was found using this construct (Figure 2A).

To verify the involvement of RNAP II, the transcription reaction was repeated in the presence of a monoclonal antibody specific for the CTD of RNAP II. However, the procedure described above was modified by performing PCR amplification on the cDNA using primer X and primers corresponding to sequences located on either side of the initiation site of transcription (Figure 1A, primers A and B). Using this approach, a specific RT–PCR product was detected when the transcription reaction was performed in the presence of X-R199G (Figure 2B). This product was not detected in transcription reactions where X-R199G was omitted, indicating the RNA template dependence of the reaction. More importantly, the product was only detected when primer B was used during the PCR amplification, in accordance with the initiation of synthesis occurring near or at position 1630. When the NE was preincubated with an antibody specific for the CTD of RNAP II, the production of the transcription product was strongly inhibited (Figure 2B). This inhibitory effect of the anti-RNAP II antibody is in accordance with previous studies reporting inhibition of DNA-directed RNAP II transcription by an antibody specific for the CTD of RNAP II (40). As a control for nonspecific effects of the antibody, the RNAP II antibody was replaced by an anti-IgG antibody. The addition of the anti-IgG antibody had no inhibitory effects on the transcription (Figure 2B). Together, these data indicate that the right terminal stem-loop domain of the genomic polarity of HDV contains an RNA promoter, and provide strong evidence that this RNA promoter is recognized by RNAP II. In addition, since the RNAP II antibody was targeted against the heptad repeats of the large subunit of RNAP II, our results indicate that the CTD of RNAP II is needed for the transcription from this RNA template.

RNAP II binds directly to R199G

For initiation to occur, the RNAP II holoenzyme must bind to promoter sequences to form a PIC. To assess the direct binding of RNAP II to R199G, we performed electrophoretic mobility shift assays (EMSAs) using purified RNAP II holoenzyme. Radiolabeled R199G was allowed to bind to increasing amounts of RNAP II and the mixture was subjected to nondenaturing PAGE. RNAP II bound directly to R199G, as illustrated by the retardation of the migration of the complex when compared to free RNA (Figure 2C). Notably, two retarded species were detected, and their formation was dependent on RNAP II concentration; a fast migrating species first formed at low concentrations of RNAP II holoenzyme (Figure 2C, C1), followed by a slower migrating species formed at higher protein concentrations (Figure 2C, C2). R199G was completely complexed in the slower migrating species when the molar ratio for RNAP II and the RNA was approximately 1:1 (i.e. 50 ng of RNAP II; Figure 2C and D).

We next investigated the specificity of the interaction between R199G and RNAP II by challenging the formation of the complex with various competitors. As an unrelated RNA competitor, we added a small RNA hairpin derived from the left terminal stem-loop domain of the peach latent mosaic viroid genome (i.e. P11.60; 5). This small RNA folds into a hairpin structure, thus providing an excellent competitive RNA to test for nonspecific binding of RNAP II to double-stranded RNA molecules or stem-loop secondary structures. As shown in Figure 2D, 50-fold molar excess of P11.60 did not displace RNAP II from R199G. However, inclusion of 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G) or a control DNA promoter fragment (i.e. CMV promoter) greatly inhibited formation of the radiolabeled R199G–RNAP II complexes. Finally, to provide evidence that RNAP II was binding to an internal site on R199G close to the initiation site, we performed EMSA with R38G, which corresponds to a 38-nt RNA located at the tip of the hairpin (Figure 1; 24). As observed for R199G, two retarded species were detected when radiolabeled R38G was allowed to bind to RNAP II (Figure 2E). These results are in accordance with our previous report indicating that R199G can be reduced to a smaller hairpin without significantly affecting its interaction with RNAP II in HeLa NE (24).

EMSAs were performed with two R38G derivatives (Figure 1B; 24). R38GPU contains mutations changing the pyrimidine-rich sequence located upstream of the terminal loop to purines, thus destabilizing the rod-like secondary structure. R38GSW is a 'flip mutant' of the region containing the initiation site. These two mutants are deficient in their interaction with the largest subunit of RNAP II in NE (24), and no transcription products were detected from these RNA species (data not shown). For both mutants, the amount of two retarded species was reduced, as compared to when R38G was used (Figure 2E). The relative intensities of the fast migrating species were reduced to 65 and 31% for R38GPU and R38GSW, respectively. The slower migrating species were reduced to 25 and 4% for R38GPU and R38GSW, respectively. These results correlate with our previous report suggesting that the maintenance of the rod-like conformation and the polarization in the purine/pyrimidine content near the tip of the rod of both polarities of HDV RNA might be required for RNAP II interaction and HDV replication (24). Taken together, these results provide direct evidence that the RNAP II holoenzyme interacts specifically with an internal site within R199G and warrant further investigation into PIC formation on R199G.

An active PIC forms on R199G

If similar mechanisms exist between DNA- and RNA-directed transcriptions by RNAP II, the specific binding of the polymerase to the HDV RNA promoter should require the coordinated assembly of RNAP II and general transcription factors. To identify the transcription machinery recruited to form a PIC on R199G, we used an affinity purification procedure that involves the cross-linking of R199G to adipic acid dehydrazide agarose beads (41). Such an approach allowed us to monitor the entire transcription machinery assembly on R199G using

active HeLa NE proteins. To compare PIC formation between R199G and a control DNA promoter, the CMV DNA promoter was biotinylated and immobilized on streptavidin agarose beads. As a negative RNA control, P11.60 was also immobilized on adipic acid dehydrazide agarose beads and subjected to the same treatment as R199G. In addition, the complexes on R199G were challenged with a large molecular excess of either the DNA promoter fragment or R199G to assess of the specificity of PIC formation. Finally, unbound adipic acid dehydrazide agarose and streptavidin agarose beads were tested for protein interaction to rule out nonspecific binding of NE proteins to the beads.

Nucleic acid-coupled beads were incubated with HeLa NE proteins and were washed to remove unbound proteins. The ribonucleoprotein complexes were eluted with SDS-loading dye, separated by SDS–PAGE, and analyzed by western blot using antibodies raised against various proteins involved in RNAP II transcription initiation. Using this approach, we detected the presence of RPB1 (RNAP II), p55 (TFIIA), TFIIB, TBP (TFIID), p56 (TFIIE), RAP74 (TFIIF), CAK (TFIIH) and TCEA3 (TFIIS) in association with both R199G and the DNA promoter (Figure 3A). As expected, no significant binding of the RNAP II transcriptional machinery to either P11.60 or to the beads alone was observed, indicating that the interactions with R199G were specific. Furthermore, molecular excess of the DNA promoter and R199G was able to abolish the protein interactions with R199G. In addition, because HDV was previously suggested to be associated with SC35-containing nuclear speckles (42), we tested for the presence of SC35 to serve as a control differentiating between proteins binding to R199G but not to the control DNA promoter. As shown in Figure 3A, SC35 interacted with R199G but not with the DNA promoter.

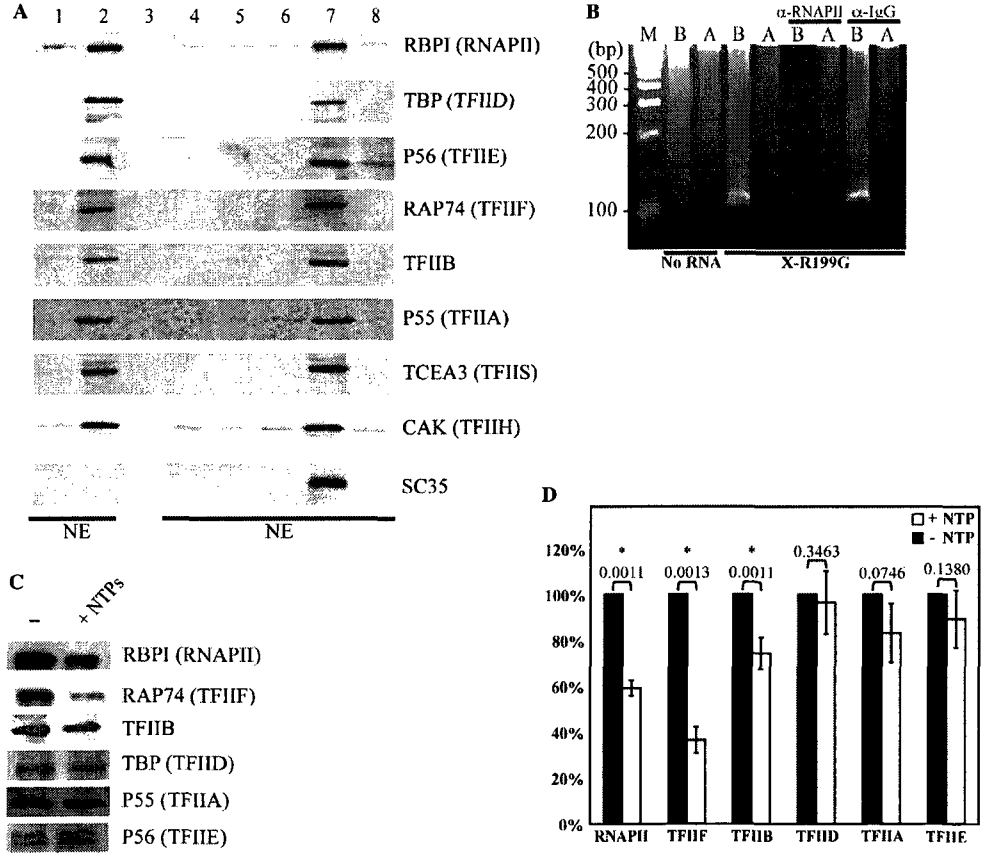


Figure 3. Formation of an active PIC on R199G. (**A**) Composition of a PIC on R199G. The samples were incubated with HeLa NE proteins in the presence or absence of either molecular excess of free DNA promoter or R199G. After elution, the proteins were detected using antibodies specific for RNAP II and its general transcription factors. Lane 1 contains the streptavidin agarose beads; lane 2 contains the biotinylated control DNA promoter bound to streptavidin agarose beads; lanes 3 and 8 contain adipic acid agarose beads; lane 4 contains P11.60 bound to adipic acid agarose beads; lane 5 contains R199G bound to adipic acid agarose beads in the presence of molecular excess of free control DNA promoter; lane 6 contains R199G bound to adipic acid agarose beads in the presence of molecular excess of free R199G; lane 7 contains R199G bound to adipic acid agarose beads. With the exception of lane 3, all the samples were incubated with HeLa NE. (**B**) Initiation of transcription on the bound PIC following the addition of NTPs. Following transcription initiation on X-R199G-coupled beads, RT–PCR products were generated using primers complementary to sequences located either before (i.e. primer B) or after (i.e. primer A) the proposed initiation site for HDAg mRNA transcription (18). Lanes M contain 1 Kb Plus DNA ladder (Invitrogen). For some reactions, monoclonal antibodies against either the CTD of RNAP II or IgG were used. (**C** and **D**) Decrease in the levels of various PIC components following initiation of transcription. The experiment was repeated three times and *P*-values (Student's *t*-test) were obtained (indicated above the bars). Asterisks indicate significant differences ($P < 0.05$).

Although R199G is an RNA promoter for RNAP II, it was possible that the observed interactions of R199G with RNAP II and its general transcription factors were not related to PIC formation and transcription initiation. To test this possibility, X-R199G-coupled beads were incubated with HeLa NE proteins, as above. After removing unbound proteins, the ribonucleoprotein complexes were equilibrated with transcription buffer and free NTPs were added to allow initiation of transcription to occur. After elution of the sample, RT–PCR was performed as above. The specific RT–PCR product was detected when the transcription reaction was performed in the presence of X-R199G (Figure 3B), demonstrating RNA synthesis by the immobilized RNAP II complex. As observed above, the transcription product from the bound RNA template was only detected when primer B was used during the PCR amplification, indicating that the initiation of synthesis occurred near the proposed initiation site for HDAg mRNA. In addition, no product was detected in absence of the X-R199G template, or when the NE was preincubated with anti-RNAP II, indicating the involvement of both RNAP II and the RNA template in this process. To further confirm the location of the initiation site of the transcription, the cDNA product was subjected to RACE, as above. Analysis of the sequence of the RACE product indicated that transcription initiated at the right location (data not shown). More importantly, these results indicate that an active RNAP II PIC forms on the bound RNA template and that this complex is able to perform the transcription reaction.

Using the same procedure, we examined the constituents of the RNAP II complex following transcription initiation. Addition of free NTPs significantly reduced the level of RPB1 (RNAP II), TFIIB and RAP74 (TFIIF) retained on R199G (Figure 3C and D). In contrast, no significant release of either p55 (TFIIA), TBP (TFIID) or P56 (TFIIE) was observed (Figure 3C and D). Such a phenomenon is in agreement with what was observed to occur on canonical DNA promoters (32). Specifically, the core subunits of RNAP II and various transcription factors leave the DNA promoters after initiation, whereas a subset of the transcriptional machinery (i.e. TFIIA, TFIID, TFIIE and TFIIH) remains to form a scaffold for assembly of a second transcription complex (32). Taken together, these results indicate that an active RNAP II PIC forms on R199G and that this complex contains the core RNAP II subunit and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TFIIS. Analogous to what is observed for DNA promoters, our results also suggest that a re-initiation complex remains on the RNA promoter following initiation of transcription.

The TATA box-binding protein binds directly to R199G

In DNA-directed RNAP II transcription, the binding of TFIID to the DNA promoter is the first step in the assembly of an active PIC (43). Because our results suggested that RNAP II is forming a PIC on R199G analogous to that observed on DNA promoters, we investigated the binding of TFIID to R199G using EMSA experiments. Radiolabeled R199G was allowed to bind to purified TFIID, and the samples were subjected to nondenaturing PAGE. Purified TFIID bound directly to R199G, as illustrated by the retardation of the migration of the complex when compared to free RNA (Figure 4A). To test the specificity of this interaction, the binding of TFIID to R199G was challenged by competition with the various competitors used above. The 50-fold molar excess of P11.60 was not able to compete for TFIID binding to R199G. However, either 50-fold molar excess amounts of nonradiolabeled R199G or the control DNA promoter fragment greatly inhibited the formation of the R199G–TFIID complex (Figure 4A). Noteworthy, when the migration of the R199G–TFIID complex was compared to the complexes formed between R199G and the RNAP II holoenzyme, we found that the migration of R199G–TFIID corresponded to the fast-migrating species (Figure 4B, C1), suggesting formation of the R199G–TFIID prior to the assembly of the RNAP II complex (Figure 4B, C2). In addition, TFIID was able to bind to R38G, suggesting that TFIID was binding to R199G close to the initiation site, as observed above with RNAP II (Figure 4C).

It is well established that TFIID interacts with DNA promoters through its TBP subunit (43); it is therefore possible that TBP serves a similar role with respect to R199G. To test this hypothesis, the formation of a complex between purified GST-tagged TBP and either R199G or R38G were analyzed by EMSA. A typical gel showing a titration of a constant amount of radiolabeled R199G with increasing amounts of purified GST–TBP is shown in Figure 4D. Purified GST–TBP bound directly to R199G only when high amount was used (i.e. 64 ng), suggesting a very low affinity when it is not part of the TFIID complex. Such low affinity was also reported between purified TBP and DNA promoters (44). Despite this low affinity, the binding of GST–TBP to R199G was shown to be specific when the interaction was challenged with various competitors. A 50-fold molar excess of P11.60 was not able to compete for GST–TBP binding to R199G. Excess of nonradiolabeled R199G or the control DNA promoter fragment inhibited the binding of GST–TBP to radiolabeled R199G (Figure 4E). Moreover, purified GST protein alone was not able to bind to R199G (data not shown) and GST–TBP was able to bind to R38G (Figure 4F). In correlation with the binding efficiency of both mutants to the RNAP II holoenzyme, the binding of GST–TBP to R38GPU and R38GSW were found to be reduced to 36 and 20%, respectively (Figure 4F). Taken together, these results provide direct evidence that TBP interacts specifically and directly with R199G close to the initiation site of the transcription. Although we cannot exclude the possibility that other TFIID component(s) might also bind directly to the RNA promoter, our results suggest that this part of TFIID is involved in the nucleation of the RNAP II PIC on R199G, similar to what occurs on DNA promoters.

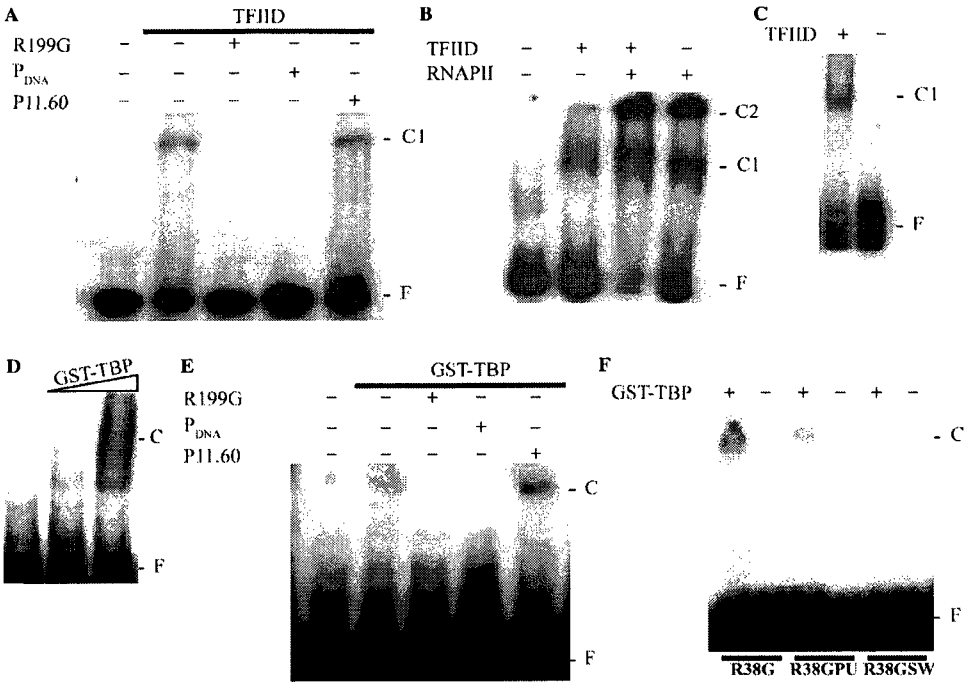


Figure 4. Specific binding of purified TFIID and GST-tagged TBP to HDV-derived RNAs. (**A**) EMSA of R199G in the presence of purified TFIID with 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G), a control DNA promoter fragment (i.e. CMV promoter) or P11.60. (**B**) EMSA of R199G in the presence purified TFIID or RNAP II. (**C**) EMSA of R38G with purified TFIID in the presence of 50-fold molar excess of P11.60. (**D**) Titration of a constant amount of radiolabeled R199G with increasing amounts of purified GST-tagged TBP (0, 32 and 64 ng). (**E**) EMSA of R199G bound to purified GST–TBP in the presence of 50-fold molar excess of either unlabeled homologous competitor RNA (i.e. R199G), a control DNA promoter fragment (i.e. CMV promoter) or P11.60. (**F**) Binding of R38G and two mutants with purified GST–TBP in the presence of 50-fold molar excess of P11.60.

DISCUSSION

RNAP II has been proposed to be involved in RNA-directed RNA polymerization of both HDV and many viroids (i.e. pospiviroids), principally based on the sensitivity of their replication/transcription to α-amanitin (21–23,45). In addition, both tomato and human RNAP II were reported to specifically coimmunoprecipitate with citrus exocortis viroid (CEVd) and HDV RNA genomes, respectively (24,46). However, these findings did not reveal either the direct interaction of RNAP II or the involvement of RNAP II transcription factors in RNA promoter recognition. To obtain insight into the mechanism of RNA promoter recognition by RNAP II, we used a transcription assay with HeLa NE proteins to demonstrate that an RNA fragment derived from the right terminal stem-loop domain of the genomic polarity of HDV acts as an RNA promoter, and that the initiation of synthesis occurred near the proposed initiation site for HDAg mRNA (18). The confirmation of RNAP II involvement was achieved by the demonstration of direct binding of the purified human RNAP II holoenzyme and the inhibition of transcription by a monoclonal antibody specific for the CTD of RNAP II.

Our results are in accordance with a previous report showing initiation of transcription from a similar HDV-derived RNA (39). In both cases, HDAg-S was not required for the reaction, and the size of the RNA products suggested that the initiation of synthesis occurred near the proposed initiation site for HDAg mRNA (i.e. position 1630; 18). Interestingly, a different transcription mechanism was reported using an RNA template derived from the left terminal stem-loop domain of the antigenomic polarity of HDV (21). In that case, the transcription reaction yielded a chimeric template/transcript product, and the reaction stopped after an elongation of only 41 nt. Addition of HDAg-S allowed the elongation to resume by binding to RNAP II directly and by displacing the elongation repressors NELF and DSIF (15,21,47). Although not tested because it was not required in our system, it is possible that the addition of HDAg-S might stimulate RNA synthesis from R199G.

Using RNA affinity chromatography, we demonstrated that RNAP II, along with general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TFIIS, binds to the RNA promoter to form a PIC similar to those observed on canonical DNA promoters (32). Moreover, this PIC was active, since addition of free NTPs resulted in production of the transcript and the release of the large subunit of RNAP II, along with both the TFIIF and TFIIB subunits. Similar to what was reported for a DNA promoter (32), our results suggest that TFIIA, TFIID and TFIIE might be part of a scaffold for a re-initiation complex. However, although the level of RNAP II, TFIIF and TFIIB were reduced,

they were not reduced to the extent reported for a DNA promoter (32). This is consistent with our observation that *in vitro* transcription initiation is reduced on such an RNA template as compared to a DNA template (i.e. it is a very weak promoter). In addition, although not required for transcription in our system using HeLa NE, additional host proteins, or the involvement of HDAg, might be required for efficient transcription initiation and/or elongation *in vivo*.

In mammalian RNAP II transcription, the formation of the TBP/DNA promoter complex is the first step in the assembly of a set of transcription factors necessary for the initiation of transcription (48). Following this binding, RNAP II and other general transcription factors are recruited to form a PIC (49). By analyzing the formation of a complex between R199G and RNAP II holoenzyme, we have observed that a fast-migrating species first formed at low concentrations of RNAP II, followed by a slower migrating species formed at higher concentrations of the enzyme. When the migration of the R199G–TFIID complex was compared to those formed between R199G and the RNAP II holoenzyme, the migration of R199G–TFIID corresponded to the fast-migrating species. Accordingly, our results suggest that formation of the R199G–TFIID complex is required to nucleate the RNAP II PIC. Because we found that purified GST–TBP binds directly and specifically to R199G, it is tempting to suggest that RNA promoter recognition by RNAP II occurs primarily though TBP.

Typically, DNA promoters recognized by RNAP II have several sequence motifs, each having specific functions that relate to the transcription process (28). Despite clear association of TBP (or TFIID) with R199G, no consensus sequence similar to the TATA box was found. However, TBP binds to canonical DNA target sites through the minor groove (50). Proteins that bind the DNA double helix through the minor groove must operate mainly by indirect readout since the symmetric positioning of donor and acceptor groups in this groove makes it difficult to differentiate an A–T base pair from a T–A pair (and likewise for G–C and C–G base pairs; 51). It is proposed that the structure of the DNA promoter region might produce intrinsic structural characteristics favoring TBP binding (52). Since TBP induces significant bending at the TATA box, sequences that are already bent, or are more flexible in a particular orientation could be energetically favored. Such a phenomenon was reported to occur in DNA templates constructed with arbitrary nonpromoter sequences and small single-stranded bubbles (53). In this case, core RNAP proteins were capable of directing RNA synthesis. Rather than being a major contributor to core promoter recognition, the Inr-like motifs found near the transcription initiation site on R199G might therefore represent preferred initiation sites for RNAP II after binding at its initial interaction site.

We recently reported that RNAP II interacts with HDV-derived RNAs at sites located within the terminal stem-loop domains of both polarities of HDV RNA (24). Analysis by base pair covariation and site-directed mutagenesis of this region revealed a strong selection to maintain a hairpin conformation (24,39,54). Wheat germ RNAP II was shown to bind to the large terminal loops of potato spindle tuber viroid genome (55), and it was recently reported that potato NE was able to initiate transcription of this viroid within a terminal stem-loop region of the RNA genome (8). Thus, it seems that a defined hairpin domain is a common feature of RNA promoters for RNAP II. Based on the conserved features of HDV RNA domains interacting with RNAP II, which are located near large terminal loops and present bulged initiation sites within a thermodynamically unstable stem (i.e. few G–C base pairs; 24), those RNA regions might mimic a prebent conformation giving those RNA molecules a general affinity for the TBP-containing TFIID. Since much of the affinity of TBP for DNA templates is mediated by electrostatic interactions with the phosphate backbone (56), an RNA template could preserve these interactions.

Further investigation is required to determine the significance of the association of TBP (and probably TFIID) with RNA promoters and to identify the RNA features involved in PIC formation. More importantly, our findings propose potential evolutionary relationships between DNA- and RNA-dependent RNAPs. Despite the unusual nature of the HDV-derived RNA promoter, our results demonstrate that an RNA template uses the same RNAP II general transcription factors in promoter recognition. Crystal structures of RNA-dependent RNAPs of positive-strand RNA and dsRNA viruses show structural similarity not only to each other, but also to DNA-dependent RNA and DNA polymerases, and reverse transcriptases (57–63). All of these polymerases share a structure similar to that of a right hand with palm, thumb and finger domains. The palm domain structure is particularly conserved and contains four sequence motifs preserved in all RNA and DNA polymerases (64,65). Because these features have been conserved throughout evolution, it is not excluded to find reminiscent RNA promoters in cellular RNAs, which therefore, would have the potential for initiation of transcription from RNA rather than DNA templates.

ACKNOWLEDGEMENTS

This work was funded by a grant from The Canadian Institutes of Health Research (CIHR) awarded to M.P. A.A. is supported by a doctoral award from the Libyan government. Funding to pay the Open Access publication charges for this article was provided by CIHR.

Conflict of interest statement. None declared.

REFERENCES

1. Konarska,M.M. and Sharp,P.A. (1989) Replication of RNA by the DNA-dependent RNA polymerase of phage T7. *Cell*, **57**, 423–431.
2. Biebricher,C.K. and Luce,R. (1996) Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase. *EMBO J.*, **15**, 3458–3465.
3. Biebricher,C.K. and Orgel,L.E. (1973) An RNA that multiplies indefinitely with DNA-dependent RNA polymerase: selection from a random copolymer. *Proc. Natl Acad. Sci. USA*, **70**, 934–938.

4. Wettich,A. and Biebricher,C.K. (2001) RNA species that replicate with DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry*, **40**, 3308–3315.
5. Pelchat,M., Grenier,C. and Perreault,J.P. (2002) Characterization of a viroid-derived RNA promoter for the DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry*, **41**, 6561–6571.
6. Pelchat,M. and Perreault,J.P. (2004) Binding site of *Escherichia coli* RNA polymerase to an RNA promoter. *Biochem. Biophys. Res. Commun.*, **319**, 636–642.
7. Wassarman,K.M. and Saecker,R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, **314**, 1601–1603.
8. Kolonko,N., Bannach,O., Aschermann,K., Hu,K.H., Moors,M., Schmitz,M., Steger,G. and Riesner,D. (2006) Transcription of potato spindle tuber viroid by RNA polymerase II starts in the left terminal loop. *Virology*, **347**, 392–404.
9. Flores,R., Daros,J.A. and Hernandez,C. (2000) *Avsunviroidae* family: viroids containing hammerhead ribozymes. *Adv. Virus Res.*, **55**, 271–323.
10. Taylor,J.M. (2006) Hepatitis delta virus. *Virology*, **344**, 71–76.
11. Wang,K.S., Choo,Q.L., Weiner,A.J., Ou,J.H., Najarian,R.C., Thayer,R.M., Mullenbach,G.T., Denniston,K.J., Gerin,J.L. and Houghton,M. (1986) Structure, sequence and expression of the hepatitis *delta* (*delta*) viral genome. *Nature*, **323**, 508–514.
12. Chang,F.L., Chen,P.J., Tu,S.J., Wang,C.J. and Chen,D.S. (1991) The large form of hepatitis *delta* antigen is crucial for assembly of hepatitis *delta* virus. *Proc. Natl Acad. Sci. USA*, **88**, 8490–8494.
13. Casey,J.L., Bergmann,K.F., Brown,T.L. and Gerin,J.L. (1992) Structural requirements for RNA editing in hepatitis *delta* virus: evidence for a uridine-to-cytidine editing mechanism. *Proc. Natl Acad. Sci. USA*, **89**, 7149–7153.
14. Polson,A.G., Bass,B.L. and Casey,J.L. (1996) RNA editing of hepatitis *delta* virus antigenome by dsRNA-adenosine deaminase. *Nature*, **380**, 454–456.
15. Yamaguchi,Y., Filipovska,J., Yano,K., Furuya,A., Inukai,N., Narita,T., Wada,T., Sugimoto,S., Konarska,M.M. and Handa,H. (2001) Stimulation of RNA polymerase II elongation by hepatitis *delta* antigen. *Science*, **293**, 124–127.
16. Sato,S., Cornillez-Ty,C. and Lazinski,D.W. (2004) By inhibiting replication, the large hepatitis *delta* antigen can indirectly regulate amber/W editing and its own expression. *J. Virol.*, **78**, 8120–8134.
17. Kos,A., Dijkema,R., Arnberg,A.C., van der Meide,P.H. and Schellekens,H. (1986) The hepatitis *delta* (*delta*) virus possesses a circular RNA. *Nature*, **323**, 558–560.
18. Gudima,S., Wu,S.Y., Chiang,C.M., Moraleda,G. and Taylor,J. (2000) Origin of hepatitis *delta* virus mRNA. *J. Virol.*, **74**, 7204–7210.
19. Nie,X., Chang,J. and Taylor,J.M. (2004) Alternative processing of hepatitis *delta* virus antigenomic RNA transcripts. *J. Virol.*, **78**, 4517–4524.
20. de Mercoyrol,L., Job,C. and Job,D. (1989) Studies on the inhibition by alpha-amanitin of single-step addition reactions and productive RNA synthesis catalysed by wheat-germ RNA polymerase II. *Biochem. J.*, **258**, 165–169.
21. Filipovska,J. and Konarska,M.M. (2000) Specific HDV RNA-templated transcription by pol II *in vitro*. *RNA*, **6**, 41–54.
22. Fu,T.B. and Taylor,J. (1993) The RNAs of hepatitis *delta* virus are copied by RNA polymerase II in nuclear homogenates. *J. Virol.*, **67**, 6965–6972.
23. MacNaughton,T.B., Gowans,E.J., McNamara,S.P. and Burrell,C.J. (1991) Hepatitis *delta* antigen is necessary for access of hepatitis *delta* virus RNA to the cell transcriptional machinery but is not part of the transcriptional complex. *Virology*, **184**, 387–390.
24. Greco-Stewart,V.S., Miron,P., Abrahem,A. and Pelchat,M. (2007) The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis *delta* virus RNA genome. *Virology*, **357**, 68–78.
25. Macnaughton,T.B., Shi,S.T., Modahl,L.E. and Lai,M.M. (2002) Rolling circle replication of hepatitis *delta* virus RNA is carried out by two different cellular RNA polymerases. *J. Virol.*, **76**, 3920–3927.
26. Modahl,L.E., Macnaughton,T.B., Zhu,N., Johnson,D.L. and Lai,M.M. (2000) RNA-Dependent replication and transcription of hepatitis *delta* virus RNA involve distinct cellular RNA polymerases. *Mol. Cell. Biol.*, **20**, 6030–6039.
27. Li,Y.J., Macnaughton,T., Gao,L. and Lai,M.M. (2006) RNA-templated replication of hepatitis *delta* virus: genomic and antigenomic RNAs associate with different nuclear bodies. *J. Virol.*, **80**, 6478–6486.
28. Smale,S.T. and Kadonaga,J.T. (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.*, **72**, 449–479.
29. Sims,R.J. III, Mandal,S.S. and Reinberg,D. (2004) Recent highlights of RNA-polymerase-II-mediated transcription. *Curr. Opin. Cell. Biol.*, **16**, 263–271.
30. Cadena,D.L. and Dahmus,M.E. (1987) Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.*, **262**, 12468–12474.
31. Orphanides,G., Lagrange,T. and Reinberg,D. (1996) The general transcription factors of RNA polymerase II. *Genes. Dev.*, **10**, 2657–2683.
32. Yudkovsky,N., Ranish,J.A. and Hahn,S. (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature*, **408**, 225–229.
33. Proudfoot,N.J., Furger,A. and Dye,M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.
34. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
35. Burke,T.W. and Kadonaga,J.T. (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.*, **10**, 711–724.
36. Deng,W. and Roberts,S.G. (2005) A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.*, **19**, 2418–2423.
37. Kuo,M.Y., Sharmeen,L., Dinter-Gottlieb,G. and Taylor,J. (1988) Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis *delta* virus. *J. Virol.*, **62**, 4439–4444.
38. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
39. Beard,M.R., MacNaughton,T.B. and Gowans,E.J. (1996) Identification and characterization of a hepatitis *delta* virus RNA transcriptional promoter. *J. Virol.*, **70**, 4986–4995.
40. Carroll,S.B. and Stollar,B.D. (1982) Inhibitory monoclonal antibody to calf thymus RNA polymerase II blocks formation of enzyme-DNA complexes. *Proc. Natl Acad. Sci. USA*, **79**, 7233–7237.
41. Vioque,A. and Altman,S. (1986) Affinity chromatography with an immobilized RNA enzyme. *Proc. Natl Acad. Sci. USA*, **83**, 5904–5908.
42. Bichko,V.V. and Taylor,J.M. (1996) Redistribution of the *delta* antigens in cells replicating the genome of hepatitis *delta* virus. *J. Virol.*, **70**, 8064–8070.
43. Verrijzer,C.P., Chen,J.L., Yokomori,K. and Tjian,R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115–1125.
44. Zhao,X. and Herr,W. (2002) A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. *Cell*, **108**, 615–627.
45. Muhlbach,H.P. and Sanger,H.L. (1979) Viroid replication is inhibited by alpha-amanitin. *Nature*, **278**, 185–188.
46. Warrilow,D. and Symons,R.H. (1999) Citrus exocortis viroid RNA is associated with the largest subunit of RNA polymerase II in tomato *in vivo*. *Arch. Virol.*, **144**, 2367–2375.
47. Yamaguchi,Y., Delehouzee,S. and Handa,H. (2002) HIV and hepatitis *delta* virus: evolution takes different paths to relieve blocks in transcriptional elongation. *Microbes Infect.*, **4**, 1169–1175.
48. Faiger,H., Ivanchenko,M., Cohen,I. and Haran,T.E. (2006) TBP flanking sequences: asymmetry of binding, long-range effects and consensus sequences. *Nucleic Acids Res.*, **34**, 104–119.
49. Verrijzer,C.P. and Tjian,R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.*, **21**, 338–342.
50. Starr,D.B. and Hawley,D.K. (1991) TFIID binds in the minor groove of the TATA box. *Cell*, **67**, 1231–1240.
51. Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.

52. Khrapunov,S. and Brenowitz,M. (2004) Comparison of the effect of water release on the interaction of the Saccharomyces cerevisiae TATA binding protein (TBP) with "TATA Box" sequences composed of adenosine or inosine. *Biophys. J.*, **86**, 371–383.
53. Helmann,J.D. and deHaseth,P.L. (1999) Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry*, **38**, 5959–5967.
54. Gudima,S., Dingle,K., Wu,T.T., Moraleda,G. and Taylor,J. (1999) Characterization of the 5' ends for polyadenylated RNAs synthesized during the replication of hepatitis *delta* virus. *J. Virol.*, **73**, 6533–6539.
55. Goodman,T.C., Nagel,L., Rappold,W., Klotz,G. and Riesner,D. (1984) Viroid replication: equilibrium association constant and comparative activity measurements for the viroid-polymerase interaction. *Nucleic Acids Res.*, **12**, 6231–6246.
56. Bewley,C.A., Gronenborn,A.M. and Clore,G.M. (1998) Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 105–131.
57. Hansen,J.L., Long,A.M. and Schultz,S.C. (1997) Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*, **5**, 1109–1122.
58. Bressanelli,S., Tomei,L., Roussel,A., Incitti,I., Vitale,R.L., Mathieu,M., De Francesco,R. and Rey,F.A. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl Acad. Sci. USA*, **96**, 13034–13039.
59. Lesburg,C.A., Cable,M.B., Ferrari,E., Hong,Z., Mannarino,A.F. and Weber,P.C. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat. Struct. Biol.*, **6**, 937–943.
60. Ago,H., Adachi,T., Yoshida,A., Yamamoto,M., Habuka,N., Yatsunami,K. and Miyano,M. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure*, **7**, 1417–1426.
61. Butcher,S.J., Grimes,J.M., Makeyev,E.V., Bamford,D.H. and Stuart,D.I. (2001) A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, **410**, 235–240.
62. Ng,K.K., Cherney,M.M., Vazquez,A.L., Machin,A., Alonso,J.M., Parra,F. and James,M.N. (2002) Crystal structures of active and inactive conformations of a caliciviral RNA-dependent RNA polymerase. *J. Biol. Chem.*, **277**, 1381–1387.
63. O'Reilly,E.K. and Kao,C.C. (1998) Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology*, **252**, 287–303.
64. Poch,O., Sauvaget,I., Delarue,M. and Tordo,N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.*, **8**, 3867–3874.
65. Xiong,Y. and Eickbush,T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.

Appendix IV: CURRICULUM VITAE

A. F. Abrahem, BSc, MSc. PhD Candidate

Education:

1. PhD Candidate (2004-2009) University of Ottawa, Ottawa, Ontario, Canada
2. MSc Parasitology (1999). Institute of Parasitology, McGill University, Montreal, Quebec, Canada
3. BSc Medical Lab Technology (1989) Higher institute of technology, Brake, Libya

Certificates:

1. Biosafety Workshop 2007, University of Ottawa
2. Sexual Harassment training course 2005, University of Ottawa
3. Radiation safety training course 2004, University of Ottawa
4. Laboratory safety workshop 2004, University of Ottawa
5. Radiation safety course 1996, McGill university

Scholarships and Fellowships:

- (2004-2009) Graduate Scholarship, Ministry of Higher Education, Libya
- (1995-1999) Graduate Scholarship, Ministry of Higher Education, Libya

Distinction Prizes, Honors and Awards:

- 23/1/2004 Atawen Sport and Social Activity Club Award, Agedabia, Libya

- 4/12/2003 Libyan Students Union Award, Tripoli, Libya
- 12/6/2003 Alfaek Institute of Computer Sciences Award, Agedabia, Libya
- 16/6/2003 Alfateh Center for Skill Development Award, Agedabia, Libya
- 17/4/2003 Students Union Award, Agedabia University, Agedabia, Libya
- 30/3/2000 Higher Institute of Medical sciences Teaching Award, Agedabia, Libya
- 26/12/1999 Sirte Oil Company Service Award, Braiga, Libya

Work experience:

1. (2001-2004) Member, Management Committee, Agedabia hospital (360 beds), Agedabia, Libya
2. (2001-2004) Medical Laboratory Supervisor, Agedabia Red Crescent, Agedabia, Libya
3. (2001-2004) Medical Laboratories supervisor, Province of Agedabia, including Agedabia hospital, Agedabia, Libya
4. (2001-2003) Hematology instructor, Higher Institute of Health Sciences, Agedabia, Libya
5. (2001-2003) Molecular Biology instructor at Agedabia University Agedabia, Libya
6. (1999-2001) Medical Laboratory Specialist, Sirte Family Clinic, Serte, Libya
7. (1996-1999) MSc student, The Institute of Parasitology, McGill University, Montreal, Quebec, Canada
8. (1993-1995) Hematology instructor, Agedabia Higher Institute of Health Sciences, Agedabia, Libya
9. (1989-1995) Medical Laboratory specialist, Sirte Family Clinic, Sirte, Libya

Other Skills:

- Proficient computer skills including power point, Word, Word Perfect, Sigma Blot, Oligo, and DNA star, Macvictor, Adobe CS3 and other web based molecular biology programs.

Publications:

1. **Abrahem A., Pelchat M.** Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis delta virus RNA genome. Nucleic Acids Res., Advance Access published on August 5, 2008
2. Greco-Stewert VS, Miron P, **Abrahem A**, Pelchat M. The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. Virology. 5; 357(1):68-78. (2007)
3. **Abrahem Abrahem**, Gabriela Ceritad, Xing-Qing Pan and Georges. Pleiotropic Resistance to Diverse antimalarials in Actinomycin D-Resistant *Plasmodium falciparum*. Biochemical Pharmacology, Vol. 59, pp.1123-1132 (2000).
4. Gabriela Ceritad, **Abrahem Abrahem** and Elias Georges. Cloning and Partial Characterization of the Proteasome S4 ATPase from *Plasmodium falciparum*. Experimental Parasitology 93, 123-133 (1999).

Conferences:

1. Ribclub sherbrooke university, (2004)
2. BMI department poster day (2006)
3. BMI department symposium (2007)
4. RNA club at university of Ottawa (2008)
6. CSBMCB at Banff (2008)
7. HBV meeting Paris, France (2009)
8. The ABC of mass spectrometry for biology workshop (2010)

Hobbies:

Soccer and Hunting.

References:

1. Dr. Martin Pelchat, PhD.
   Department of Biochemistry, Microbiology and Immunology
   Faculté de médecine / Faculty of Medicine
   Université d'Ottawa / University of Ottawa

2. Dr. Earl Brown Professor
   Department of Biochemistry, Microbiology and Immunology
   Faculté de médecine / Faculty of Medicine
   Université d'Ottawa / University of Ottawa

3. A. Mhanni, MD, FRCPC, FCCMG, FACMG, PhD Medical Geneticist and Assistant Professor Department of Pediatrics and Child Health & Biochemistry and Medical Genetics University of Manitoba Genetics and Metabolism Program,

I would like to have your permission to use the FIGURE 8. (Asimplified model for AG HDV RNA-templated transcription *in vitro*) that is in your paper titled Specific HDV RNA-templated transcription by pol II in vitro, for my thesis writing.

Thank you

Abrahem Abrahem

B.Sc., M.Sc., Ph.D. Candidate

Faculty of Medicine- BMI
University of Ottawa
Dr. Martin Pelchat Lab

RE: Permission to use a figure

From:**CDC-INFO**

Sent: Monday, January 11, 2010 5:15:36 PM

Thank you for your inquiry to CDC-INFO. We are pleased to provide you with the following information.

General text information, publications available for download, and charts developed by CDC and presented on CDC's website are works of the United States Government and in the public domain, which means that they are meant for public use and are not subject to copyright law protections. Permission is not required for use of public domain items, but CDC does ask that you credit the original institution and contributor, when known, whenever the item is used in any publicly distributed media.

You are also free to adapt and revise these materials, provided the information is distributed free of cost; however, you must remove the CDC name and logo if changes are made. Additionally, in accordance with 42 U.S.C. Section 1320b-10, no person may, for a fee, reproduce, reprint, or distribute any item consisting of a form, application, or other publication of the Department of Health and Human Services unless such person has obtained specific, written authorization for such activity. Therefore, if you wish to sell the CDC materials presented on CDC's website, you must first obtain

permission to do so from CDC.

You may also encounter information on the website that is owned/created by others, including copyrighted materials and other materials sponsored by private companies or non-government organizations. Those other parties retain all rights to publish or reproduce those documents or to allow others to do so, in accordance with 17 U.S.C. Sections 106-120. Any copyrighted materials included on this site were used with the permission of the copyright holder and are not in the public domain; such materials remain the property of their respective owners/creators and should not be reproduced or otherwise used.

Images, illustrations, assessment tools, and applications on CDC web pages may be copyright protected and require permission from the content provider/web page owner. Use the "contact us" feature on the relevant web page to request permission and obtain the rules for use of the item.

Thank you for contacting CDC-INFO Contact Center. Please call 1-800-CDC-INFO, e-mail or visit http://www.cdc.gov if you have any additional questions.

CDC-INFO is a service of the Centers for Disease Control and Prevention (CDC) and the Agency for Toxic Substances and Disease Registry (ATSDR). This service is provided by Vangent, Inc. under contract to CDC and ATSDR.

Please visit CDC-INFO's Customer Satisfaction Survey and tell us about your recent experience with CDC-INFO. The survey results will help improve our services. The address for the survey is http://www.emtservices.org/cdcinfo/en/email_satisfaction_survey.htm. Thank you.

AM/5457/vv

[THREAD ID:1-1NHYUO] [SR No.:1-99940888]

Hi,

I would like to have your permission to use the Geographic distribution of HDV infection map (figure) in my thesis.

Thank you

Abrahem Abrahem

B.Sc., M.Sc., Ph.D. Candidate

Faculty of Medicine- BMI
University of Ottawa
Dr. Martin Pelchat Lab