INVOLVEMENT OF THE POLYPYRIMIDINE TRACT-BINDING PROTEIN-ASSOCIATED SPLICING FACTOR (PSF) IN THE HEPATITIS DELTA VIRUS (HDV) RNA-TEMPLATED TRANSCRIPTION

Da Jiang Zhang

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the degree of Masters in Biochemistry

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

University of Ottawa

Ottawa, Ontario, Canada

Abstract

Hepatitis delta virus (HDV) is the smallest known mammalian RNA virus, containing a genome of ~ 1700 nt. Replication of HDV is extremely dependent on the host transcription machinery. Previous studies indicated that RNA polymerase II (RNAPII) directly binds to and forms an active preinitiation complex on the right terminal stem-loop fragment (R199G) of HDV genomic RNA, and that the polypyrimidine tract-binding protein-associated splicing factor (PSF) directly binds to the same region. Further studies demonstrated that PSF also binds to the carboxyl-terminal domain (CTD) of RNAP II. In my thesis, co-immunoprecipitation assays were performed to show that PSF stimulates the interaction of RNAPII with R199G. Results of co-immunoprecipitation experiments also suggest that both the RNA recognition motif 2 (RRM2) and N-terminal proline-rich region (PRR) of PSF are required for the interaction between PSF and RNAPII, while the two RNA recognition motifs (RRM1 and RRM2) might be required for the interaction of PSF with R199G. Furthermore, *in vitro* run-off transcription assays suggest that PSF facilitates the HDV RNA transcription from the R199G template. Together, the above experiments suggest that PSF might act as a transcription factor for the RNAPII transcription of HDV RNA by linking the CTD of RNAPII and the HDV RNA promoter. My experiments provide a better understanding of the mechanism of HDV RNA-dependent transcription by RNAP II.

Acknowledgments

I would like to sincerely thank my thesis supervisor Dr. Martin Pelchat for the support of my M.Sc study and research, for his immense knowledge and patience. His mentorship helped me throughout the research and writing of my thesis. I would also like to thank my thesis advisory committee: Dr. Earl Brown and Dr. Lionel G. Filion for their help and advice throughout my thesis. I would also like to thank my thesis examiners: Dr. Marc-André Langlois and Dr. Jean-François Couture. Particularly, I would like to thank Dr. Dorota Sikora for helping me to edit my thesis.

I would like to thank my previous and current fellow lab members: Yasnee, Dorota, Youser, Ali, Alycia, Salma and Mary for the stimulating and fun discussions. In addition, I would like to thank previous members of Dr. Earl Brown: Jianjun Jia, Jihui Ping and Nicole E Forbes for their help and suggestion.

I would also like to sincerely thank my parents and my family for their ongoing support and encouragement.

Table of Contents

List of Abbreviations

(+/-) - highly charged basic and acidic residues rich region

ADAR-1 - Adenosine deaminase acting on RNA-1

ARMs - two arginine-rich motifs

ASF/SF2 - alternative splicing factor, also known as pre-mRNA-splicing factor SF2

BrUTP - 5-Bromouridine 5'-Triphosphate

BSA - bovine serum albumin

CCD - coiled-coil domain

Cdk9 - Cyclin-dependent kinase 9

cDNA - complementary DNA

CDS - coding sequence

CKII - Casein kinase II

c-RBD - cryptic RNA binding domain

CTD - carboxy-terminal domain

DBHS - drosophila behavior/human splicing family of proteins

DMEM - Dulbecco's modified eagle medium

DRB - 5,6-dichloro-1-4-D-ribofuranosyl benzimidazol

DSIF - DRB sensitivity inducing factor

DTT - dithiothreitol

E.coli - *Escherichia coli*

eEF1A1 - eukaryotic elongation factor 1A1

ERK1/2 - Extracellular signal-related kinases 1 and 2

FBS - fetal bovine serum

FHL2 - four-and-a-half LIM-only protein 2

FTase – Farnesyltransferase

GAPDH - glyceraldehydes 3-phosphate dehydrogenase

GST - glutathione S transferase

HBsAg - HBV surface antigen

HBV - Hepatitis B virus

HDAg - Hepatitis delta virus antigen

HDAg-L - large isoform of HDAg

HDAg-S - small isoform of HDAg

HDV - Hepatitis delta virus

HEK293 - Human Embryonic Kidney 293 cells

His6-PSF - hexahistidine–tagged PSF

His6 – hexahistidine tag

HLH - helix-loop-helix

hnRNP-L - heterogeneous nuclear ribonucleoprotein L

HRP - Horseradish Peroxidase

IgG H&L - Immunoglobulin G heavy chains and light chains

IP - immunoprecipitation

kb - Kilobase

kDa - Kilodalton

LB - Luria broth medium

L-HBsAg - large HBV surface antigen

M-HBsAg - middle HBV surface antigen

m^7G - 7-methylguanosine

mRNA - messenger RNA

ncRNA - noncoding RNA

NE - nuclear extract

NELF - negative elongation factor

NELF-A - subunit A of NELF

NESI - Nuclear export signal-interacting protein

NLS - nuclear localization signal

NOPS - NONA/paraspeckle domain

nt – nucleotide

ORF - open reading frame

PRR - proline rich region

P/Q - proline/glutamine rich region

p54nrb/NONO - 54kDa nuclear RNA-binding protein, also known as Non-POU domain-containing octamer-binding protein

PBS - phosphate buffered saline

PCR - polymerase chain reaction

Phe (F) - Phenylalanine

PIC - pre-initiation complex

PKC - Protein kinase C

PKR - Double-stranded RNA-activated kinase

PLMVd - peach latent mosaic viroid

PMSF - phenylmethylsulfonyl fluoride

poly (A) tail - a long chain of multiple adenosine monophosphates

Pro (P) - Proline

PSF/SFPQ - polypyrimidine tract- binding protein-associated splicing factor, also known as splicing factor proline / glutamine-rich

PSPC1 - paraspeckle component 1

p-TEFb - positive transcription elongation factor b

PVDF - polyvinylidene difluoride

qRT-PCR - quantitative reverse transcription polymerase chain reaction

Rbp1-12 - RNA binding proteins (1-12), subunits of RNA polymerase II

RBS - receptor binding site

RdRP - RNA-depended RNA polymerase

RGG - arginine glycine glycine repeat motif

RIPA - ribonucleoprotein immunoprecipitation assay

RNAP IIA - hypophosphorylated RNA polymerase II

RNAP IIO - hyperphosphorylated RNA polymerase II

RNAPI - RNA polymerase I

RNAPII - RNA polymerase II

RNAPIII - RNA polymerase III

RNase - ribonuclease

RNP - ribonucleoprotein

RRM - RNA recognition motif

RT - reverse transcription

SDS - sodium dodecyl sulfate

SDS-PAGE - sodium dodecyl sulfate polyacrylamide gel electrophoresis

SELEX - systematic evolution of ligands by exponential enrichment

S-HBsAg - small HBV surface antigen

shRNA - small hairpin RNA

siRNA - small interfering RNA

TBST - Tris-buffed saline containing 0.1% (v/v) Tween

TET – tetracycline

TFII – the RNAPII general transcription factor

Urea-PAGE - urea polyacrylamide gel electrophoresis

VAS - viral assembly signal

WCE - whole cell extract

wt – wild type

List of Figures

Chapter1. Introduction

1.1 History and epidemiology

Hepatitis delta virus (HDV) was firstly reported when a nuclear antigen, named 'delta', was identified in severe chronic hepatitis B patients (Rizzetto *et al*., 1977). Follow-up studies showed that this antigen was a protein produced by a pathogen requiring the hepatitis B virus as a helper virus for infection (Rizzetto *et al*., 1980). In 1986, cloning and sequencing of the entire virus was completed, and HDV obtained its own genus, the *Deltavirus* (Wang *et al*., 1986). HDV has a worldwide distribution. Currently, approximately 15 to 20 million people are infected with HDV (Pascarella and Negro, 2011). Eight genotypes of HDV have been discovered through restriction fragment length polymorphism analysis, fragment sequencing and full-length genome sequencing of reverse transcription polymerase chain reaction (RT-PCR) products of various HDV RNA strains isolated from sera of infected patients (Le Gal *et al*., 2006).

There are two major patterns of HDV infection: co-infection, in which HDV and HBV individually infect the host cell at the same time; super-infection, in which HDV infects the chronic HBV carrier (Taylor, 2003). Less than 5% of acute hepatitis Delta will turn into chronic hepatitis Delta when co-infection occurs, while the percentage of progression to chronicity is up to 80% in the case of super-infection (Pascarella and Negro, 2011). HDV usually induces severe hepatitis and most of chronic hepatitis D patients may develop cirrhosis; a long-term and high-dose administration of interferon-α can only improve symptoms of hepatitis D, but not eradicate HDV (Farci *et al*., 1994; Pascarella and Negro, 2011). Liver transplantation is the only effective treatment for

fulminant and end-stage chronic hepatitis D patients (Samuel *et al*., 1995).

1.2 Virion structure

HDV is the smallest mammalian virus pathogen known in the world. HDV virion is a roughly spherical particle approximately 36 nm in size (Hughes *et al*., 2011) (Figure 1.1). The outer envelope of the virion is made up of HBV surface antigen (HBsAg) proteins and host lipids, while the inner core is an about 19-nm diameter ribonucleoprotein (RNP) complex which contains a circular single strand of RNA genome and approximately 200 molecules of hepatitis delta antigen (HDAg) (Ryu *et al*., 1993; Gudima *et al*., 2002). The assembly of an HDV particle mainly involves the large isoform of HDAg (HDAg-L) and HBsAgs, though the small isoform of HDAg (HDAg-S) increases packaging efficiency (Chen *et al*., 1992; Wang *et al*., 1994).

The genome of HDV does not encode viral coat protein. HDV utilizes the HBsAg proteins, the HBV envelope proteins, for its encapsidation and transmission (reviewed in Sureau, 2006). There are three HBsAg proteins: the small HBsAg (S-HBsAg), the middle (M-HBsAg) and the large (L-HBsAg). These proteins are translated from three in-frame start codons to the same stop codon. The open reading frame (ORF) of HBsAg mRNA contains three contiguous regions from 5'-end to 3'-end: pre-S1, pre-S2 and S. The L-HBsAg is encoded by all three regions, the M-HBsAg is encoded by the pre-S2 and S regions, and the S-HBsAg is encoded by the S region (Ganem and Varmus, 1987; reviewed in Sureau, 2006). The relative proportion of three HBsAgs in the envelope of HDV virion is 95:5:1 (S:M:L) (Bonino *et al*., 1986). Studies using an *in vitro* culture system susceptible to HDV infection demonstrated that the S-HBsAg is sufficient for the

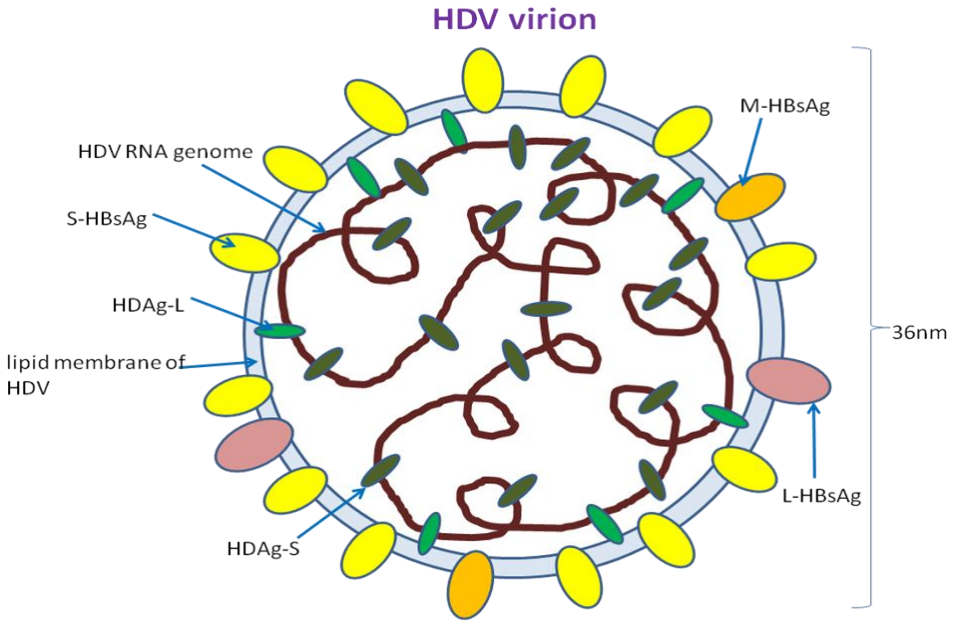

HDV virion
M-HBsAg
HDV RNA genome
S-HBsAg
HDAg-L
lipid membrane of HDV
36nm
L-HBsAg
HDAg-S

Figure 1.1: Schematic representation of the HDV virion.

The diameter of the HDV virion is approximately 36 nm. The inner HDV viral ribonucleoprotein complex is composed of a circular single-strand genomic RNA and HDAg proteins; the outer viral envelope is made up of host lipid membrane and HBV surface proteins (L-HBsAg, M-HBsAg and S-HBsAg). HDV RNA genome is shown in brown, HDAg-L proteins in green, HDAg-S proteins in dark-olive, S-HBsAg in yellow, M-HBsAg in goldenrod, L-HBsAg in lightcoral and viral lipid membrane in gray. (Hughes *et al*., 2011; Ryu *et al*., 1993; 241 Gudima *et al*., 2002; Chen *et al*., 1992; Wang *et al*., 1994)

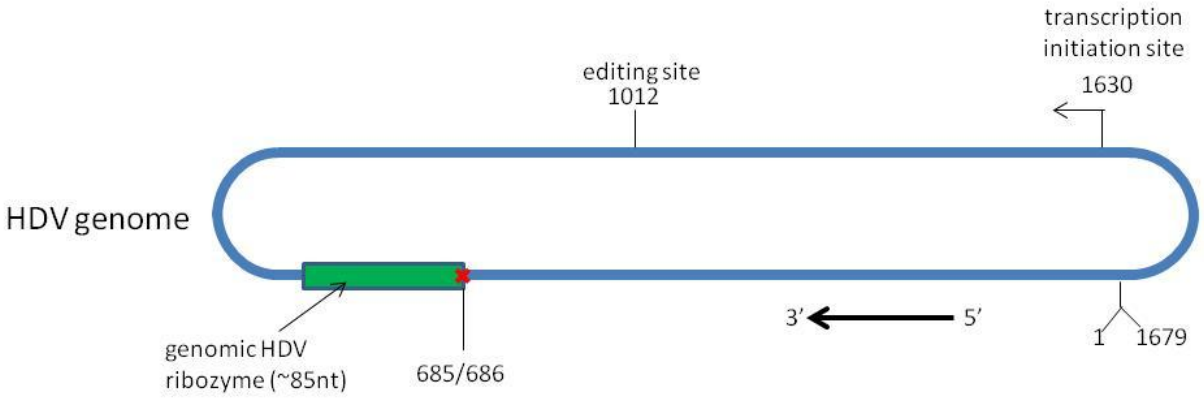
assembly of the HDV particles; however, the particles are non-infectious without the L-HBsAg (Sureau *et al*., 1993; Sureau *et al*., 1994). Thus, the S-HBsAg and the L-HBsAg ensure the assembly and the infectivity of the virions, respectively. On the other hand, M-HBsAg is not required for the HDV particle assembly and infectivity (Sureau *et al*., 1994).

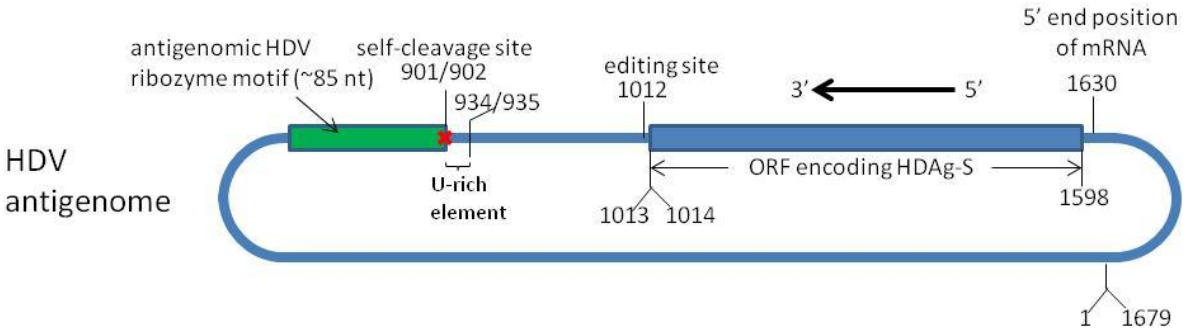
1.3 HDV genome, antigenome and mRNA

By definition, the HDV genome is the full length RNA included in the HDV virion. It is a circular negative-sense single-stranded RNA of approximately 1700 nucleotides (Taylor, 2006). On the other hand, there are many copies of the exact complement of the genomic RNA in cells replicating HDV. This complementary RNA is called the antigenome (Chen *et al*., 1986). The HDV antigenome is an RNA intermediate during the replication of the genome.

Both genome and antigenome fold into unbranched double-stranded rod-like structures due to 74% internal base pairing (Wang *et al*., 1986; Taylor *et al*., 2006) (Figure 1.2). Each of them contains a ribozyme domain of approximately 85 nucleotides, allowing the self-cleavage of HDV RNA (Ferre-D'Amare *et al*., 1998; Sharmeen *et al*., 1988). The two terminal stem-loop domains of both genome and antigenome contain putative RNA promoter elements (Abrahem and Pelchat 2008; Beard *et al*., 1996; Filipovska and Konarska 2000).

The antigenome contains a single open reading frame (ORF) encoding the HDAg. HDAg mRNA has the same polarity as the antigenome and is a linear RNA of ~ 800 nucleotides, containing a 5'-end 7-methylguanosine (m^7G) cap structure and a 3'-end


transcription
initiation site
1630
editing site
1012
HDV genome
3′
5′
1
1679
genomic HDV
ribozyme (~85nt)
685/686


5′ end position
of mRNA
1630
antigenomic HDV
ribozyme motif (~85 nt)
self-cleavage site
901/902
934/935
editing site
1012
3′
5′
HDV
antigenome
ORF encoding HDAg-S
1598
U-rich
element
1013 1014
1
1679

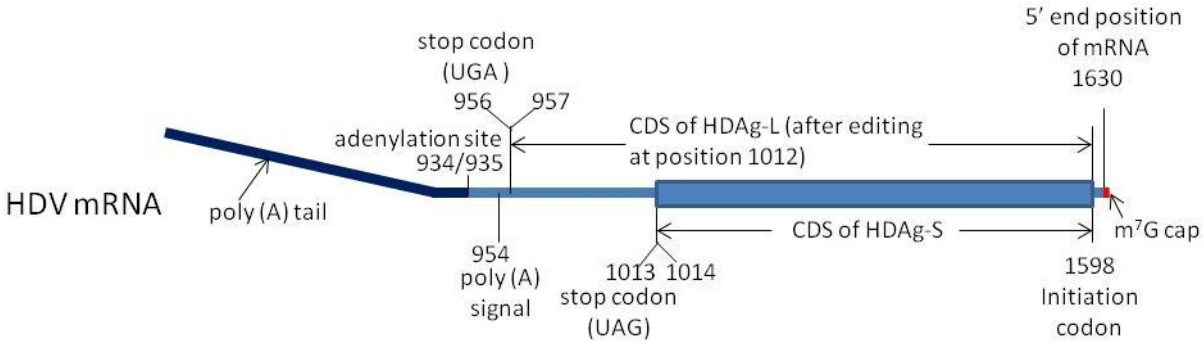

5′ end position
of mRNA
1630
stop codon
(UGA)
956
957
adenylation site
934/935
CDS of HDAg-L (after editing
at position 1012)
HDV mRNA
poly (A) tail
CDS of HDAg-S
m7G cap
954
poly (A)
signal
1013 1014
stop codon
(UAG)
1598
Initiation
codon

Figure 1.2: Schematic representation of HDV genome, antigenome and mRNA.

The HDV genome and its complementary antigenome are depicted in circular unbranched rod-like conformation. The HDV ribozyme motifs (~85 nt; shown in green box) and self-cleavage sites (shown in red X) are indicated The blue box of the HDV antigenome scheme indicates the ORF encoding HDAg-S. The ADAR-1 editing site of UAG to UIG at position 1012 of HDV antigenomic RNA is indicated. The 5'-end of HDAg mRNA corresponds to position 1630 of the antigenome. The corresponding transcription initiation site at position 1630 of the genome is indicated with an arrow. A U-rich element is located between position 902 and 934. The 5'-end m^7G cap is shown in red in the scheme of the HDV mRNA. The CDS of HDAg-S is from position 1598 to 1014. The CDS of HDAg-L is from 1598 to 957 after the first stop codon (UAG; position 1013 to 1011) is edited to the tryptophan codon (UGG). The second stop codon is the UGA (position 956 to 954). From 954 to 949 is the poly (A) signal, AAUAAA. A poly (A) tail of ~150 adenosines is added at the adenylation site of the mRNA, which is the CA sequence at position 935 to 934. Numbering is in accordance with (Kuo *et al*., 1988a). (Taylor, 2006; Ferre-D'Amare *et al*., 1998; Kuo *et al*., 1988b; Hsieh *et al*., 1990; Casey and Gerin, 1995; Gudima *et al*., 1999; Abrahem and Pelchat, 2008)

poly (A) tail of ~ 150 adenosines (Taylor, 2006). Consistently, the HDAg mRNA contains a poly (A) signal, AAUAAA, located downstream of the coding sequence (CDS) while the antigenome contains a U-rich element, which is necessary for adenylation, located downstream of the adenylation site (Nie *et al*., 2004; Figure 1.2).

During the HDV life cycle, approximately 300,000 copies of genomic RNA, 30,000 copies of antigenomic RNA and 600 copies of HDAg mRNA per cell are generated (Chen *et al*., 1986; Nie *et al*., 2004). Both the antigenome and mRNA of HDV are transcribed from the same template, the HDV genome (Taylor, 2006). Unlike HBV replication, which involves reverse transcription, HDV genome replication is dependent on an RNA template, without any DNA intermediates (Chen *et al*., 1986; Beck and Nassal 2007).

1.4 HDV antigen

HDV expresses two antigens: HDAg-S (195 amino acids, 24 kDa) and HDAg-L (214 amino acids, 27kDa) (Casey and Gerin 1995). HDAg-S is produced earlier during the infection while HDAg-L appears later. During HDV replication, an RNA-editing enzyme, the adenosine deaminase acting on RNA-1 (ADAR-1), converts the adenosine (nucleotide position 1012; see Figure 1.2) at the amber termination codon (UAG) of some antigenomic RNAs into an inosine (I). Following the next round of replication of antigenome to genome, the AUC anticodon in the genome is replaced by ACC. Therefore, the mRNA transcribed from this modified genome contains a tryptophan codon (UGG) instead of UAG stop codon in the HDAg-S mRNA. The translation of this mRNA terminates at the next stop codon (UGA) and generates the HDAg-L, which is 19 amino acids longer than the HDAg-S at the C-terminus (Casey and Gerin 1995).

HDAg-S is necessary for HDV replication, while HDAg-L inhibits replication but is essential for the assembly of the HDV particle (Lai, 1995). The two isoforms of HDAg contain several common functional regions except the C-terminal 19 residues of HDAg-L (reviewed in Sikora, 2012). Figure 1.3 shows the structural arrangement of HDAg-S and HDAg-L. A leucine-rich coiled-coil domain spanning amino acid residues 12 – 60 in HDAg's N-terminus is required for HDAg oligomerization (Rozzelle *et al*., 1995; Zuccola *et al*., 1998), which is necessary for trans-activation of HDAg-S and trans-dominant inhibition activities of HDAg-L during HDV transcription (Xia and Lai, 1992). A nuclear localization signal (NLS) is found between residues 66 – 77, allowing the HDV RNP complex to enter the nucleus (Alves *et al*., 2008). A cryptic RNA binding domain (RBD) is located between residues 2 – 27 (Poisson *et al*., 1993), mediating nuclear import of HDV RNA and also acting as a RNA chaperone to modulate the ribozyme activity of HDV (Huang and Wu 1998; Wang *et al*., 2003; Chou *et al*., 1998). Two conserved arginine-rich motifs (ARM), KERQDHRRRKA and EDEKRERRIAG (residues 97 – 107 and 136 – 146, respectively), also bind HDV RNA (Lee *et al*., 1993). These interactions between HDAg and HDV RNA are caused by electrostatic forces between highly basic amino acid clusters and RNA (Wang *et al*., 2003). These interactions are required for viral RNA packaging (Wang *et al*., 1994) and nuclear import (Chou *et al.,* 1998). The spacer between the two ARMs contains a helix-loop-helix (HLH) motif (residues 105 – 138), which is also important for RNA binding (Lee *et al*., 1993; Chang *et al*., 1995). In addition to the common motifs owned by both small and large antigens, HDAg-L contains a unique viral assembly signal (VAS) located in the 19 C-terminal residues (Lai, 2005). This VAS is

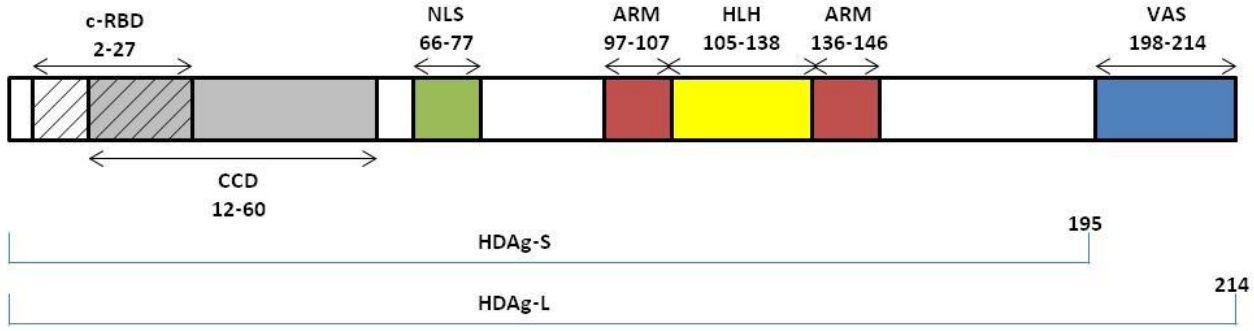

c-RBD
2-27
NLS
66-77
ARM
97-107
HLH
105-138
ARM
136-146
VAS
198-214
CCD
12-60
195
HDAg-S
214
HDAg-L

Figure 1.3: Schematic representation of HDAg domain organization.

Motifs are shown as indicated: cryptic RNA binding domain (c-RBD), coiled-coil domain (CCD), nuclear localization signal (NLS), two arginine-rich motifs (ARMs), helix-loop-helix motif (HLH) and viral assembly signal (VAS). HDAg-S and HDAg-L contain 195 and 214 amino acid residues, respectively, as indicated. (Rozzelle et al., 1995; Zuccola et al., 1998; Alves et al., 2008; Poisson et al., 1993; Lee et al., 1993; Chang et al., 1995; Lai, 2005; Lee et al., 2001; Glenn et al., 1992)

composed of a proline-rich nuclear export signal (NES; residues 198 – 210) (Lee *et al*., 2001) and an isoprenylation site (the last 4 residues, Cys-Arg-Pro-Gln, at the C-terminus) (Glenn *et al*., 1992). Isoprenylation occurs on Cys211, and is catalyzed by protein-farnesyltransferase, which adds a farnesyl lipid (Otto and Casey, 1996; Glenn *et al*., 1992). Farnesylation facilitates the direct interactions between HDAg-L and HBsAg and is required for viral packaging (Hwang and Lai, 1993a; Lee *et al*., 1994). Moreover, farnesylation causes trans-dominant repression by HDAg-L on HDV transcription, by masking a conformation corresponding to HDAg-S's 9E4 epitope (amino acid 164-195) which is important for the trans-activating activity of HDAg-S (Hwang and Lai, 1994; Hwang and Lai, 1993b).

1.5 HDV replication

HDV RNA replication follows the symmetrical double rolling circle mechanism (Taylor, 2006) (Figure 1.4). In the nucleus, a monomeric circular HDV genomic RNA is firstly transcribed into linear multimeric antigenomic transcripts. These linear multimeric antigenomic RNAs are auto-cleaved into unit length monomeric RNAs by the endogenous ribozyme motif. Linear monomeric antigenomic RNAs are then ligated by unknown cellular ligases to form circular antigenomic monomers. Following the same three steps (transcription, cleavage and ligation), the genomic HDV RNAs are generated from the circular monomeric antigenomic templates (Taylor, 2006).

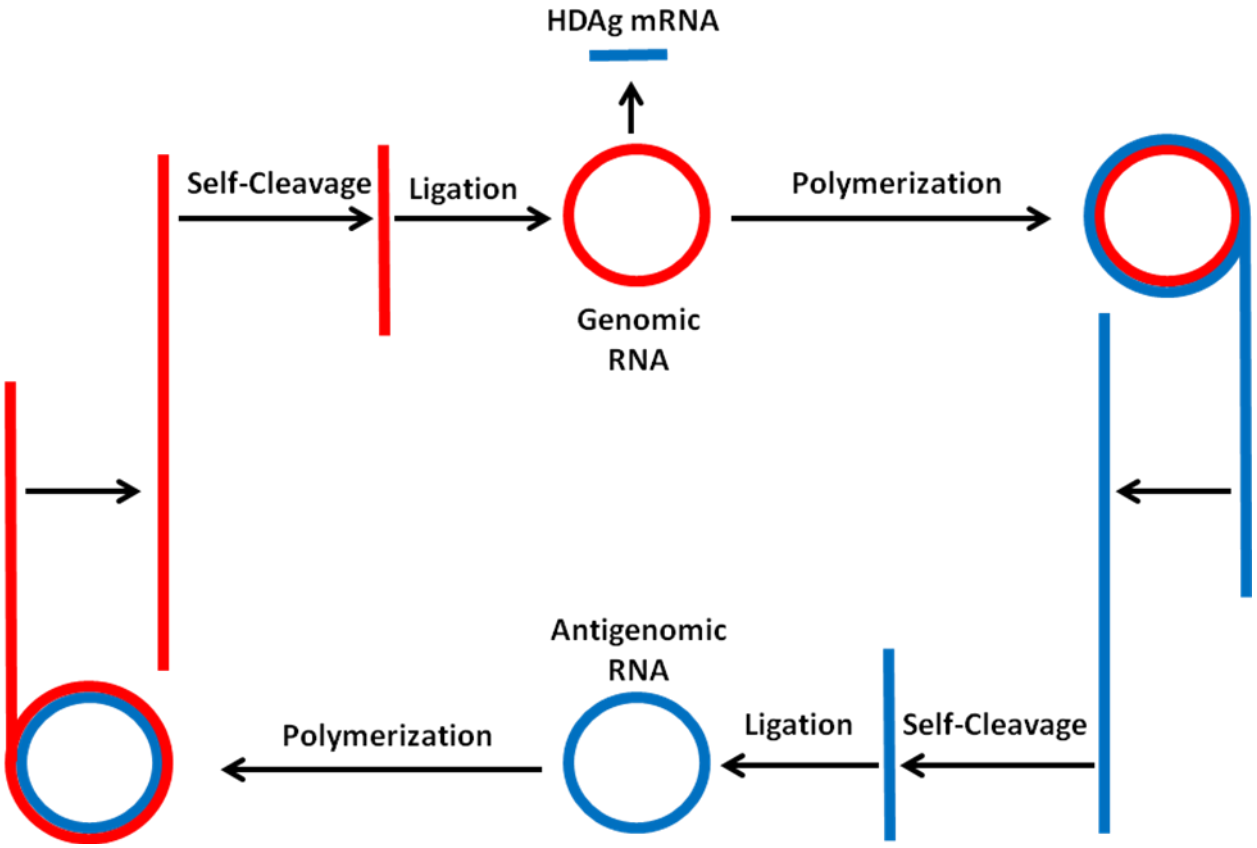

HDAg mRNA
Self-Cleavage
Ligation
Genomic
RNA
Polymerization
Antigenomic
RNA
Polymerization
Ligation
Self-Cleavage

Figure: 1.4: Schematic representation of the symmetrical rolling cycle model of HDV replication.

Linear antigenomic RNA multimers are polymerized by a host RNA polymerase from the template of circular genomic monomeric strand. The RNA multimers are then self-cleaved into monomers by HDV ribozyme motifs. The linear antigenomic monomers are then ligated to generate circular antigenomic RNA monomers via unknown host ligases. The circular genomic monomeric strands are then transcribed from the templates of antigenomic RNA monomers via steps of polymerization, self-cleavage and ligation. The mRNA for the HDV delta antigen is transcribed from the template of circular genomic RNA. Adapted with permission from http://subviral.med.uottawa.ca /~pelchatlab/cgi-bin/recherche.cgi?langue=2&section=1.

1.5.1 RNA polymerases involved in HDV replication

Since HDV RNA does not encode an RNA polymerase, the replication of HDV must use host cell transcription machinery. Many studies have demonstrated that RNAPII is a critical enzyme for HDV RNA replication. First of all, HDAg mRNA is transcribed by RNAPII because the mRNA contains not only a 5'-end m^7G cap and a 3'-end poly (A) tail, which are essential characteristics of RNAPII transcripts, but also a consensus sequence AAUAAA, typically signalling host mRNA cleavage and polyadenylation, at 15 to 20 bases upstream of the 3'-end poly (A) addition site (Hsieh *et al*., 1990; Gudima *et al*., 1999; Nie and Taylor, 2004). Secondly, a series of studies suggested, by use of α-amanitin, that HDV RNA is replicated by RNAPII machinery. α-amanitin is a mycotoxin which strongly inhibits RNAPII (~1μg/ml drug), moderately inhibits RNAPIII (~500μg/ml drug), but does not affect RNAPI (Weinmann and Roeder, 1974; de Mercoyrol *et al*., 1989). RNAPII has 12 subunits (Rbp1-12): 10 subunits associate tightly, while Rpb4 and Rpb7 form a dissociable heterodimeric sub-complex (Meyer *et al*., 2009). Rbp1 and Rbp2, two largest subunits, extensively interact and combine with other subunits to form a transcription active site cleft. α-amanitin inhibition of RNAPII is caused by interaction with a bridge α-helix domain of Rbp1 in this cleft and further preventing translocation of nucleic acid for the transcription process (Figure 1.5) (Bushnell *et al*., 2002; Klug, 2001).

In nuclear run-on experiments by MacNaughton *et al*., HDV RNA transcription using intact nuclei of transfected HepG2 cells was sensitive to low levels (1 μg/ml) of α-amanitin (MacNaughton *et al*., 1991) but not sensitive to actinomycin D (2 μg/ml) (Macnaughton *et al.,* 1990). Actinomycin D inhibits DNA replication and DNA-dependent

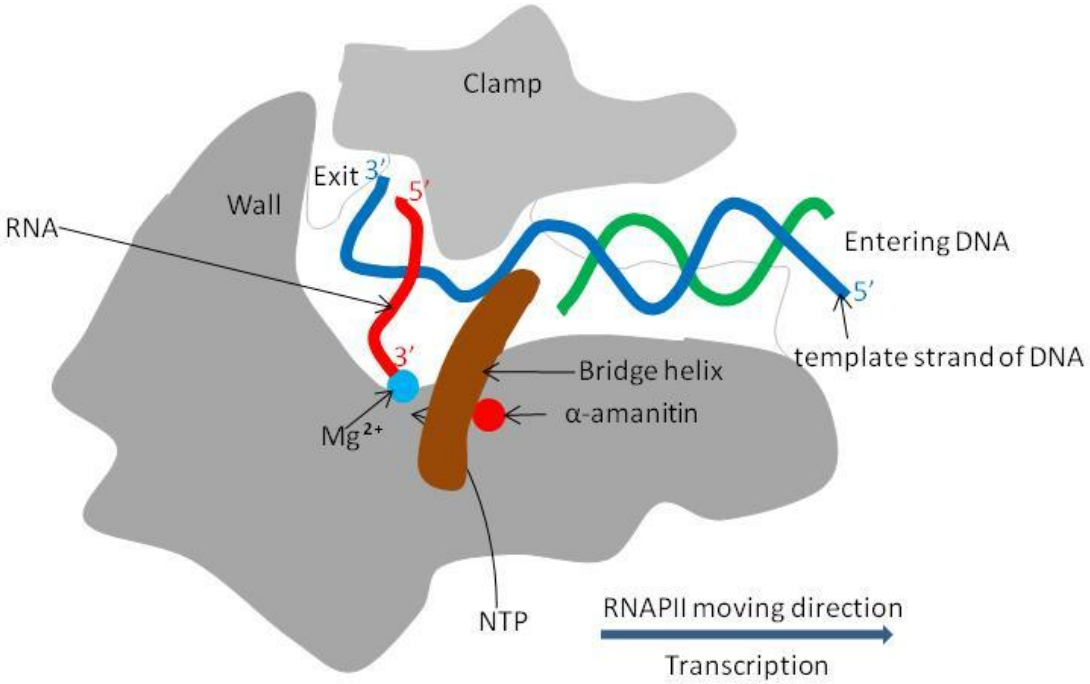

Clamp
Exit 3′
5′
Wall
RNA
Entering DNA
5′
3′
template strand of DNA
Bridge helix
α-amanitin
Mg 2+
NTP
RNAPII moving direction
Transcription

Figure 1.5: A cutaway scheme of the RNAPII transcription complex

The scheme shows the α-amanitin binding position and the path of DNA and RNA in the active site cleft formed by Rbp1 and Rbp2. Moving direction of RNAPII is from left to right. The template strand of DNA is indicated in cyan, the non-template strand of DNA is in green and generated RNA is in red. α-amanitin (shown as a red sphere) interacts with a bridge α-helix domain of Rbp1 and prevents translocation of nucleic acid in the transcription process. Mg^{2+} is indicated as a sky blue sphere at the active site, which the 3' end of RNA is close to. Nucleoside triphosphates (NTPs) get close to the active centre through a "funnel"-shaped structure and a "pore" (not shown in the figure) at the bottom of the cleft. A protein "wall" bends the DNA template to a right angle. The 5' end of RNA and 3' end of template strand of DNA exit the complex from the top of the cleft. Adapted with permission from the American Association for the Advancement of Science: Science 288 (5466): 640–649, copyright 2000. (Bushnell *et al*., 2002; Klug, 2001)

RNA transcription by intercalating at a GpC base pairs of DNA (Sobell and Jain, 1972; Kamitori and Takusagawa, 1992). Therefore, MacNaughton *et al*. (1991) deduced that the synthesis of HDV RNA might proceed via RNA-dependent transcription by RNAPII. By comparing the levels of HDV RNA transcribed *in vitro* using HDAg-positive or HDAg-negative cell nuclear extract, they also ruled out the possibility of HDAg's function as an RNA-dependent RNA polymerase (RdRP) for HDV RNA replication. However, their experiment did not identify whether the transcribed HDV RNA was genomic or antigenomic (MacNaughton *et al*., 1991). To solve this problem, Taylor's laboratory performed an improved run-on experiment by use of HDV RNA templates and nuclear extracts from different cell lines (Fu and Taylor, 1993). They reported that both HDV genomic and antigenomic full length (1.7 Kb) RNA could be synthesized from the corresponding complementary templates; moreover, HDV RNA (genomic and antigenomic) synthesis was inhibited by α-amanitin (1 μg/ml) but not inhibited by actinomycin D (20 μg/ml). Here, actinomycin D was used to inhibit the preparation of HDV RNA templates from any remaining HDV cDNA. Therefore, they indicated that the replication of HDV genomic and antigenomic RNA proceeds via RNA-dependent RNA transcription by RNAPII (Fu and Taylor, 1993). Taylor's laboratory also performed a nuclear run-on experiment by using a 293-HDV cell line containing an endogenous HDV RNA template, whose replication was under the control of a tetracycline-inducible HDAg-S cDNA (Chang *et al*., 2008). This study demonstrated that accumulation of both genomic and antigenomic HDV RNA was inhibited by a low concentration of α-amanitin. Additionally, *ex vivo* studies using Northern blot analysis demonstrated that the

accumulation of HDV genomic RNA in Huh7 cells transiently transfected with HDV RNA was sensitive to a low concentration of α-amanitin (1-3 μg/ml; Modahl *et al*., 2000; Moraleda and Taylor, 2001). Moreover, Northern blot analysis demonstrated that the accumulation of HDV mRNA in 293-HDV cells was sensitive to 1 μg/ml of α-amanitin, and the accumulation of genomic and antigenomic RNA was sensitive to 3 μg/ml α-amanitin (Chang *et al*., 2006). Furthermore, an antibody against the carboxy-terminal domain (CTD) of RNAPII inhibited HDV RNA transcription *in vitro* (genomic, antigenomic and mRNA) (Abrahem and Pelchat, 2008; Fu and Taylor, 1993) and RNAPII bound HDV RNA of both polarities (Abrahem and Pelchat, 2008; Greco-Stewart *et al*., 2007). Therefore, RNAPII is suggested to be involved in the synthesis of all three HDV RNAs.

On the other hand, several studies suggest that other RNA polymerases might be involved in the synthesis of HDV RNA. Firstly, our laboratory demonstrated that RNAPI and RNAPIII interact with both polarities of HDV genome *in vitro* and in human cells (Greco-Stewart *et al*., 2009). Secondly, the research from Lai's laboratory demonstrated that genomic, antigenomic and mRNA of HDV were synthesized by different host cell RNA polymerases at distinct sub-nuclear sites (Li *et al*., 2006; Macnaughton *et al*., 2002; Modahl *et al*., 2000; Tseng *et al*., 2008). To avoid DNA templates and any possible HDV cDNA intermediate, Lai's laboratory performed a series of experiments by use of a technique that relied on transfecting cells with HDV RNA templates and metabolic labeling of RNA products with [^{32}P] orthophosphate (Modahl *et al*., 2000; Macnaughton *et al*., 2002). They found that transcription of HDV mRNA was sensitive to a low level of α-amanitin (< 3 μg/ml) and could be partially recovered by an α-amanitin-resistant

mutant of RNAPII. Interestingly, transcription of full length (1.7 kb) antigenomic HDV RNA was not affected by high levels of α-amanitin (25 to 100 μg/ml), while synthesis of full length genomic RNA was inhibited by a low level of α-amanitin (2.5 μg/ml). According to these findings, Lai's laboratory suggested that HDV mRNA and genomic RNA are synthesized by host cell RNAPII, while antigenomic RNA is transcribed via other enzymes (Modahl *et al*., 2000; Macnaughton *et al*., 2002). To further explore HDV RNA transcription, Lai's laboratory labeled *de novo*-synthesized HDV RNA with BrUTP (5-Bromouridine 5'-Triphosphate) in HDV RNA-replicating cells (hepatocytes either containing a stably intergrated HDV cDNA or transiently transfected with HDV RNA), and then performed immunofluorescence staining (Li *et al*., 2006). They found that the synthesis of HDV genomic RNA was sensitive to α-amanitin (10 μg/ml) and the synthesized genomic RNA was more diffusely localized in the nucleoplasm, while the synthesis of antigenomic RNA was carried out in nucleolus and was not affected by α-amanitin (Li *et al.,* 2006). Furthermore, Lai's laboratory demonstrated that HDAg could co-immunoprecipitate with RNAPII and SL1, an RNAPI-associated transcription factor (Li *et al.,* 2006). Moreover, *in vitro* transcription assays using purified HDV viral particles and HeLa nuclear extract (NE) pre-treated with antibodies against various nuclear factors showed that antigenomic RNA accumulation was resistant to α-amanitin (10 μg/ml) and the use of anti-RNAPII antibody, but was inhibited when an anti-SL1 antibody was used. On the other hand, genomic RNA accumulation was sensitive to α-amanitin and an anti-RNAPII antibody (Li *et al.,* 2006). Based on the above results, Lai's laboratory suggested that HDV antigenomic RNA is synthesized in the nucleolus by RNAPI, and

genomic RNA is synthesized in the nucleoplasm by RNAPII (Li *et al.,* 2006).

The difference between the results from Taylor's laboratory and Lai's laboratory might be due to the different HDV replication systems adopted; one utilized a cell line containing an endogenous HDV template while the other utilized cells transiently transfected with an HDV template (Chang *et al*., 2008; Li *et al*., 2006). Probably the true mechanism includes the opinions from both sides, as (1) researches of our laboratory demonstrated that RNAPI, II and III interact with both HDV genome and antigenome (Greco-Stewart *et al*., 2007; Greco-Stewart *et al*., 2009), (2) Lai's experimental system is similar to the situation of acute hepatitis during the early stage of infection, while Taylor's system is more like that in chronic hepatitis Delta (Chang *et al*., 2008; Li *et al*., 2006). Therefore, HDV RNA replication might not only be spatially regulated but also temporally dependent. Whatever enzyme is involved in HDV replication, it is commonly accepted that HDV genomic RNA and mRNA are transcribed by RNAPII.

To determine the direct interaction between RNAPII and HDV RNA in cells, our laboratory performed a ribonucleoprotein immunoprecipitation assay (RIPA) (Greco-Stewart *et al*., 2007). Formaldehyde was used as a heat-reversible cross-linking agent. Following treatment with formaldehyde, HDV-replicating HeLa cells were lysed, and immunoprecipitation was performed with an antibody against the CTD of RNAPII (8WG16). Immunoprecipitated complexes of RNAPII and other cross-linked molecules, including nucleic acids and proteins, were then heated to reverse cross-linking. RT-PCR was then performed to detect the cross-linked HDV RNA. This experiment showed that RNAPII interacts with both polarities of HDV RNA *ex vivo* (Greco-Stewart *et al*., 2007).

Furthermore, co-immunoprecipitation experiments utilizing HeLa NE, radiolabeled RNA fragments of various regions of HDV, and the same antibody as in the above RIPA indicated that RNAPII specifically interacts with both right and left terminal stem-loop domains of HDV genome and antigenome (Greco-Stewart *et al*., 2007).

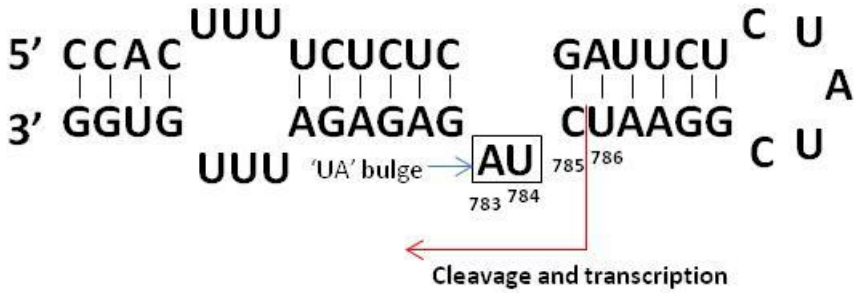
1.5.2 HDV RNA promoters

The extensive intramolecular base-pairing forces HDV genome/antigenome to fold into a double-stranded rod-like structure in the middle with two terminal loops (Taylor, 2006). This unique feature allows HDV to recruit host RNA polymerase machinery and redirect DNA-dependent transcription toward HDV RNA replication and accumulation. Specifically, RIPA experiments demonstrated that, in addition to RNAPII, RNAPI and III individually interact with both polarities of HDV RNA in cells. *In vitro* co-immunoprecipitation and RNA affinity chromatography studies indicated that RNAPI and III independently interact with the same terminal stem-loop domains as RNAPII was shown to bind to (Greco-Stewart *et al.,* 2007; Greco-Stewart *et al*., 2009). Therefore, terminal stem-loop domains of HDV RNA might contain HDV promoters.

Firstly, the left terminal stem-loop domain of HDV antigenome (nt 687 to 900) has been proposed to contain an initiation site for the transcription of HDV genomic RNA, as this region interacts with RNAP II (Greco-Stewart *et al*., 2007), and contains an active promoter for RNAP II *in vitro*, which is sensitive to α-amanitin (Filipovska and Konarska, 2000). However, the transcription starting from this promoter is proposed to occur through cleaving the antigenomic template near the left stem-loop and extending the new 3’ end by transcribing the opposite strand. This conclusion was made from a series

of detailed RNase digestion analyses of gel-purified RNA products generated from HeLa nuclear extract transcription using different mutant templates (Filipovska and Konarska, 2000). Moreover, RNase digestion analysis of RNA generated in transcription reactions containing different α-^{32}P-NTPs demonstrated that the cleavage position is between U_{786} and C_{785}, and the first incorporated nucleotide is a new C residue replacing the C_{785} of the template (Filipovska and Konarska, 2000) (see Figure 1.6 A). The secondary structure of the $U_{784}A_{783}$ bulge (see Figure 1.6 A) determines the start site for transcription, as mutant RNA templates missing this UA bulge but containing a GC or UA bulge at the opposite side of the terminal stem did not support the nuclear extract transcription, while mutant RNA templates keeping this $U_{784}A_{783}$ bulge or containing a GC bulge replacing the UA bulge efficiently supported transcription starting upstream of the bulge (Filipovska and Konarska, 2000). RNase digestion analysis also demonstrated that this RNA-primed transcription could extend about 40 nucleotides in HeLa NE (Filipovska and Konarska, 2000) and up to approximately 100 nucleotides in HeLa NE containing HDAg-S, as HDAg-S directly binds to RNAPII and promotes transcription elongation by displacing the negative elongation factor (NELF) (Yamaguchi *et al*., 2001). Moreover, Lehmann *et al*. (2007) showed that the general transcription factor IIS (TFIIS), in a complex with RNAPII, facilitates the previously reported cleavage of the HDV antigenomic template, creating a reactive stem-loop in the hybrid site and extending the new 3' end. A crystallography study suggested that cleavage occurs when the HDV bulge $U_{784}A_{783}$ passes the bridge helix of RNAP II and the Mg^{2+} active site (Figure1.6 B) (Lehmann *et al*., 2007). The study also showed that there is no base-specific contact between RNAP II and RNA in RdRP

(A)


5′ CCAC UUU UCUCUC GAUUCU C U A
3′ GGUG AGAGAG CUAAGG C U
UUU
'UA' bulge
AU
783 784
785 786
Cleavage and transcription

(B)

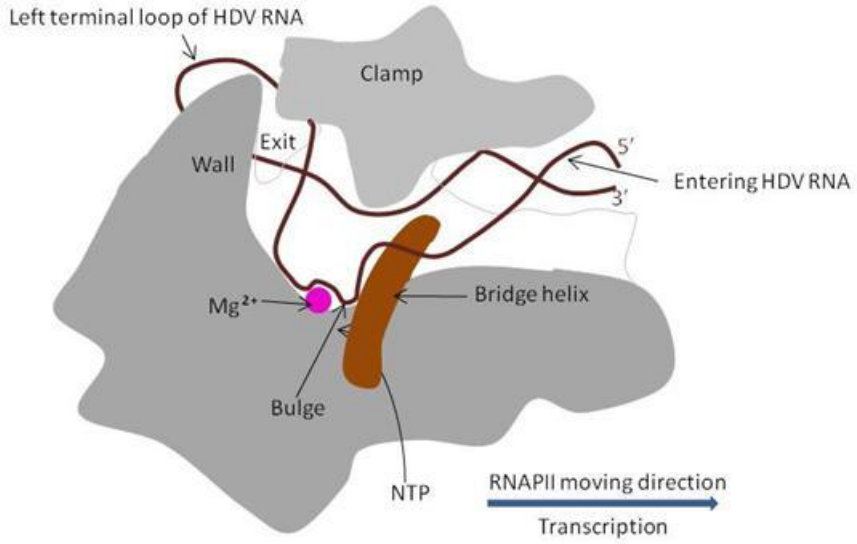

Left terminal loop of HDV RNA
Clamp
Exit
Wall
5′
3′
Entering HDV RNA
Mg2+
Bridge helix
Bulge
NTP
RNAPII moving direction
Transcription

Figure1.6: Schematic representation of the transcription starting from the left terminal stem-loop of HDV antigenomic RNA.

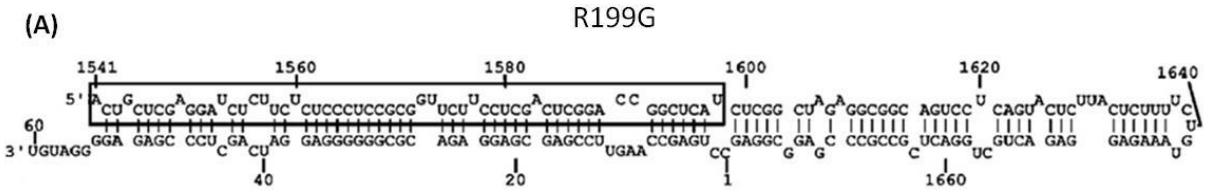
(A) Extreme tip of left terminal stem-loop domain of HDV antigenomic RNA possessing promoter activity *in vitro* (Filipovska and Konarska, 2000; Lehmann *et al*., 2007). $U_{784}A_{783}$ bulge is outlined. Red arrow indicates transcription via cleavage occurs between U_{786} and C_{785} and extends the new 3' end by copying the opposite strand. (Figure adapted with permission from Elsevier Inc.: Virology 357: 68–78, copyright 2006.)

(B) Schematic representation of the model of the initial interaction of RNAPII complex with the left terminal stem-loop domain of HDV antigenome. The HDV bulge is predicted to pass the bridge helix and Mg^{2+} at the active site, causing the cleavage between U_{786} and C_{785}. The cut-view of RNAPII complex is shown the same way as in Figure 1.5. Adapted with permission from Macmillan Publishers Ltd: Nature 450 (7168):445-449, copyright 2007. (Bushnell *et al*., 2002; Klug, 2001)

elongation complex structures; moreover, RNAP II recognizes the extreme left stem structure of HDV rather than a particular RNA sequence, consistent with previous observations that HDV secondary structure is more important than the primary sequence for RNA transcription and RNAPII-RNA interaction (Filipovska and Konarska, 2000; Greco-Stewart *et al*., 2007). As of now, *de novo* synthesis of HDV genomic RNA has not been reported in reconstituted RNAP II systems *in vitro*; therefore, this reaction might be modulated by some unknown critical cofactors in cells or initiate from an unknown promoter site.

On the other hand, using an *in vitro* transcription assay with HeLa nuclear extract, Beard *et al.* (1996) reported that an HDV antigenomic RNA transcript could be synthesized from a 199nt genomic RNA fragment (including 5'- nt 1541 through 1679 to 60 -3'; called "R199G" in my thesis) (see Figure 1.7A) containing the right terminal stem-loop region of HDV. This R199G fragment also includes the HDAg mRNA transcription initiation site. Using primer extension analysis, the 5'-end of HDAg mRNA extracted from cells generating HDV RNA was initially found to correspond to nt 1631±1 (A_{1631}) (Hsieh *et al*., 1990; Beard *et al*., 1996), while the 5'-end of an RNA transcribed off the R199G template was found within 1 to 13 nt of that of the mRNA (Beard *et al*., 1996). However, the new refined data obtained by 5' rapid amplification of cDNA ends assay (5' RACE) indicated that the transcription of both the HDV mRNA and the RNA transcribed off the R199G template predominantly started at U1630 (Gudima *et al*., 1999; Abrahem and Pelchat, 2008). U1630 is in the middle of a 3-nt external bulge ($A_{1631}U_{1630}U_{1529}$, see Figure 1.7 B), consistent with the preference of RNAPII to initiate DNA-dependent RNA

(A)


R199G
1541
1560
1580
1600
1620
1640
60
40
20
1
1660

(B)

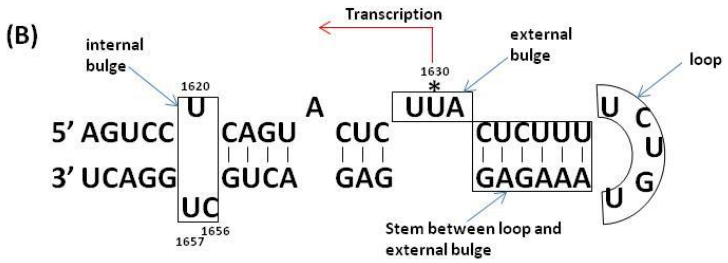

Transcription
internal bulge
external bulge
loop
1620
1630
1656
1657
5' AGUCC
3' UCAGG
CAGU
GUCA
UC
A
CUC
GAG
UUA
CUCUUU
GAGAAA
U C
U
U G
Stem between loop and external bulge

Figure 1.7: Schematic representation of the secondary structure of right terminal stem-loop domain of HDV genomic RNA.

(A) Secondary structure of R199G, a 199-nt fragment corresponding to the right terminal stem-loop domain of HDV genomic RNA. The region corresponding to the 5'-end of the ORF is indicated with outline. (Figure adapted with permission from Elsevier Inc.: Virology 356: 35–44, copyright 2006.)

(B) Right extreme tip of R199G. Structures of bulges, stem and loop are indicated with outline. Red arrow indicates the transcription direction. Transcription initiates at U_{1630}. (Figure adapted with permission from Elsevier Inc.: Virology 357: 68–78, copyright 2006.)

transcription by a purine (Darnell et al, 1986). These results suggested that the right terminal stem-loop domain of the HDV genome contains an RNA promoter for the transcription of the mRNA. However, we do not know if the synthesis of the antigenome would start from the same promoter because the exact transcription initiation site of the full length antigenomic RNA has never been found.

By comparing 8 sequences of HDV genotypes I, II and III, Beard *et al*. (1996) found that the region corresponding to nts 1608 to 1669 within R199G contains highly conserved sequence and secondary structures. In order to characterize the involvement of this region in HDV replication, mutagenesis analysis of this region was performed (Beard *et al*., 1996). *In vitro* HeLa nuclear extract transcription followed by RNase protection analysis showed that the transcription level of antigenomic HDV RNA was reduced when using HDV genomic RNA templates mutated on the external bulge ($A_{1631}U_{1630}U_{1529}$) or internal bulge (formed by U_{1620} and $C_{1656}U_{1657}$) (see Figure1.7 B) (Beard *et al*., 1996). This experiment directly demonstrated that these bulge structures are necessary for the HDV RNA promoter activity. Furthermore, Northern blotting and 5' RACE of the polyadenylated RNA extracted from the transfected cells determined that mutations around the external bulge ($A_{1631}U_{1630}U_{1529}$) and at the right tip of the HDV genomic RNA (including the 6-nt stem and the loop) (see Figure1.7 B) could impair the accumulation of antigenomic RNA and/or change the position of the 5' end (Gudima *et al*., 1999; Wang *et al*., 1997; Wu *et al*., 1997). All the above results obtained from mutagenesis analysis indicated that the sequence and secondary structure at the right terminal stem-loop of genomic RNA are important for promoter activity of HDV mRNA.

Moreover, base-pair covariation analysis of 81 variants of different HDV genotypes demonstrated a strong selection to maintain the hairpin conformation of HDV RNA tips (Greco-Stewart *et al*., 2007). This conformation contains a conserved triple base pair motif (CUC/GAG) and a strong polarization in the purine/pyrimidine content near the tip of the rod up to the CUC/GAG stem (Figure1.8). Mutation studies showed that the secondary structure and the polarization in the purine/pyrimidine in these regions are necessary for RNAPII interaction with HDV RNA, as disruption of this hairpin secondary structure or switching of the polarization strongly inhibited the interaction (Greco-Stewart *et al*., 2007).

Furthermore, RNA affinity chromatography and *in vitro* nuclear extract transcription showed that an active pre-initiation complex (PIC) formed at the R199G region (Abrahem and Pelchat, 2008). This PIC included the core RNAPII subunit and general transcription factors (TFIIA, B, D, E, F, H and S). Moreover, binding assays showed that the TATA-binding protein (TBP) directly binds the HDV RNA promoter, suggesting the requirement of TBP for nucleation of the RNAP II complex on this promoter (Abrahem and Pelchat, 2008). These results suggest that the formation of the PIC on the HDV RNA promoter is generally similar to that on the DNA promoter. Together, all these studies demonstrated that the sequence and secondary structure of the right tip of HDV genome are important for this RNA promoter to recruit RNAPII and transcription factors to form a PIC.

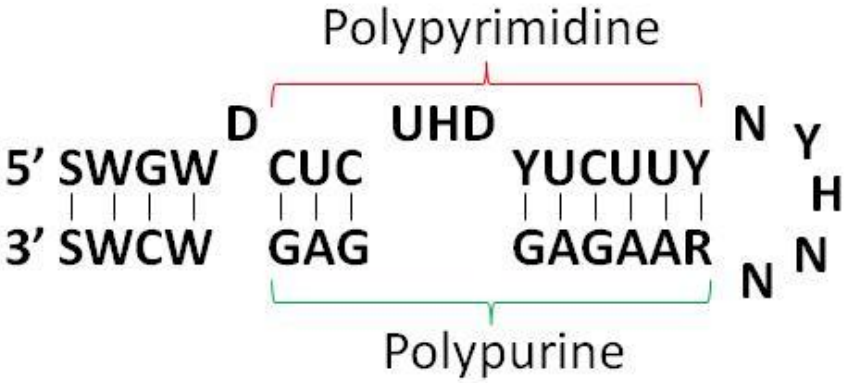

Polypyrimidine
D
UHD
N
Y
5’ SWGW
CUC
YUCUUY
H
3’ SWCW
GAG
GAGAAR
N
N
Polypurine

Figure 1.8: Schematic representation of the conserved secondary structure of the right extreme tip of genomic HDV RNA.

Secondary structure was derived by covariation analysis of 81 variants. The conserved CUC/GAG, polypyrimidine and polypurine sequences are shown as indicated. Nucleotides on the secondary structure are represented by IUPAC-letter code abbreviations. (Figure adapted with permission from Elsevier Inc.: Virology 357: 68–78, copyright 2006.)

1.5.3 Transcription regulation by HDAg

A heterodimer of DSIF (5,6-dichloro-1-4-D-ribofuranosyl benzimidazol (DRB) sensitivity inducing factor) and NELF (negative elongation factor) represses the transcription elongation activity of RNAPII (reviewed in Nechaev and Adelman, 2011; Wada *et al*., 1998; Yamaguchi *et al*., 1999). The sequences of the N-terminus of the HDAg and the NELF-A subunit contain weak homology (Yamaguchi *et al*., 2001). Immunoprecipitation and *in vitro* transcription assays showed that both isoforms of HDAg (more efficiently HDAg-S) directly bind RNAPII and promote transcription elongation by displacing NELF (Yamaguchi *et al*., 2001). Particularly, the C-terminus (amino acid 130 - 195) of HDAg-S binds RNAPII, and this region is necessary for HDV replication *ex vivo* and effective elongation by RNAPII *in vitro* (Yamaguchi *et al*., 2001). Using HeLa NE *in vitro* transcription, it was demonstrated that (1) HDAg, especially HDAg-S, reversed the trans-repression effect of DSIF/NELF by displacing NELF when NELF was present in the reaction system; (2) it also directly stimulated RNAPII transcription elongation when NELF was immunodepleted (Yamaguchi *et al*., 2001). Further studies of specific photocrosslinking and proteolytic mapping with hydroxylamine showed that the C-terminus of HDAg-S interacts with the RNAPII clamp formed by the largest subunits Rpb 1 and Rbp 2 (Yamaguchi *et al*., 2007). Moreover, evidence from the transcription assay using reconstituted elongation complexes demonstrated that HDAg-S stimulates the rate of transcription while affecting transcription fidelity; therefore, Yamaguchi et al. (2007) proposed that HDAg might bind to and loosen the clamp, and thereby promote forward translocation of RNAPII while sacrificing fidelity. However, this hypothesis has no

direct evidence so far.

1.6 Interaction of HDV with host proteins

Because of its extremely limited protein coding capacity, HDV elaborately interacts with and utilizes host cellular proteins for its transcription and replication. Because HDAg-S plays a key role in HDV replication, Kay's laboratory created a stable 293 cell line expressing a functional N-terminal Flag-tagged HDAg-S. By use of this cell line, they performed mass spectrometric analysis following anti-Flag immunoprecipitation, and identified that more than 100 proteins interacted with HDAg-S (Cao *et al*., 2009). These proteins include 9 out of 12 subunits of RNAPII, heterogeneous nuclear ribonucleoproteins (hnRNPs), RNA helicases, splicing-related proteins, transcription-related proteins, RNA binding and processing proteins, etc. (Cao *et al*., 2009). Among these proteins were included previously reported HDAg-S-interacting proteins SPT5 (DSIF large subunit) (Yamaguchi *et al*., 2001), double-stranded RNA-activated protein kinase R (PKR) (Chen *et al*., 2002) and MOV-10, which might be involved in HDV replication (Haussecker *et al*., 2008). It should be pointed out that the interactions in this paper did not discriminate between direct or indirect, RNA-dependent or RNA-independent interactions (Cao *et al*., 2009). Further, a qRT-PCR study following siRNA(small interfering RNA)-mediated knockdown of 65 selected proteins showed that 18 knockdowns significantly affected HDV replication (Cao *et al*., 2009). Other host cellular proteins interacting with HDV include those that interact with HDAg-L, such as PKC, FTase, nuclear export signal-interacting protein (NESI), and those that interact with HDV RNA, such as ADAR1, PKR, RNAPI, RNAPII and RNAPIII (reviewed

in Greco-Stewart and Pelchat, 2010). Although some functions of interactions of HDV with host proteins have been reported, the biological significance of most of these interactions remains unknown.

1.7 Host proteins interacting with R199G

Our laboratory's previous research focused on the right terminal stem-loop region of HDV genomic RNA (R199G), which carries a putative promoter for HDV mRNA transcription by RNAPII (see Figure 1.7). In order to identify the R199G-interacting host nuclear proteins, our laboratory used R199G as bait to perform UV-crosslinking experiment with HeLa NE (Greco-Stewart *et al*., 2006; Sikora *et al*., 2009). Following UV-crosslinking, mass spectrometry analysis of the obtained ribonucleoprotein complex revealed that the polypyrimidine tract-binding protein-associated splicing factor (PSF), the eukaryotic elongation factor 1A1 (eEF1A1) and the 54kDa nuclear binding protein (p54nrb) interacted with R199G (Greco-Stewart *et al*., 2006; Sikora *et al*., 2009). Alternately, using R199G affinity chromatography followed by mass spectrometry analysis, our laboratory found that the heterogeneous nuclear ribonuceloprotein L (hnRNP-L) interacted with R199G (Sikora *et al*., 2009). In addition, our laboratory performed electrophoretic mobility shift assay (EMSA) experiments to screen several purified recombinant RNA-binding proteins for their interaction with R199G (Sikora *et al*., 2009). By this strategy, our laboratory determined that the alternative splicing factor (ASF/SF2) interacted with R199G (Sikora *et al*., 2009).

In order to confirm the above identified proteins' interactions with R199G, co-immunoprecipitation experiments were performed in HeLa NE with R199G and

antibodies against these proteins (Greco-Stewart *et al*., 2006; Sikora *et al*., 2009). The interaction between glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and R199G was also examined by co-immunoprecipitation (Sikora *et al*., 2009). GAPDH was previously reported to bind the UC-rich region (nt 379 – 414) of HDV antigenomic RNA to enhance HDV ribozyme activity (Lin *et al*., 2000). Co-immunoprecipitation experiments confirmed that PSF, eEF1A1, p54nrb, hnRNP-L, ASF/SF2 and GAPDH interacted with R199G *in vitro* (Greco-Stewart *et al*., 2006; Sikora *et al*., 2009). Furthermore, a series of RIPA studies were performed with HeLa cells replicating HDV RNA to confirm these HDV RNA-protein interactions *ex vivo* (Greco-Stewart *et al*., 2006; Sikora *et al*., 2009). During RIPA, HDV-replicating cells were treated with formaldehyde, a heat-reversible RNA-protein crosslinking reagent. Reverse transcription polymerase chain reaction (RT-PCR) following co-immunoprecipitation indicated that all of these proteins interacted with both polarities of HDV RNA *ex vivo* (Sikora *et al*., 2009). Recently, our laboratory found that ASF/SF2 binds to a purine-rich region inside R199G but far from both the HDV mRNA transcription initiation site and the RNAPII binding site, and shRNA(small hairpin RNA)-mediated knockdown suggested that ASF/SF2 is not involved in HDV replication *ex vivo* (Sikora *et al*., 2013). The functions of these interactions of host proteins with R199G need to be further explored.

The HDV-interacting protein investigated in my project is PSF, also referred to as the splicing factor proline / glutamine-rich (SFPQ) (see Figure 1.9). PSF is a 76 kDa nuclear protein (Patton *et al*., 1993). The N-terminus of PSF is comprised of an RGG motif and a proline/glutamine-rich domain, while the C-terminus of PSF possesses two RNA

polypyrimidine tract- binding protein-associated splicing factor (PSF)

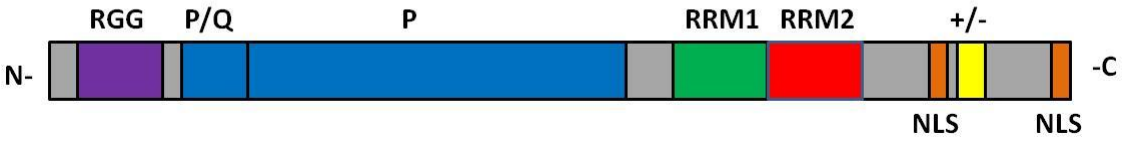


Figure 1.9: Schematic representation of the polypyrimidine tract-binding protein associated splicing factor (PSF).

PSF is a 76kDa multifunctional nuclear protein. It includes two RNA recognition motifs (RRM1 and RRM2), a highly charged domain (+/-), two nuclear localization signals (NLS), an arginine (R) and glycine (G) rich domain (RGG), a proline (P) and glutamine (Q) rich domain (P/Q) and a proline rich domain (P), shown as indicated. (Figure adapted with permission from Elsevier Ltd: FEBS Letters 531 109–114, copyright 2002).

recognition motifs (RRMs), two nuclear localization signals (NLSs) and a region of highly charged residues (Patton *et al*., 1993; Shav-Tal and Zipori, 2002). PSF participates in many cellular processes, such as pre-mRNA splicing, transcriptional regulation, nuclear retention of defective RNAs, DNA unwinding and pairing, and maintenance of pH homeostasis (Shav-Tal and Zipori, 2002). PSF has been found to be distributed in the nuceloplasm, nuclear matrix, nucleolus and nuclear membrane (Shav-Tal and Zipori, 2002). Moreover, PSF can form a heterodimer with the 54kDa nuclear RNA-binding protein (P54nrb), and both proteins belong to a *Drosophila* behavior/human splicing (DBHS) family of RNA-binding proteins, which contain two highly homologous RRMs and a carboxy-terminal dimerization coiled-coil domain (Dong *et al*., 1993). Both PSF and P54nrb were observed to bind double-stranded DNA, single-stranded DNA, and RNA (Shav-Tal and Zipori, 2002). They were also found to co-localize with paraspeckle component 1 (PSPC1; another protein of DBHS family) in the paraspeckles, which are RNA-protein bodies located in the interchromatin space of mammalian nuclei, important for regulation of gene expression through the RNA nuclear retention mechanism (Fox and Lamond, 2010).

To identify the HDV region which PSF interacts with, a series of *in vitro* co-immunoprecipitation assays were performed by use of HeLa NE, an antibody against PSF and radiolabeled RNA fragments containing different regions of HDV, including the central region, and the right or left terminal stem-loop domain of both genomic and antigenomic HDV (Greco-Stewart *et al*., 2006). These studies indicated that PSF specifically interacts with both right and left terminal stem-loop domains of both

polarities of HDV (Greco-Stewart *et al*., 2006). Furthermore, EMSA experiments utilizing purified recombinant PSF and radiolabeled HDV RNA fragments demonstrated that this PSF-HDV interaction is direct and specific (Greco-Stewart *et al*., 2006). Together, these studies indicated that PSF directly and specifically interacts with terminal regions of both polarities of HDV, including R199G.

1.8 Interaction of PSF with the CTD tail of RNAPII

Eukaryotic RNAPII is responsible for the transcription of mRNA and many non-coding RNAs (ncRNA) (Hsin and Manley, 2012). Rbp1, the largest subunit of RNAPII, contains a C-terminal domain (CTD) consisting of multiple repeats of the consensus sequence YSPTSPS. The mammalian CTD contains 52 heptapeptide repeats. The heptapeptide motifs are closer to consensus at the N-terminal half of the CTD than at the C-terminal half (Hsin and Manley, 2012). The CTD undergoes extensive and dynamic post-translational modifications, including phosphorylation at the hydroxylated residues (Tyr1, Ser2, Thr4, Ser5, Ser7), isomerisation at the Pro3 and Pro6, glycosylation of the serine and threonine residues, and methylation of lysine and arginine residues of non-consensus heptad repeats (Heidemann *et al*., 2013; Hsin and Manley, 2012). The combination of various modifications composes the so called “CTD code”, spatially and temporally coordinating the recruitment of specific transcription factors interacting with RNAPII during transcription (Heidemann *et al*., 2013). Among all of the modifications, phosphorylation is the most prominent. The dynamic phosphorylation pattern comprises part of the “CTD code”, and phosphorylation and dephosphorylation of the CTD are catalyzed by various CTD kinases and phosphatases, respectively (Hsin and Manley,

2012). Phosphorylation of Ser-5 recruits the 5’ capping machinery to process the uncapped mRNA, while the recruitment of factors for transcription enlongation and mRNA splicing corresponds to the phosphorylation level of Ser2/Ser5 (Heidemann *et al*., 2013; Hsin and Manley, 2012). In addition, the initiating (RNAP IIA) and elongating (RNAP IIO) forms of RNAPII contain hypophosphylated and hyperphosphylated CTDs, respectively (Hsin and Manley, 2012).

A series of studies demonstrated that PSF binds to the CTD of RNAPII. Affinity chromatography experiments combined with N-terminal sequencing (Edman degradation) and Western blotting showed that PSF and p54nrb specifically interacted with immobilized recombinant mouse CTD with an N-terminal polyhistidine tag in HeLa cell extracts (Emili *et al*., 2002). Mass spectrometry following glutathione S transferase (GST) affinity chromatography found that PSF and p54nrb interacted with both nonphosphorylated and phosphorylated GST-CTD in HeLa nuclear extracts (Emili,A. 2002). Furthermore, co-immunoprecipitation using monoclonal antibodies 8WG16 and B3, which specifically recognize hypophosphorylated CTD of RNAPIIA and hyperphosphorylated CTD of RNAPIIO, respectively, indicated that PSF and p54nrb bind to both hypo- and hyperphosphorylated forms of the CTD (Emili *et al*., 2002). Interestingly, Western blotting analysis of the RNAPIIA holoenzyme, which was purified by immobilized human TFIIS affinity chromatography, showed that the stoichiometries of both PSF and p54nrb in the RNAPIIA holoenzyme were greater than 1:1 relative to RNAPII, consistent with the presence of multiple heptad repeats of the RNAPII CTD (Emili *et al*., 2002). Moreover, GST-CTD affinity chromatography showed that PSF and p54nrb

can facilitate the interaction between RNA and the CTD (Emili *et al*., 2002). In this experiment, GST columns were used to immobilize recombinant GST-CTD. GST-CTD columns were then pre-incubated with or without HeLa PSF and p54nrb proteins, which were purified to near homogeneity by GST-CTD affinity chromatography. Pre-incubated GST-CTD columns were then incubated with radiolabeled RNA transcripts. The GST-CTD column containing PSF and p54nrb retained many more RNA transcripts than GST-CTD column only. This result suggested that PSF and p54nrb might play a role in the association of RNA with the CTD of RNAPII (Emili *et al*., 2002). Together, all these results suggest that PSF and p54nrb link the CTD of RNAPII and the RNA during RNAPII transcription initiation and elongation (Emili *et al*., 2002).

1.9 Involvement of PSF in HDV RNA replication and the interaction of RNAPII with HDV RNA

Since both PSF and RNAPII bind to the same region of HDV RNA (Greco-Stewart *et al*., 2006; Greco-Stewart *et al*., 2007), and PSF binds to the CTD of RNAPII (Emili *et al*., 2002), PSF might act as the key link for the interaction of RNAPII with the HDV RNA promoter (Greco-Stewart, 2009). To confirm this hypothesis, our laboratory utilized an anti-RNAPII (8WG16) antibody to co-immunoprecipitate radiolabled R199G from HeLa NE with competitors (Greco-Stewart, 2009). The competitors included purified recombinant CTD and three PSF-binding RNAs identified by SELEX (systematic evolution of ligands by exponential enrichment). R199G could not be co-immunoprecipitated with RNAPII when an excess of each competitor for CTD- or PSF-binding was added into the reaction mixture. These results suggested that both the PSF-R199G and the PSF-CTD

associations are necessary for the interaction of RNAPII with the HDV promoter R199G (Greco-Stewart, 2009).

To further indicate that HDV RNA binds to RNAPII and PSF simultaneously, our laboratory performed RNA affinity chromatography followed by co-immunoprecipitation assays. Firstly, a RNA affinity tag (called S1; Srisawat & Engelke, 2002) against streptavidin was connected to the 3'-end of the R199G (Greco-Stewart, 2009). The S1-tagged R199G was then coupled to streptavidin beads and acted as bait to fish the proteins in HeLa NE. Co-immunoprecipitation assays using an anti-PSF or anti-RNAPII antibody were performed on the samples of the R199G-nucleoprotein complexes eluted from the above RNA affinity chromatography. Following co-immunoprecipiation, Western blotting analysis indicated that PSF co-immunoprecipitated with RNAPII and vice versa. Therefore, R199G interacts with both PSF and RNAPII simultaneously (Greco-Stewart, 2009).

In order to determine the biological significance of PSF in HDV RNA replication *ex vivo*, our laboratory performed an RNAi-mediated PSF knockdown experiment (Al-Ali, 2011). The experiment was performed with a HEK293 cell line (called 293-HDV) replicating HDV RNA genome (Chang *et al*., 2005). To construct this cell line, HEK293TRex cells were stably transfected with a copy of HDAg-S cDNA which is controlled by a tetracycline (TET) inducible promoter. These cells (called 293-HDAg-S) were further co-transfected with a mutant HDV RNA genome which cannot generate a functional HDAg-S (Chang *et al*., 2005; Chao *et al*., 1990). Thus the replication of HDV genome depends on the expression of the HDAg-S cDNA. RT-qPCR determined that the

TET-induced HDV RNA accumulation was significantly reduced by knockdown of PSF (Al-Ali, 2011). Therefore, PSF might directly or indirectly facilitate HDV RNA accumulation in cells (Al-Ali, 2011).

1.10 Statement of purpose and hypothesis

Previous studies demonstrated that RNAPII transcribed HDV mRNA and genomic RNA (reviewed in Taylor, 2006). RNAPII and PSF specifically interact with genomic and antigenomic HDV RNA at the right and left terminal stem-loop domains, two putative HDV promoter regions (Greco-Stewart *et al*., 2006; Greco-Stewart *et al*., 2007). Furthermore, the right terminal stem-loop domain (R199G) of HDV genomic RNA contains an active promoter for the RNAPII transcription of HDV RNA (Abrahem and Pelchat, 2008). The RNAPII transcription preinitiation complex formed on R199G is similar to that of DNA-dependent transcription (Abrahem and Pelchat, 2008). On the other side, PSF interacts with the hypo- and hyperphosphorylated forms of the CTD of RNAPII and facilitate the interaction of RNAPII with RNA transcripts (Emili *et al*., 2002). Moreover, PSF simultaneously interacts with HDV RNA promoter (R199G) and the CTD of RNAPII (Greco-Stewart, 2009). Knockdown experiment showed that PSF facilitates HDV accumulation *ex vivo* by direct or indirect effects (Al-Ali, 2011). Therefore, I hypothesize that *PSF might act as a transcription factor for the RNAPII transcription of HDV RNA through linking the HDV RNA promoter (R199G) and the CTD tail of RNAPII.*

In order to verify this hypothesis, the objectives of my thesis were:

(1) To determine whether PSF can facilitate the interaction of RNAP II with R199G.

(2) To identify which domains of PSF interact with RNAPII.

(3) To identify which domains of PSF interact with HDV RNA promoter (R199G).

(4) To examine whether PSF can facilitate the transcription from R199G.

Chapter 2. Materials and Methods

2.1 Synthesis of RNA

PCR amplification was performed off the pHDVd2 plasmid to generate DNA templates used for synthesis of HDV-derived RNAs. The pHDVd2 plasmid is a derivative of pBluescriptKS$^+$ (Stratagene), carrying a dimeric HDV cDNA (Kuo *et al*., 1988b). A short sequence containing the T7 promoter region was added to the 5' end of the forward PCR primer. The DNA template used for synthesis of P11.60 RNA was generated by annealing and extension of two primers. Sequences of primers used for generating DNA templates are listed in Appendix I.

RNA was synthesized using 1µg of each DNA template by *in vitro* run-off transcription using 200 units of T7 RNA polymerase (New England BioLab; NEB) and 8 units of human placental RNase inhibitor (Biobasic) in 80mM Tris-HCl, pH 7.9, 40mM dithiothreitol, 20mM $MgCl_2$, 2mM spermidine and 1.25mM of each NTP. Following 4 hours of incubation at 37^{o}C, DNA templates were digested with 1 unit of DNase I (Promega) at 37^{o}C for 30 min. RNA products were resolved by denaturing urea polyacrylamide gel electrophoresis (Urea-PAGE) (10% polyacrylamide (19:1, acrylamide: bis-acrylamide) / 7 M urea) in 1x TBE buffer (90mM Tris-HCl, pH 8.3, 90mM boric acid, 2mM EDTA). Following Urea-PAGE, RNA bands were visualized by UV shadowing, excised and eluted in elution buffer (500mM ammonium acetate, 0.1% SDS, 10mM EDTA) overnight at 4^{o}C. The eluted RNA was then ethanol-precipitated with 0.3M NaOAc (pH5.0), dissolved in 100µl of H_2O, passed through Sephadex G-50 column (GE Healthcare), ethanol-precipitated and stored at −80°C. Before using, RNA was

resuspended in 20µl of H_2O and quantified by UV spectrophotometry at 260nm.

2.2 Radiolabeling

Radiolabeled R199G RNA was synthesized by *in vitro* run-off transcription with all components of reaction (see section 2.1) and 10µCi of [α-^{32}P] GTP (PerkinElmer; PE). A radiolabeled Y-primer (see Appendix I) was used for synthesizing the radiolabeled cDNA of transcription product from a Y-R199G RNA template (see Figure 3.9). 5'-radiolabeling of the Y-primer was carried out by incubating 3 pmoles of Y-primer with 10 units of T4 polynucleotide kinase (NEB) and 10µCi of [γ-^{32}P] ATP (PE) in 15µl of reaction buffer (70 mM Tris-HCl, pH7.6, 10mM $MgCl_2$ and 5mM dithiothreitol) at 37^{o}C for 1 hour. Following the reaction, radiolabeled nucleic acids were phenol-chloroform extracted, passed through Sephadex G-50 columns (GE Healthcare), ethanol-precipitated, resuspended in 20µl of H_2O and store at -20^{o}C.

2.3 Expression of Myc epitope-tagged wild type and deletion mutant proteins

N-terminal Myc epitope-tagged wild type and deletion mutant PSF plasmids (gifts of Dr. Benjamin Blencowe, University of Toronto, Figure 3.3) were transformed into *Escherichia coli* (*E.coli*) strain BL21(DE3)pLysS (Novagen). Transformed bacteria were then inoculated onto 10 cm LB agar plates supplemented with ampicillin (100µg/ml) and incubated at 37^{o}C for 24 to 48 hours. Individual colonies were inoculated into 5ml of LB ampicillin medium and incubated overnight at 37^{o}C with agitation. Plasmids were extracted and purified with the QIAprep Miniprep kit (Qiagen) according to manufacturer's protocol.

Following quantification by UV spectrophotometry at 260nm, plasmids were

transfected into HEK293T cells (HEK293) with LipoD293 kit (SignaGen Laboratories) according to manufacturer's protocol. Briefly, cells (~1x10^6/well) were seeded on a 6-well plate (Corning) containing Dulbecco's modified eagle medium – high glucose (DMEM-high glucose, Invitrogen) supplemented with 10% fetal bovine serum (FBS) and then incubated overnight at 37^{o}C in an incubator with 5% CO_2 until they reached ~80% confluency. LipoD293 / plasmids mixture was prepared by mixing dilute solution of LipoD293 (containing 3µl of LipoD293 and 50µl of serum-free DMEM – high glucose) with dilute solution of plasmids (containing 1µg of plasmids and 50µl of serum-free DMEM – high glucose). The mixture was then transferred drop-wise to the cells. The transfected cells were incubated at 37^{o}C for 48 hours in an incubator with 5% CO_2. Following incubation, media in each well was removed by aspiration; cells were rinsed twice with 1X phosphate buffered saline (1XPBS, Invitrogen), collected in individual 1.5ml eppendorf tubes with 1ml of 1X PBS and then spun at 1500rpm for 5 min at 4°C using the Sorvall microcentrifuge (Fisher Scientific). The supernatant was removed, and the cell pellet was washed twice with 1ml of 1X PBS and spun down at 4°C between each step. The cell pellet was directly used for protein extraction or snap-frozen with liquid nitrogen and stored at – 80^{o}C.

Whole cell extraction was performed using the following method: the cell pellet was resuspended in 200µl of non-denaturing immunoprecipitation buffer (IP buffer) (50mM Tris, pH 7.0, 0.5% Triton X-100 or Nonidet P40, 150mM NaCl, 1mM EDTA, 10% glycerol) supplemented with 1% (v/v) protease inhibitor cocktail for mammalian cell extracts (Sigma-Aldrich), 1% (v/v) phosphatase inhibitor cocktail 1 (Sigma-Aldrich) and

1mM phenylmethylsulfonyl fluoride (PMSF). The cell suspension was passed through a 26-gauge needle (BD) 10 times, cooled on ice for 30 min with mild interval vortex and spun for 20 min at 15000 x g and 4^{o}C. Following spinning, pellet was discarded and clear supernatant contained the whole cell extract (WCE). Concentration of total proteins in WCE was measured by Bradford assay. WCE was then divided into several aliquots, snap-frozen in liquid nitrogen and then stored at – 80^{o}C.

2.4 Hexahistidine–tagged PSF expression and isolation

N-terminal hexahistidine–tagged PSF (His6-PSF) was expressed from plasmid pET15b-PSF (Patton *et al*., 1993) transformed into *E.coli* OverExpress C43 (DE3) competent cells (Lucigen). Bacterial cultures were incubated at 37^{o}C in Terrific Broth (Gibco) supplemented with 100µg/ml ampicillin until OD600 (optical density at 600nm) reached ~0.7. The protein expression of His6-PSF was then induced with 1mM IPTG. Following continued incubation at 37^{o}C for 4 hours, bacterial cultures were collected in 50 ml tubes. Following centrifugation (6000 x g) at 4^{o}C for 15 min, supernatant was discarded and cell pellets were washed twice with 1x PBS. The washed cell pellets were resuspended in 1.4 ml of non-denaturing lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole pH8.0) supplemented with 1% (v/v) protease inhibitor cocktail for mammalian cell extracts (Sigma-Aldrich), 1% (v/v) phosphatase inhibitor cocktail 1 (Sigma-Aldrich) and 1mM PMSF. Cell lysis was then performed using probe-sonication (twice, 30 seconds each time) on ice. Following centrifugation, His6-PSF protein was isolated from supernatant using Ni-NTA spin kit (Qiagen) according to manufacturer's protocols. Briefly, the supernatant was loaded onto a Ni-NTA spin column (Qiagen),

spun and washed 3X with 600µl of wash buffer (50mM NaH_2PO_4, 300mM NaCl, 20mM imidazole pH8.0). Flow-through was discarded after each centrifugation and the bound proteins were eluted with elution buffer containing 250 to 500mM imidazole. Following elution, eluates were desalted with Microcon YM-10 column (Millipore) and resuspended in non-denaturing IP buffer supplemented with 1% (v/v) protease inhibitor cocktail for mammalian cell extracts (Sigma-Aldrich), 1% (v/v) phosphatase inhibitor cocktail 1 (Sigma-Aldrich) and 1mM PMSF.

2.5 Immunodepletion

Immunodepletion of PSF in protein extracts was performed with Dynabeads Protein G for Immunoprecipitation (Invitrogen) and anti-PSF/SFPQ antibody (mouse monoclonal, [B92], Abcam). 50 µl of the magnetic beads were resuspended in 200 µl of IP buffer and incubated with 20 µg of anti-SFPQ antibody at room temperature for 30 min. Following incubation, the reaction buffer was removed. The antibody-coupled beads were then washed 3X with 200µl of IP buffer (supernatant was discarded each time) and incubated with 100µl (~1000 µg) of protein extract (either HEK293 whole cell extract or 293-HDAg-S nuclear extract) at 4^{o}C for 60 min. The protein extract solution was separated from the beads and subjected to another round of immunodepletion by following the same procedure as above. To quantify the immunodepleted protein exact, Western blotting analysis was performed with anti-PSF/SFPQ antibody (mouse monoclonal, [B92], Abcam) and anti-RNAP II CTD repeat YSPTSPS antibody (mouse monoclonal, [8WG16], Abcam).

2.6 Western blotting

Protein samples were mixed with equal volume of 2 X Laemmli sample buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125M pH6.8 Tris-HCl), heated at 95°C for 5 min and then cooled on ice for 2 min. Proteins were then resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (8%) in running buffer (25mM Tris-HCl, 200mM Glycine, 0.1% SDS). Following electrophoresis, proteins in the SDS-PAGE gel were transferred overnight at 4^{o}C and 40 volts to a PVDF membrane in transfer buffer (48mM Tris, 39mM glycine, 20% methanol, 0.037% SDS; not require to adjust pH). Following transfer, the PVDF membrane was blocked for 1 hour at room temperature using 5% skimmed milk in 1x Tris-buffered saline (0.8% NaCl, 0.24% Tris, pH7.6) containing 0.1% (v/v) Tween (TBST) on a rocking shaker. The blocked membrane was then incubated for 1 hour at room temperature in a dilute solution of primary antibody (1⁄1000 antibody and 3% skimmed milk in 1x TBST) on the shaker. Following the incubation with the primary antibody, the membrane was rinsed twice with 1x TBST and washed 4 times (10min/time) with 1x TBST on the shaker. The membrane was then incubated for 1 hour at room temperature in a dilute solution of Horseradish Peroxidase (HRP) conjugated species-specific secondary antibody (1⁄20000 antibody and 3% skimmed milk in 1x TBST). The membrane was then rinsed and washed with 1x TBST by following the same procedure as above. Following the last wash, the membrane was incubated for 5 min in the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific). An X-ray film was then exposed against the membrane. Antibodies used for Western blotting were as follows: (1) Primary antibodies: RNAPII was

detected with anti-RNAP II CTD repeat YSPTSPS antibody (mouse monoclonal, [8WG16], Abcam); endogenous PSF was detected with anti-PSF/SFPQ antibody (mouse, monoclonal, [B92], Abcam); c-Myc epitope-tagged wild type and mutant PSF proteins (Δ+/-,ΔRGG, ΔRRM1, ΔRRM1+2, ΔP or ΔN) were detected with anti-Myc epitope antibody (mouse monoclonal, [9E10], Abcam) in most cases, except that anti-Myc epitope antibody (rabbit polyclonal, ab11917, Abcam) was used to detect ΔP and ΔN proteins in co-immunoprecipitations and immunoprecipiations. (2) Secondary antibodies: rabbit anti-mouse IgG H&L (Immunoglobulin G heavy chains and light chains) (HRP) antibody (polyclonal, ab6728, Abcam) was used to detect mouse primary antibody; goat anti-rabbit IgG H&L (HRP) antibody (polyclonal, ab6721, Abcam) was used to detect rabbit primary antibody.

2.7 Protein co-immunoprecipitation assay

Co-immunoprecipitation experiments were carried out using Dynabeads Protein G for Immunoprecipitation (Invitrogen) with a modified manufacturer's protocol. 20µl of the magnetic beads were resuspended in 200µl of IP buffer supplemented with 20µl of BSA (1mg/ml in H_2O) and incubated with 2µg of antibody at room temperature for 30min. The antibodies were as follows: anti-RNAP II CTD repeat YSPTSPS antibody (mouse monoclonal, [8WG16], Abcam), anti-Myc epitope antibody (mouse monoclonal, [9E10], Abcam) or goat anti-mouse IgG H&L antibody (polyclonal, ab6708, Abcam, used as negative control). The antibody-coupled beads were then isolated, washed one time with 200µl of IP buffer and isolated from washing buffer. The beads were then incubated with protein extract at room temperature for 1 hour in 200µl of IP buffer supplemented

with 20µl of BSA (1mg/ml in H_2O). The protein extracts were as follows: HeLa NE (Accurate Chemical & Scientific Co.), HEK293 WCE or WCE of HEK293 expressing c-Myc epitope-tagged wild type or deletion mutant PSF proteins. 200µg (total proteins) of protein extract was used in most cases, except that 2000µg (total proteins) of ΔP or ΔN protein extract was used. 1µl of Ribonuclease A (RNase A, 10mg/ml, Qiagen) was added to some samples to eliminate the RNA-protein interactions. Following incubation, the beads were isolated, washed 4X with 200µl of IP buffer (supernatant was discarded after each wash). Following the last wash and isolation, bound proteins were eluted with 20µl of 1x laemmli sample buffer and detected using Western blotting. The procedure of immunoprecipitation for investigating the accessibility of c-Myc epitope tag of recombinant PSF proteins with anti-Myc epitope antibody was same as above (see Figure 3.7).

2.8 RNA co-immunoprecipitation assay

20µl of the magnetic beads were resuspended in 200µl of IP buffer supplemented with 20µl of BSA (1mg/ml in H_2O) and incubated with 2µg of antibody at room temperature for 30min. The antibody-coupled beads were isolated, washed one time with 200µl of IP buffer and then isolated from washing buffer. The beads were then incubated at 4°C for 4 hours to overnight with ~3 pmoles of ^{32}P-radiolabeled R199G RNA and 200µg of total protein extract in 150µl of IP buffer supplemented with 20µl of BSA (1mg/ml in H_2O) and 40 units of human placental RNase inhibitor (Biobasic). P11.60 RNA (150 pmoles) or unlabeled R199G RNA (150 pmoles) were added as competitors to co-immunoprecipitation reactions, where indicated. Following incubation, the beads

were washed 3 x with 500µl of RIPA buffer (50mM Tris–HCl, pH 7.5, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.05% SDS, 1mM EDTA, 150mM NaCl) and then 1 x with 500µl of IP buffer (supernatant was discarded after each wash). Following the last wash and isolation, bound RNAs were eluted with 20µl of acrylamide loading dye (95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, 0.5mM EDTA), resolved by Urea-PAGE (10% polyacrylamide (19:1, acrylamide:bis-acrylamide) / 7M urea) and visualized by phosphorimaging.

2.9 Cell nuclear extraction

Cell nuclear extraction was performed using CelLytic™ NuCLEAR™ Extraction Kit (Sigma-Aldrich) according to manufacturer's protocol with small modifications. About $1x10^6$ HEK293 or 293-HDAg-S (Chang *et al*., 2005) cells were seeded in 100mm x 20mm cell culture dishes (Corning) containing 15ml of culture medium, and incubated until cells reached ~90% confluency (~2days) at 37^oC in an incubator with 5% CO_2. The culture medium of 293-HDAg-S cells was DMEM-High glucose medium (Invitrogen) supplemented with 10% FBS, selection antibiotics (200µg/ml hygromycin B and 5µg/ml blasticidin, Invitrogen) for maintaining the HDAg-S and TET repressor genes (Chang *et al*., 2005). The expression of HDAg-S was induced with 10µg/ml TET. The culture medium of HEK293 cells was DMEM-High glucose medium supplemented with 10% FBS. Following incubation, media in dishes were aspirated and adherent cells were rinsed twice with 1ml of 1xPBS. The cells in each dish were then scraped into a 1.5ml eppendorf tube with 1ml of 1xPBS. Following centrifugation at 450 x g and 4^oC for 5 min, supernatant was discarded, 100µl of cell pellet was resuspended in 500µl of 1x Lysis Buffer (hypotonic;

supplemented with 1mM DTT and 1% (v/v) protease inhibitor cocktail, Sigma-Aldrich) and incubated for 15 min on ice. IGEPAL CA-630 (final concentration 0.6%) was then added into the mixture. Following vortex and centrifugation, supernatant was removed and the pellet of cell nuclei was extracted using ~ 70µl of extraction buffer (20 mM HEPES pH7.9, 1.5mM $MgCl_2$, 0.42M KCl, 0.2mM EDTA and 25% (v/v) glycerol, 1mM DTT, 1% (v/v) protease inhibitor (Sigma)). The mixture was agitated at medium speed and 4^{o}C for 30 min and then centrifuged at 20000 x g and 4^{o}C for 5 min. The supernatant containing cell nuclear extract (NE) was divided into several aliquots, snap-frozen with liquid nitrogen and stored at -80^{o}C. Concentration of total proteins in cell NE was measured by Bradford assay kit (Bio-rad).

2.10 *In vitro* transcription assay

150 pmoles of RNA templates were incubated with 75µg of 293-HDAg-S NE in 100µl of *in vitro* transcription buffer (20mM HEPES pH 7.9, 50mM KCl, 0.2 mM EDTA, 3.0mM $MgCl_2$, 0.15mM DTT, 20% glycerol, 1.5mM each NTP, 0.5unit/µl human placental RNase inhibitor (Biobasic)). Alternatively, HEK293 NE, PSF-depleted 293-HDAg-S NE or PSF-depleted 293-HDAg-S NE supplemented with purified His6-PSF was used. Transcription reactions were carried out at 30^{o}C for 30 min, and purified with phenol-chloroform extraction and ethanol-precipitation. Purified RNA was resuspended in 20µl of H_2O and subjected to reverse transcription (RT) using the SuperScript® III Reverse Transcriptase (Invitrogen), according to manufacturer's instruction. During the RT reaction, a radiolabeled 'Y-primer' (see Appendix I) was used to specifically recognize transcription product of Y-R199G RNA template (see Figure 3.9). Following the RT

reaction, RNA was digested with 1µl of RNase A (10mg/ml, Qiagen) for 30 min at 37^{o}C. Radiolabeled cDNA corresponding to transcription product was resolved by Urea-PAGE (10% polyacrylamide (19:1, acrylamide:bis-acrylamide) / 7M urea) and visualized by phosphorimaging.

Chapter 3. Results

3.1 PSF might stimulate RNAPII interaction with R199G

Based on the findings that RNAPII interacts with R199G (Greco-Stewart *et al*., 2007), PSF directly binds to R199G (Greco-Stewart *et al*., 2006), and PSF binds to the CTD tail of RNAPII (Emili *et al*., 2002), PSF might facilitate RNAPII interaction with R199G. In order to verify this hypothesis, I utilized an anti-RNAPII (8WG16) antibody to co-immunoprecipitate ^{32}P-radiolabeled R199G from HEK293 WCE, which was immunodepleted of endogenous PSF, or from PSF-depleted HEK293 WCE complemented with recombinant PSF. If PSF is able to facilitate RNAPII interaction with R199G, depletion of PSF from the WCE should reduce the level of the co-immunoprecipitated radiolabeled R199G, while adding recombinant PSF into the PSF-depleted WCE should restore this level. Therefore, to perform this experiment, I firstly expressed an N-terminal-hexahistine-tagged PSF (His6-PSF) in *E.coli*, and then isolated the recombinant protein by Ni^{2+} affinity chromatography. Quality control by Coomassie blue staining following SDS-PAGE showed that the isolation process removed most *E.coli* endogenous proteins (Figure 3.1 A(1)); Western blotting with an anti-PSF/SFPQ (B92) antibody showed that His6-PSF was enriched in the final protein fraction (Figure 3.1 A(2)). However, some other bands (~75kDa and ~50kDa) were also present, possibly representing premature translational termination products which containing the N-terminal-hexahistine-tag because (1) in my experiment the N-terminal-hexahistine-tag of the recombinant proteins bound to the Ni^{2+} on the column, and then the proteins were eluted with imidazole buffer, and (2) the anti-PSF/SFPQ (B92) antibody was

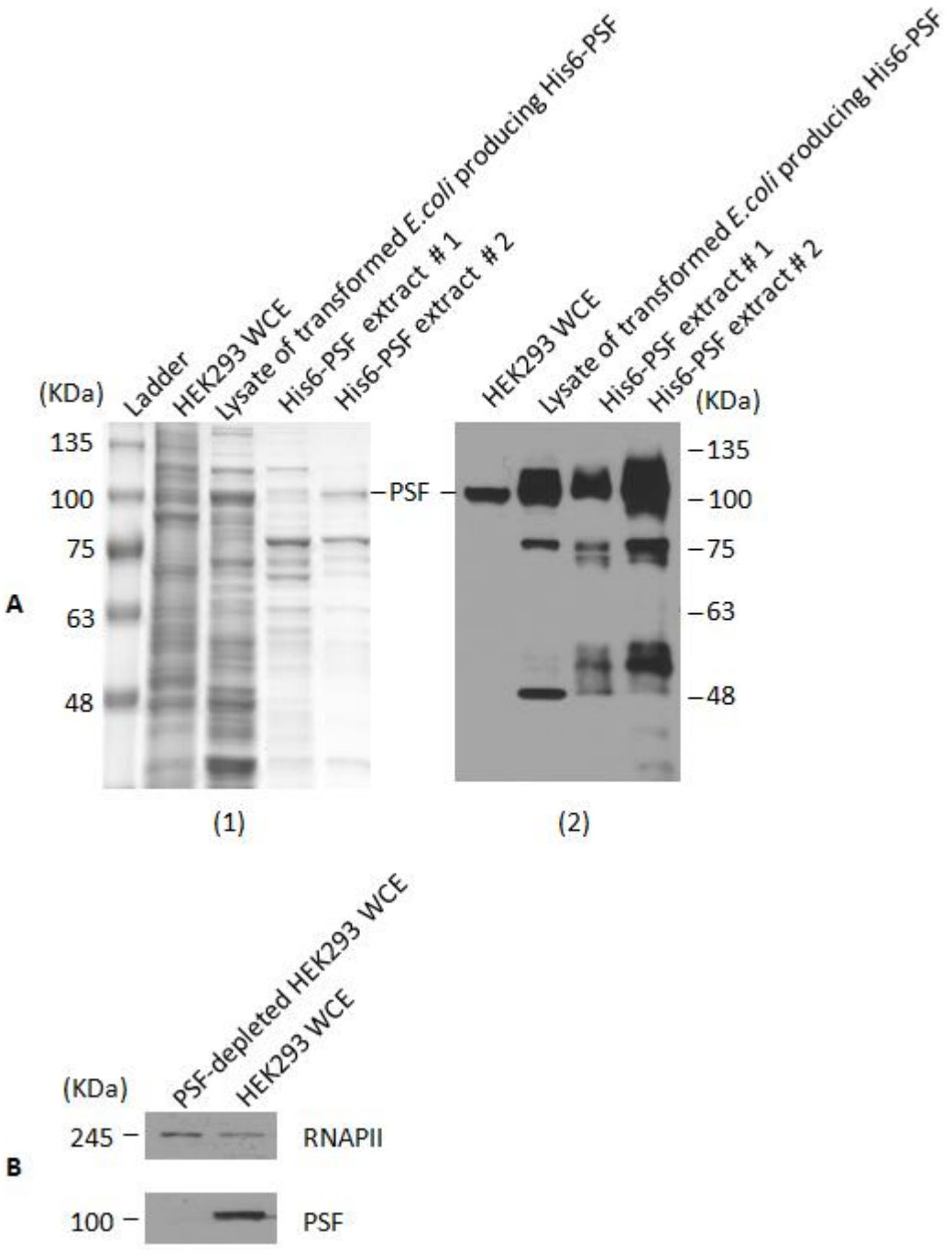

A
(KDa)
Ladder
HEK293 WCE
Lysate of transformed E.coli producing His6-PSF
His6-PSF extract # 1
His6-PSF extract # 2
135
100
75
63
48
PSF
(1)
HEK293 WCE
Lysate of transformed E.coli producing His6-PSF
His6-PSF extract # 1
His6-PSF extract # 2
(KDa)
135
100
75
63
48
(2)
B
(KDa)
PSF-depleted HEK293 WCE
HEK293 WCE
245
RNAPII
100
PSF

Figure 3.1 Preparation of N-terminal-hexahistidine-tagged PSF (His6-PSF) and PSF-depleted HEK293 WCE

(A) Quality control of recombinant His6-PSF protein. Lysate of *E.coli* producing His6-PSF was prepared and the recombinant protein was isolated using Qiagen Ni-NTA spin columns. Two His6-PSF extracts, #1 and #2, were eluted from Ni-NTA columns with 250mM and 500 mM imidazole buffer, respectively. A sample of HEK293 WCE (positive control), *E.coli* lysate containing His6-PSF and the two His6-PSF protein extracts were resolved by SDS-PAGE. The SDS-PAGE-resolved proteins were (1) visualized by Coomassie blue staining or (2) identified by Western blotting using an anti-PSF/SFPQ (B92) antibody. **(B)** Quality control of PSF-immunodepleted HEK293 NE. HEK293 WCE was immunodepleted of PSF with anti-PSF/SFPQ antibody-coupled protein G beads in co-immunoprecipitation buffer. Proteins in PSF-depleted HEK293 WCE or untreated WCE were resolved by SDS-PAGE and identified by Western blotting with an anti-PSF/SFPQ (B92) antibody or an anti-RNAP II (8WG16) antibody.

reported binding to the N-terminus of PSF (Shav-Tal *et al*., 2000; Shav-Tal *et al*., 2001); therefore, the Western blotting should detect the N-terminal part of the recombinant proteins which included the full length His6-PSF and the proteins of premature translational termination. Secondly, I utilized anti-PSF/SFPQ (B92) antibody to immunodeplete endogenous PSF in the HEK293 WCE. Western blotting showed that PSF had been depleted while RNAPII was still present in the WCE after immunodepletion (Figure 3.1 B); therefore, this PSF-depleted WCE could be used in the co-immunoprecipitation assay.

For the next step, anti-RNAP II (8WG16) or anti-IgG (negative control) antibody was coupled to the protein G magnetic beads to co-immunoprecipitate ^{32}P-radiolabeled R199G from either untreated or PSF-depleted HEK293 WCE. His6-PSF extract was added into some reactions in PSF-depleted WCE. An excess (50-fold) of either unlabeled R199G (specific RNA competitor) or P11.60 (non-specific RNA competitor) was added into reaction mixtures of the competition assays. P11.60 RNA is a small peach latent mosaic viroid (PLMVd)-derived RNA, which folds into a hairpin secondary structure (Pelchat and Perreault, 2004). It is an excellent competitive RNA to test for non-specific affinity of the proteins to double-stranded RNA or RNA possessing a stem–loop secondary structure. After extensive washing of beads, bound RNAs were eluted, then subjected to Urea-PAGE and visualized by autoradiography. Figure 3.2 (A) shows that radiolabeled R199G was co-immunoprecipited from untreated HEK293 WCE containing endogenous PSF (lane2); moreover, the radiolabeled R199G was not detected when unlabeled R199G competitor was present (lane5), but the appearance of the radiolabeled R199G band was

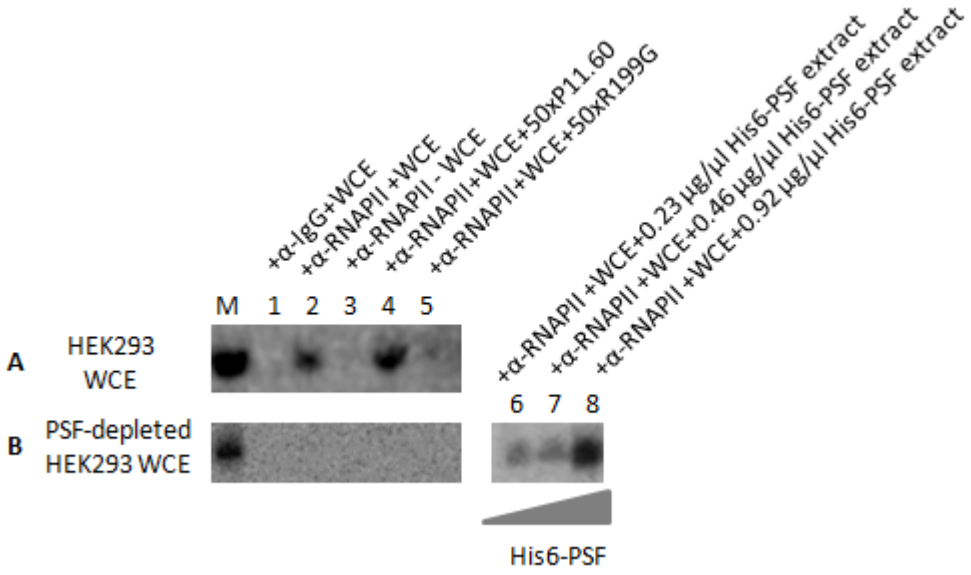

+α-IgG+WCE
+α-RNAPII +WCE
+α-RNAPII - WCE
+α-RNAPII+WCE+50xP11.60
+α-RNAPII+WCE+50xR199G
+α-RNAPII +WCE+0.23 µg/µl His6-PSF extract
+α-RNAPII +WCE+0.46 µg/µl His6-PSF extract
+α-RNAPII +WCE+0.92 µg/µl His6-PSF extract
M 1 2 3 4 5
6 7 8
A
HEK293 WCE
B
PSF-depleted HEK293 WCE
His6-PSF

Figure 3.2 PSF might stimulate RNAPII interaction with R199G

Co-immunoprecipitation assays of radiolabeled R199G with RNAPII were performed using anti-RNAPII antibody-coupled protein G beads. Row A: untreated HEK293 WCE was used. Row B: PSF-depleted HEK293 WCE was used and His6-PSF protein extract was added where indicated. Lane M contains a 1:100 dilution of radiolabeled R199G. Reaction components of the assays were shown, as indicated. The protein concentration shown in the figure was total protein concentration of His-PSF extract in each co-immunoprecipitation reaction mixture.

not affected by adding the non-specific P11.60 RNA (lane 4). The result of Figure 3.2 (A) confirmed the previously-reported specific interaction of RNAPII with R199G (Greco-Stewart *et al*., 2007), and validated the co-immunoprecipitation procedure. Figure 3.2 (B) shows that radiolabeled R199G could not be detected in the co-immunoprecipitation assay in PSF-depleted HEK293 WCE; however, R199G was clearly identified when 0.23 – 0.92 µg/µl (35 µg/ 150 µl to 140 µg/ 150 µl) of His6-PSF extract was added back (lane6, 7, 8), and the level of co-immunoprecipitated R199G increased with the addition of more recombinant PSF. Therefore, **PSF might stimulate RNAPII interaction with R199G**. Also, negative control co-immunoprecipitation was performed by using anti-IgG (goat polyclonal to mouse IgG H&L, Abcam) antibody-coupled protein G beads incubating with radiolabeled R199G and WCE (lane 1, Figure 3.2), or using anti-RNAPII (mouse monoclonal, 8WG16) antibody-coupled beads with radiolabeled R199G but without adding WCE (lane 3, Figure 3.2). No radiolabeled R199G was detected in negative control tests.

3.2 Identification of PSF domains needed for the interaction with RNAPII

Previous reports showed that PSF interacts with both R199G and the CTD tail of RNAPII (Greco-Stewart *et al*., 2007; Emili *et al*., 2002). Moreover, my co-immunoprecipitation results (Section 3.1) suggested that PSF might facilitate RNAPII interaction with R199G. However, we still do not know which regions of PSF are involved in the interactions with R199G and RNAPII. Therefore, in my thesis, a series of N-terminal Myc epitope-tagged deletion derivatives of PSF was used to determine which domains of PSF interact with RNAPII and HDV RNA. The derivatives of PSF were constructed to specifically remove the RGG box (ΔRGG, residues 9 to 27), RRM1 (ΔRRM1, residues 298 to 365),the basic/acidic rich domain +/- (Δ+/-, residues 575 to 611), the entire N terminal half of PSF (ΔN, residues 9 to 264), the proline-rich region (ΔP, residues 26 to 267) and both RRMs (ΔRRM1+2, residues 298 to 429) (Rosonina *et al*., 2005) (Figure 3.3). In order to determine which domain of PSF is involved in the interaction between PSF and RNAPII, I utilized WCE of HEK293 cells expressing above Myc epitope-tagged mutants and wild type (wt) PSF protein to perform a series of co-immunoprecipitation experiments followed by Western blot assays. If the specific domain of PSF is important for the interaction between PSF and RNAPII, the corresponding PSF mutant protein will not interact with RNAPII.

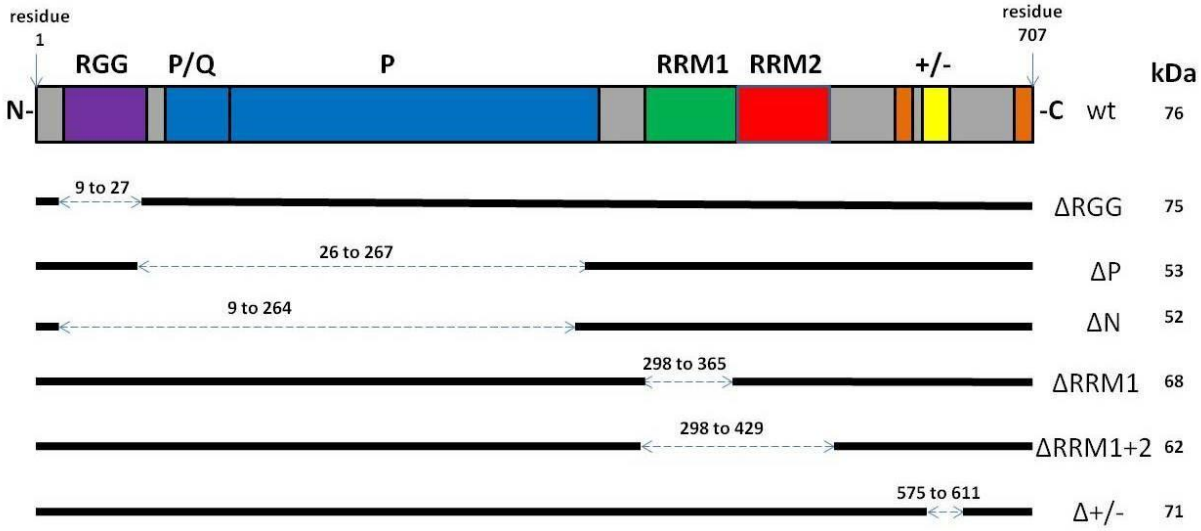

residue
1
residue
707
RGG
P/Q
P
RRM1
RRM2
+/-
kDa
N-
-C
wt
76
9 to 27
ΔRGG
75
26 to 267
ΔP
53
9 to 264
ΔN
52
298 to 365
ΔRRM1
68
298 to 429
ΔRRM1+2
62
575 to 611
Δ+/-
71

Figure 3.3: Schematic representation of the domain arrangement of wild type and deletion mutants of PSF.

A scheme of the domain arrangement of wild type PSF is shown at the top. Multiple motifs of PSF are shown as indicated: the arginine glycine glycine repeat motif (RGG), the proline/glutamine rich region (P/Q), the proline rich region (P), the highly charged basic and acidic residues rich region (+/-) and two RNA recognition motifs (RRM1 and RRM2). The construction of each deletion mutant is shown below the scheme of wild type PSF. The deletion part of each mutant is shown as a dash arrow. Above the dash arrows are the amino acid residue numbers. Molecular weights of proteins (not including Myc epitope tag, ~1.2kDa) are shown at the right side of the figure. The molecular weights are calculated by using human PSF sequence from http://www.ncbi.nlm.nih.gov/protein /NP_005057.1 and online protein molecular weight calculator: http:// www.bioinformatics.org/sms/ prot_mw.html. (Rosonina *et al*., 2005)

Because both RNAPII and the Myc epitope-tagged PSF proteins must be present in the WCEs, which are used for the co-immunoprecipitation experiments to examine the PSF-RNAPII interaction, I performed Western blotting assays to verify the presence of the above proteins in these WCEs. Western blotting assays utilizing an anti-RNAPII antibody or anti-Myc epitope antibody indicated that the transfected HEK293 WCE contained RNAPII and the Myc epitope-tagged PSF proteins (Figure 3.4). The apparent sizes of the wild type PSF, Δ+/-, ΔRGG, ΔRRM1 and ΔRRM1+2 mutant PSF proteins on Western blot was between ~100 kDa and ~80 kDa according to the size marker, while the apparent sizes of ΔP and ΔN proteins were ~60 kDa and ~55 kDa, respectively (Figure 3.4). These sizes agree with the reported results of each recombinant PSF protein (Rosonina *et al*., 2005). Figure 3.4 also shows that the expression levels of the ΔP and ΔN mutant proteins were much lower (~10-fold determined by densitometry quantification using ImageJ software) than those of the other recombinant PSF proteins; therefore, more whole cell extract containing ΔP or ΔN mutant protein was used for immunoprecipitation.

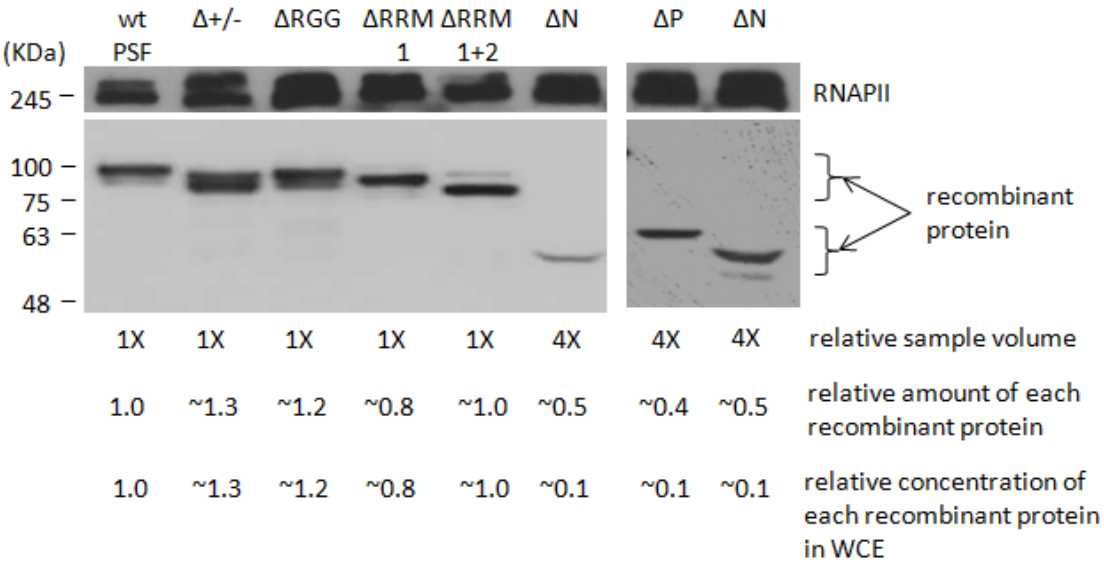

wt PSF
Δ+/-
ΔRGG
ΔRRM 1
ΔRRM 1+2
ΔN
ΔP
ΔN
(KDa)
245
100
75
63
48
RNAPII
recombinant protein
1X 1X 1X 1X 1X 4X 4X 4X
relative sample volume
1.0 ~1.3 ~1.2 ~0.8 ~1.0 ~0.5 ~0.4 ~0.5
relative amount of each recombinant protein
1.0 ~1.3 ~1.2 ~0.8 ~1.0 ~0.1 ~0.1 ~0.1
relative concentration of each recombinant protein in WCE

Figure 3.4 Expression of c-Myc epitope-tagged wild type and mutant PSF proteins in transfected HEK293 cells.

Western blotting using anti-RNAPII and anti-Myc epitope antibodies were performed to confirm the expression of the RNAPII and recombinant PSF proteins, respectively. The relative loaded volumes of WCEs are shown under the images of the Western blotting assays. Either 2 µl (1x) or 8 µl (4X) of WCE was loaded on the SDS-PAGE gel. The relative loaded amount of each of recombinant c-Myc epitope-tagged proteins was determined by densitometry quantification using ImageJ software. All other protein bands were compared with the protein band of wild type PSF in the first sample lane. Then, the relative concentration of each protein was calculated by dividing the relative loaded protein amount with the relative loaded sample volume. The results of calculation are shown under the images of the Western blotting. The protein expression levels of wild type PSF, Δ+/-, ΔRGG, ΔRRM1 and ΔRRM1+2 were similar, but those of ΔN and ΔP were ~10-fold lower.

Before I started to identify the interaction between Myc epitope-tagged mutant PSF and RNAPII, I performed preliminary experiments to verify my co-immunoprecipitation protocol. Here, I utilized Protein G magnetic beads, which were coupled to an anti-RNAPII antibody or anti-IgG antibody (negative control) to immunoprecipitate the RNAPII complex from HeLa NE (Figure 3.5A). After extensive washing of beads, bound proteins were resolved by SDS-PAGE and identified by Western blotting assays with an anti-PSF/SFPQ antibody. Because PSF binds to nucleic acids (Shav-Tal and Zipori, 2002), I added 1µl of RNase A (10mg/ml, Qiagen) in a co-immunoprecipitation assay using anti-RNAPII antibody in order to eliminate RNA-protein interactions. Figure 3.5A shows that anti-RNAPII antibody is able to co-immunoprecipitate PSF with RNAPII but anti-IgG antibody cannot co-immunoprecipitate PSF. The Western blot shows a protein band corresponding to PSF, which migrates at ~100kDa according to the size marker (Figure 3.5A). This is consistent with the reported apparent molecular weight of PSF determined by SDS-PAGE (Patton *et al*., 1993). The result of negative control test with anti-IgG antibody verified the co-immunoprecipitation experiment was specific. The results of above tests validated my co-immunoprecipitation protocol, except that the co-immunoprecipitated PSF bands were slightly lower than PSF band in the input due to an unknown reason.

Next, I tested two possible ways to perform the co-immunoprecipitation assays: by utilizing an anti-Myc epitope antibody to co-immunoprecipitate RNAPII with the c-Myc epitope-tagged PSF proteins, or by utilizing anti-RNAPII antibody to co-immunoprecipitate each of the recombinant PSF proteins with RNAPII from the

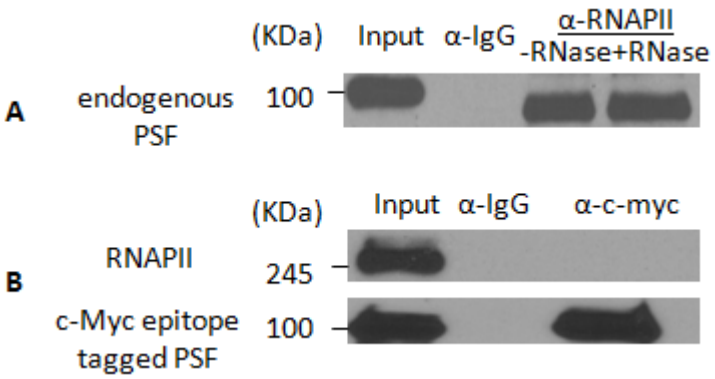

A
(KDa)
Input
α-IgG
α-RNAPII
-RNase
+RNase
endogenous PSF
100
B
(KDa)
Input
α-IgG
α-c-myc
RNAPII
245
c-Myc epitope tagged PSF
100

Figure 3.5 Using anti-RNAPII or anti-Myc antibody to co-immunoprecipitate PSF or RNAPII.

(A) Co-immunoprecipitation assay using anti-RNAPII antibody was performed in HeLa NE. Protein G magnetic beads coupled to either anti-IgG antibody (negative control) or anti-RNAPII CTD antibody were incubated with HeLa NE in IP buffer. 1µl of RNase A (10mg/ml) was added in a co-immunoprecipitation assay using anti-RNAPII CTD antibody to eliminate RNA-protein interactions. Bound proteins were analyzed by Western blotting using anti-PSF/SFPQ antibody.
(B) Co-immunoprecipitation assay was performed in a similar way as (A), but with anti-Myc antibody and protein extract of c-Myc epitope-tagged PSF. Bound proteins were analyzed by Western blotting using an anti-RNAPII CTD or anti-Myc antibody.

HEK293 WCEs. To verify the first method, I utilized protein G magnetic beads, which were coupled to anti-Myc epitope antibody or anti-IgG antibody to co-immunoprecipitate RNAPII and c-Myc epitope-tagged wild type PSF in a HEK293 WCE. Then I performed Western blotting using an anti-RNAPII antibody or anti-c-Myc antibody to detect RNAPII or c-Myc epitope-tagged PSF, respectively. Results are shown in Figure 3.5B. Both RNAPII and c-Myc epitope-tagged PSF were detectable in the input lane, but were not detectable in the lane corresponding to co-immunoprecipitation utilizing anti-IgG antibody (negative control) (Figure 3.5B). Following immunoprecipitation with anti-Myc epitope antibody, c-Myc epitope-tagged PSF was detected, but RNAPII was not detected (Figure 3.5B). Therefore, anti-Myc epitope antibody could immunoprecipitate the c-Myc epitope-tagged PSF, but was not able to co-immunoprecipitate RNAPII. However, previous experiments determined that the anti-RNAPII antibody could co-immunoprecipitate RNAPII and PSF, and that the anti-PSF/SFPQ (B92) antibody could co-immunoprecipitate PSF and RNAPII (Greco-Stewart, 2009). Therefore, my results suggest **that RNAPII might bind to the region close to the N-terminus of PSF, and the binding of anti-Myc epitope antibody on the N-terminal c-Myc epitope-tag might block the interaction of RNAPII with Myc epitope-tagged PSF.**

Based on above results, I decided to take the second approach to perform co-immunoprecipitation assays. I utilized protein G magnetic beads, which were coupled to anti-RNAPII antibody (mouse, [8WG16], Abcam) or anti-IgG antibody (goat anti-mouse, ab6708, Abcam, negative control) to co-immunoprecipitate RNAPII and each of the c-Myc epitope-tagged mutant and wild type PSF proteins in the HEK293 WCEs.

RNase A was added in the samples of one group of co-immunoprecipitation reactions utilizing anti-RNAPII antibody, in order to remove the RNA in the WCEs and eliminate the RNA-protein interactions. In addition, when I was doing co-immunoprecipitation, I added 10-fold more HEK293 WCE containing ΔP or ΔN protein than other WCEs in order to provide approximately equal amount of ΔP or ΔN protein. Following the co-immunoprecipitation, I performed Western blotting using anti-Myc epitope antibody to detect the proteins. During western blotting, the c-Myc epitope-tagged wild type PSF and the Δ+/-,ΔRGG, ΔRRM1, ΔRRM1+2 mutant proteins were detected with mouse anti-Myc primary antibody (monoclonal, [9E10], Abcam), but the ΔP and ΔN proteins were detected with rabbit anti-Myc primary antibody (polyclonal, ab11917, Abcam). Here I am explaining why I used rabbit anti-Myc antibody to detect ΔP and ΔN proteins. When mixing the IgG antibody-coupled protein G beads with the Laemmli sample buffer, the disulfide bonds connecting heavy and light chains of antibody will be reduced by the 2-mercaptoethanol in the Laemmli buffer. Then, these IgG heavy chain (~53kDa) and light chain (~25kDa) will be eluted together with the immunoprecipitated proteins. Because the molecular weights of ΔP (~53kDa) and ΔN (~52kDa) proteins and the IgG heavy chain are very close, it might be difficult to separate these proteins by SDS-PAGE. Therefore, I used a rabbit primary antibody to bind to the Myc epitope-tag of ΔP or ΔN protein. Then, I used a goat secondary antibody to bind to the rabbit primary antibody, but not bind to the mouse IgG heavy chain. By this way, the Myc epitope-tagged ΔP and ΔN proteins were detected, but the protein band of mouse IgG heavy chain did not appear. Western blotting results are shown in Figure 3.6. All target proteins were present

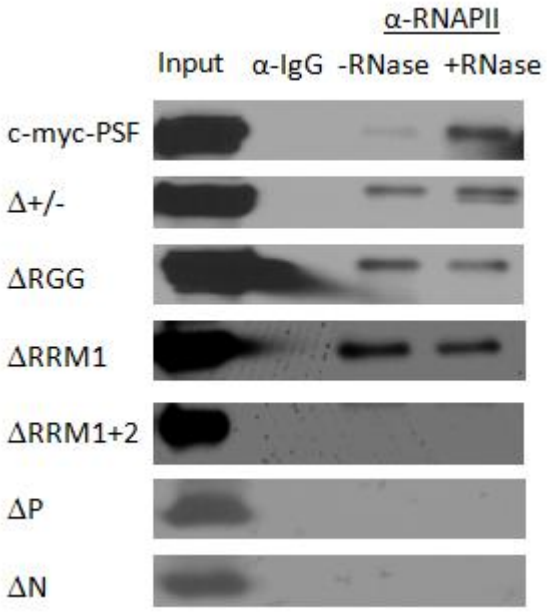

α-RNAPII
Input
α-IgG
-RNase
+RNase
c-myc-PSF
Δ+/-
ΔRGG
ΔRRM1
ΔRRM1+2
ΔP
ΔN

Figure 3.6 Identification of the PSF domains required for the interaction with RNAPII.

Co-immunoprecipitations using an anti-RNAPII antibody were performed on the protein extracts of HEK 293 cells transfected with each of the c-Myc epitope-tagged PSF constructs. For some assays, RNase A was added in order to eliminate the RNA-protein interactions. The co-immunoprecipitation with anti-IgG antibody was used to assess of the specificity of the interaction. Following the co-immunoprecipitations, bound proteins were subjected to SDS-PAGE and detected by Western blotting using anti-Myc epitope antibody. Identified proteins are shown in each lane, as indicated.

in the WCEs (input lane), but not present in the negative control (anti-IgG). By treating with RNase A, more c-Myc epitope-tagged wild type PSF proteins were co-immunoprecipitated, probably because RNase A eliminated the RNA-protein interactions, and thus more Myc epitope-tagged PSF interacted with the RNAPII. In both groups of co-immunoprecipitation assays (anti-RNAPII) with or without RNase A treatment, c-Myc epitope-tagged wild type PSF, Δ+/-, ΔRGG, ΔRRM1 mutant proteins were detectable, but ΔRRM1+2, ΔP and ΔN proteins were not detectable. Thus I suggest that (1) the **proline-rich region (PRR, residues 26 to 267) of PSF is needed for the interaction between PSF and RNAPII**, which is consistent with the above deduction from Figure 3.5. Because the PRR is close to the N-terminus of PSF, the binding of anti-Myc epitope antibody on the N-terminal c-Myc epitope-tag might hinder RNAPII binding to the PRR; (2) **RRM2 (residues 375 to 429) of PSF is needed for the interaction between PSF and RNAPII**, consistent with the previous report that both c-Myc epitope-tagged ΔRRM2 (residues 375 to 429) mutant and ΔRRM1+2 mutant did not interact with purified recombinant CTD of RNAPII (Rosonina *et al*., 2005).

3.3 Identification of PSF domains required for the interaction with R199G

Co-immunoprecipitation assays using anti-Myc epitope antibody were performed to determine which domains of PSF interact with R199G. Because the c-Myc epitope tag of a properly folded recombinant protein under native conditions might be buried inside the protein and not be accessible to the anti-Myc epitope antibody, it is necessary to verify the accessibility of the antibody to the c-Myc epitope tag before doing above co-immunoprecipitation experiments. For this reason, I coupled anti-Myc epitope antibody or anti-IgG antibody to the protein G magnetic beads, and then performed immunoprecipitations in WCEs from HEK 293 cells expressing the c-Myc epitope-tagged wild type PSF and each of the mutant proteins (Δ+/-, ΔRGG, ΔRRM1, ΔRRM1+2, ΔP and ΔN). Because protein expression levels of ΔP and ΔN were ~10-fold lower than those of other proteins, 10-fold more HEK293 WCE containing ΔP or ΔN protein than other WCEs was used to acquire approximately equal amount of ΔP or ΔN protein. Beads were then washed and bound proteins were eluted and identified by Western blotting using anti-Myc epitope antibody. During western blotting, the ΔP and ΔN proteins were detected with rabbit anti-Myc primary antibody while the other recombinant proteins were detected by mouse anti-Myc primary antibody (the reason for using rabbit anti-Myc antibody to detect ΔP and ΔN proteins has been explained in the page 74). Figure 3.7 shows the results of immunoprecipitation assays. All c-Myc epitope-tagged proteins in the input lanes were detectable, while they could not be found in the lanes corresponding to the negative control (anti-IgG). Following immunoprecipitation, wild type, Δ+/-, ΔRGG, ΔRRM1 and ΔRRM1+2 recombinant PSF proteins were detectable, but

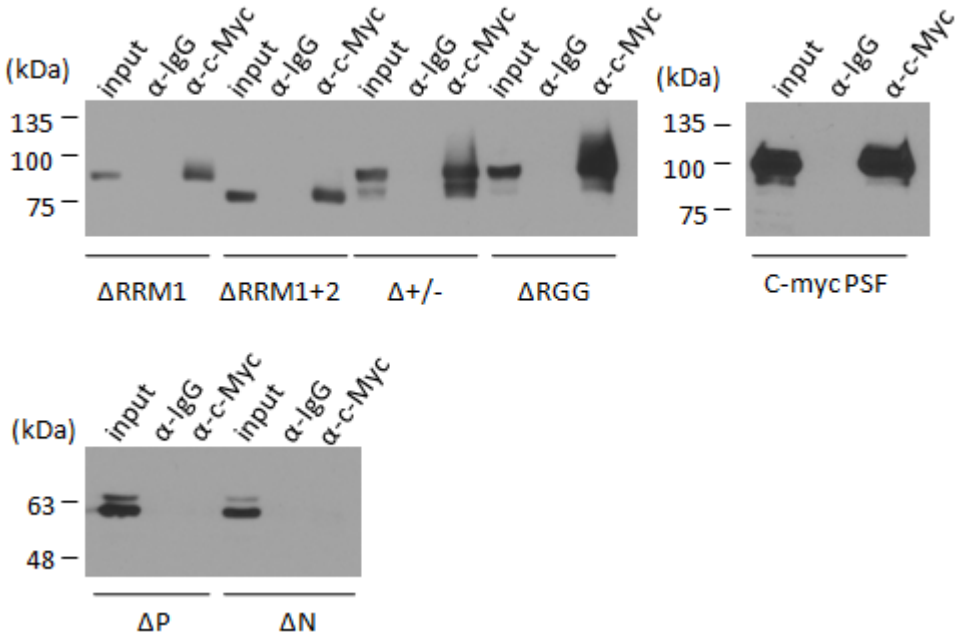

(kDa)
input
α-IgG
α-c-Myc
135
100
75
ΔRRM1
ΔRRM1+2
Δ+/-
ΔRGG
C-myc PSF
63
48
ΔP
ΔN

Figure 3.7 Accessibility of c-Myc epitope tag during immunoprecipitation using anti-c-Myc epitope antibody.

After coupling with either anti-IgG or anti-c-Myc epitope antibody, protein G magnetic beads were incubated with WCE from HEK293 cells expressing the various c-Myc epitope-tagged PSF constructs. Bound proteins were eluted and analyzed by Western blotting using the anti-c-Myc antibody. Identified proteins are shown as indicated.

ΔP and ΔN could not be detected (Figure 3.7). Since I had provided sufficient amounts of recombinant proteins for each immunoprecipitation, it is reasonable to suggest that the c-Myc epitope tag of wild type, Δ+/-, ΔRGG, ΔRRM1 or ΔRRM1+2 recombinant PSF protein is accessible by the anti-c-Myc antibody during immunoprecipitation, while c-Myc epitope tag of ΔP or ΔN mutant protein is not accessible.

The above results showed that c-Myc epitope tags of recombinant wild type PSF, Δ+/-, ΔRGG, ΔRRM1 and ΔRRM1+2 mutants are accessible to the anti-Myc epitope antibody during immunoprecipitation. Thus, I coupled anti-c-Myc epitope antibody or anti-IgG antibody (negative control) to the protein G beads, and then performed co-immunoprecipitations with ^{32}P-radiolabeled R199G and each of WCEs from the HEK293 cells expressing these c-Myc epitope-tagged recombinant proteins, in order to identify which domain of PSF is involved in the PSF-R199G interaction. In some reactions, an excess (50x) of either unlabeled R199G or P11.60 was added to act as a specific or non-specific RNA competitor, respectively. After extensive washing of beads, bound RNA-protein complexes were eluted. Eluted RNAs were resolved by Urea-PAGE and visualized by phosphorimaging. In all tests, radiolabeled R199G was identified in each lane of positive load control (lane M, Figure 3.8), but was undetectable in the negative control using anti-IgG antibody (lane 1, Figure 3.8) or missing WCE (lane 3, Figure 3.8), except for a faint band detected in lane 1 of the Δ+/- mutant probably due to insufficient washing. Radiolabled R199G was also undetectable in all lanes of specific competition tests with unlabeled R199G (lane 5, Figure 3.8). R199G RNA was co-immunoprecipitated with wild type PSF, Δ+/- or ΔRGG in the absence of competitor (Figure 3.8, lane 2), or in

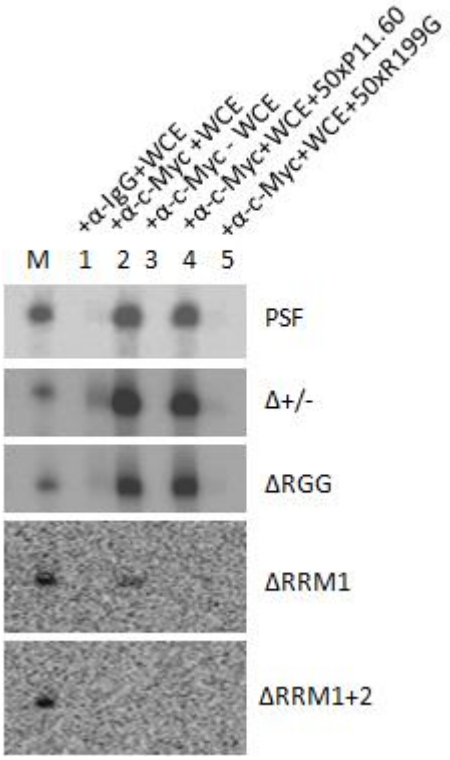

+α-IgG+WCE
+α-c-Myc+WCE
+α-c-Myc - WCE
+α-c-Myc+WCE+50xP11.60
+α-c-Myc+WCE+50xR199G
M 1 2 3 4 5
PSF
Δ+/-
ΔRGG
ΔRRM1
ΔRRM1+2

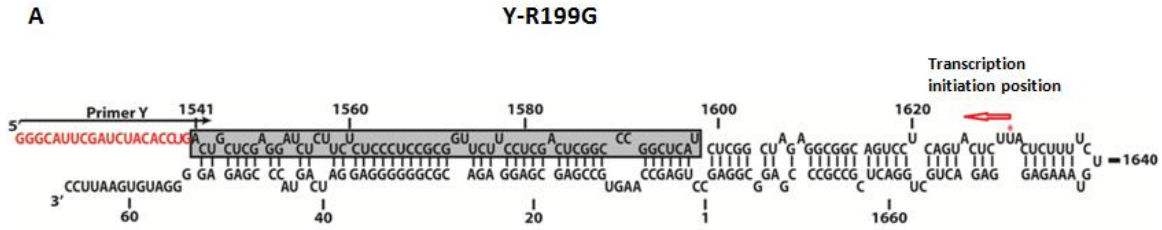
Figure 3.8 Identification of domains of PSF required for the interaction between R199G and PSF

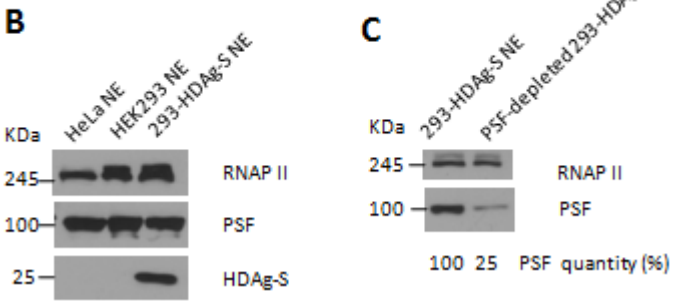
Anti-IgG (negative) or anti-c-Myc epitope antibody-coupled protein G magnetic beads were used to co-immunoprecipitate radiolabeled R199G with each of c-Myc epitope-tagged wild type and deletion mutant PSF proteins. WCE from HEK293 cells expressing each of c-Myc epitope-tagged proteins was added into each reaction mixture as protein source. Bound RNAs were resolved by Urea-PAGE and visualized by phosphorimaging**.** Lane M contains 1⁄100 radiolabeled R199G of that used for co-immunoprecipitation assay. Reaction components of the co-immunoprecipitations were shown, as indicated.

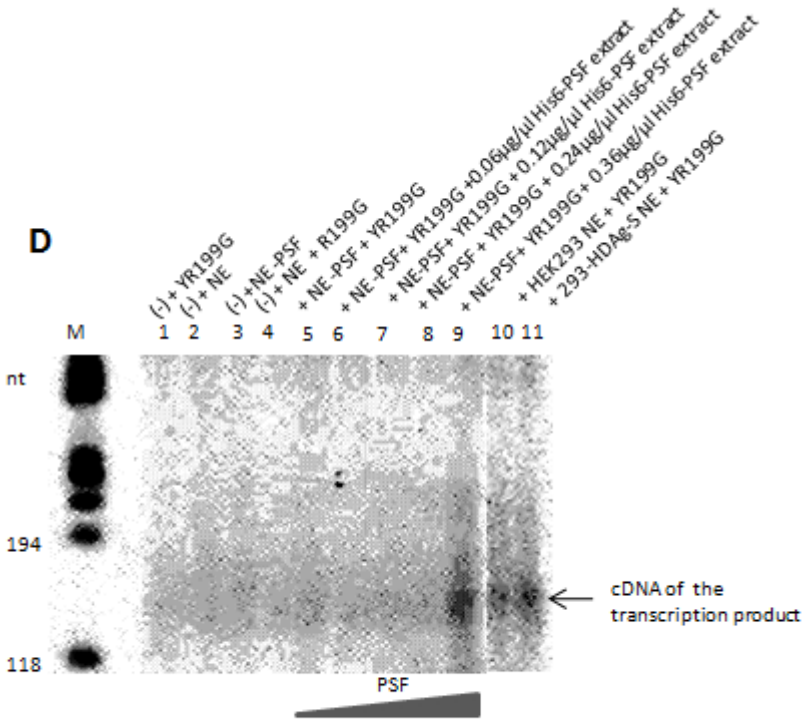
the presence of the non-specific competitor P11.60 (Figure 3.8, lane 4). Conversely, R199G was undetectable following co-immunoprecipitation with the ΔRRM1 or ΔRRM1+2 mutant, except for a faint band detected in lane 2 following co-immunoprecipitation with ΔRRM1. However, in the ΔRRM1 experiment, the band corresponding to radiolabeled R199G was undetectable when the non-specific competitor was present (Figure 3.8, lane 4); therefore, this band in lane 2 might be due to the non-specific interaction between the radiolabeled R199G and the ΔRRM1 mutant. Because the above co-immunoprecipitation results showed that ΔRRM1+2 and ΔRRM1 mutants did not specifically interact with R199G, **RRM1 or both RRM1 and RRM2 motifs of PSF might be needed for the interaction between R199G and PSF**.

3.4 PSF might facilitate *in vitro* transcription from R199G

A previous study demonstrated that R199G contains an RNA-dependent RNAPII promoter (Abrahem and Pelchat, 2008). My co-immunoprecipitation studies suggested that PSF might stimulate the interaction between RNAPII and R199G, and specific domains of PSF might be involved in the interactions of PSF with RNAPII and R199G (Sections 3.1 to 3.3). In order to further examine whether PSF is involved in HDV RNA transcription from the R199G template, a series of *in vitro* run-off transcription assays were performed, and the transcription products were then detected by using reverse transcription (RT) with a radiolabeled primer. His6-PSF protein extract was also added to some transcription reactions in endogenous PSF-depleted NE in order to assess whether recombinant PSF is able to stimulate HDV RNA transcription. However, because the 5' and 3' ends of R199G are highly complementary, simply using R199G as transcription template and using a primer binding to the 3' end of the transcript to do RT will not be able to tell the difference between the cDNA product from the transcript of R199G and that from the R199G template (Sikora, 2012). To overcome this problem, we construct a modified R199G, containing a 20-nt non-HDV sequence of 'Y' (5'-GGGCAUUCGAUCUACACCUG-3') at the 5' end of R199G (Sikora, 2012; Figure 3.9A). By using a 'Y' DNA oligonucleotide as an RT primer, the 'Y'-primed cDNA can be synthesized from the run-off transcription product, but cannot be directly synthesized from the 'Y-R199G' template (Sikora, 2012). Therefore, I used a 'Y-R199G' RNA fragment as the RNA template for the run-off transcription assay. I performed the transcription reaction in nuclear extract from tetracycline-induced 293-HDAg-S (Chang *et al*., 2005)







Lane 1-9: NE= 293-HDAg-S NE
NE-PSF = endogenous PSF-depleted 293-HDAg-S NE
(-) = negative control

Figure 3.9 Involvement of PSF in the transcription from the right terminal stem-loop region of HDV genomic RNA

(A) Sequence and predicted secondary structure of Y-R199G RNA (a modified R199G) template used for the run-off transcription assay. The gray boxed region corresponds to the 5' end of HDAg ORF. HDAg mRNA is initially transcribed from the U_{1630}, which is indicated with a red asterisk. A 20-nt non-HDV 'Y' sequence (shown in red) was added to the 5' end of R199G in order to detect the transcription product by reverse transcription with a radiolabeled primer 'Y'. (Reprinted with permission from the Figure 3.6 A of Sikora, PhD thesis, 2012). **(B)** Western blotting result of the protein extracts used for the transcription assay. Anti-RNAPII, anti-PSF/SFPQ or anti-HDAg-S antibody was used as primary antibody to detect the RNAPII, PSF or HDAg-S in the 293-HDAg-S NE and HEK 293 NE. **(C)** Western blotting result of the PSF-depleted 293-HDAg-S NE used for the transcription assay. Endogenous PSF was immunodepleted from 293-HDAg-S NE with an anti-PSF/SFPQ antibody. Western blotting with anti-RNAPII or anti-PSF/SFPQ antibody was performed to detect RNAPII or PSF in untreated and PSF-depleted 293-HDAg-S NE. **(D)** Phosporimaging results of Urea-PAGE of transcription products detected by RT. *In vitro* transcription assays were performed in 293-HDAg-S NE or HEK293 NE using Y-R199G RNA as template. Transcription products were subjected to reverse transcription (RT) using a radiolabeled 'Y' DNA oligonucleotide as the RT primer. cDNAs of RT products were resolved by urea-PAGE and visualized by autoradiography. Lane M contains the loading ladder of ΦX174 DNA/Hae III fragments radiolabeled with [γ-^{32}P] ATP. Lane 1-11 represent each test containing different transcription reaction components, as indicated.

cells containing over-expressed HDAg-S hoping it would increase the transcription efficiency, because HDAg-S is required for HDV RNA replication (Lai, 1995) and was reported to facilitate transcription elongation (Yamaguchi *et al*., 2001). HEK293 NE was also used for a parallel control reaction to evaluate the possible stimulation of HDAg-S on the *in vitro* transcription from the Y-R199G template. Following *in vitro* transcription reaction, the transcription product was then subjected to RT using a radiolabeled 'Y' DNA as the RT primer. The cDNA of RT product was resolved by urea-PAGE and visualized by phosphorimaging.

Figure 3.9 B and C shows the Western blotting results of each protein extract used for the transcription assay. RNAPII and PSF were present in HEK293 NE (Figure 3.9 B); RNAPII, PSF and HDAg-S were present in 293-HDAg-S NE (Figure 3.9 B); after immunodepletion, endogenous PSF in the 293-HDAg-S NE was significantly depleted (-75%, quantified by ImageJ software), while RNAPII was not affected (Figure 3.9 C).

In Figure 3.9 D, a band was observed between the 118-nt and 194-nt markers and the average molecular size of the band was ~130-nt (lanes 9, 10, 11). The appearance of the band was PSF-dependent, because the band was not observed in the control test using endogenous PSF-immunodepleted 293-HDAg-S NE (Figure 3.9 D, lane 5), but it was observed when 0.36 µg/µl His6-PSF protein extract was added back (Figure 3.9 D, lane 9) or when the untreated NE containing endogenous PSF was used for transcription (Figure 3.9 D, lane 10, 11). Adding back a small amount of His6-PSF might not be able to restore the *in vitro* transcription in endogenous PSF-depleted NE, because the band was not observed when 0.06, 0.12 or 0.24 µg/µl His6-PSF protein extract was put into the

endogenous PSF-depleted 293-HDAg-S NE (lane 6, 7, 8; Figure 3.9 D). The appearance of the band was unlikely due to 'crowding' because all protein extracts and transcription reaction mixtures were clear solution, and the maximum total protein concentration in each reaction was approximately 0.8 - 1.1 μg/μl, which was not very high. By comparing lane 10 (transcription in HEK293 NE) with lane 11 (transcription in 293-HDAg-S NE), we can find that HDAg-S does not affect the appearance and the intensity of the band (Figure 3.9 D). By using a series of negative control tests, the specificity of this band was examined. Firstly, the band was template-dependent because it was missing from the negative control tests, in which *in vitro* transcription was carried out in 293-HDAg-S NE (lane 2) or PSF-depleted 293-HDAg-S NE (lane 3) but without adding template (Figure 3.9 D). Secondly, the band corresponded to a product of the *in vitro* transcription reaction but not the RT product of the Y-R199G template, because it was not detectable in the negative control test which was missing NE (Figure 3.9 D, lane1). Thirdly, the band corresponded to a specific 'Y'-primed cDNA of the transcription product by using Y-R199G template, because it was not observed in the negative control transcription assay using R199G template which does not contain a 5'-terminal 'Y' sequence (Figure 3.9 D, lane 4). Therefore, the appearance of the bands in lane 9, 10 and 11 (Figure 3.9 D) suggests that *in vitro* transcription is able to start from the right terminal stem-loop region of HDV genome by using a NE including endogenous PSF or recombinant PSF. Altogether, these results suggests that **PSF facilitates the transcription from the R199G template and HDAg-S is not required for this reaction under my experimental conditions**.

Chapter 4. Discussion

The synthesis of HDV RNA takes advantage of the host transcription machinery, particularly via RNAPII (Taylor, 2006). Usually, RNAPII catalyzes transcription using a DNA template. It is still not clear how HDV manipulates cellular RNAPII machinery to recognize the HDV RNA promoters and synthesize viral RNA. Fortunately, it has been confirmed that the HDV mRNA is synthesized by RNAPII and the initiation site is located within the R199G region, which is the right terminal stem-loop domain of HDV genomic RNA (Beard *et al*., 1996; Gudima *et al*., 1999; Abrahem and Pelchat, 2008). Although the RNAPII holoenzyme can bind the R199G to form an active PIC, which is similar to that formed on a DNA promoter, the RNAPII transcription of HDV RNA in cells might require other transcription factors because the *in vitro* transcription from R199G was inefficient, and many proteins other than RNAPII and general transcription factors interact with R199G (Abrahem,A. 2008; Greco-Stewart *et al*., 2006; Sikora *et al*., 2009). In order to identify transcription factors involved in the RNAPII transcription of HDV RNA, our laboratory previously investigated host proteins associated with R199G (Sikora *et al*., 2009; Greco-Stewart *et al*., 2006). Because PSF interacts directly and specifically with R199G (Greco-Stewart *et al*., 2006), and PSF interacts with the CTD tail of RNAPII and facilitates the interaction of the CTD with RNA (Emili *et al*., 2002), PSF might play a role in the RNAPII transcription of HDV RNA. My co-immunoprecipitation experiments and transcription assay suggested that PSF stimulates the interaction between hypophosphorylated RNAPII and R199G through linking R199G and the CTD tail of RNAPII, and acts as a transcription factor for the RNAPII transcription of HDV RNA.

In my thesis, I performed co-immunoprecipitation assays using an antibody (8WG16) against the CTD tail of RNAPII and radiolabeled R199G to investigate whether PSF is able to facilitate the interaction of RNAPII with R199G. When the PSF-immunodepleted protein extract was used in the assay, interaction of RNAPII with R199G was undetectable. Complementation of the PSF-depleted extract with recombinant His6-PSF allowed the recovery of R199G-RNAPII interaction. This suggests that PSF is necessary for the efficient interaction between RNAPII and R199G. My result corresponds well with an earlier study indicating that PSF and p54nrb were able to facilitate the interaction between RNA and the recombinant GST-tagged CTD of RNAPII (Emili *et al*., 2002). However, Emili *et al.* reported that the GST-CTD itself could interact with smaller amount of RNA, while my result showed that RNAPII was not detected to interact R199G when PSF was depleted. The possible reason for this disagreement is that I used RIPA buffer containing 0.1% SDS to wash the beads following the co-immunoprecipitation. The SDS detergent in RIPA buffer efficiently removed non-specific interaction between R199G RNA and RNAPII. My result is also in accordance with a previous finding indicating that R199G simultaneously interacts with both PSF and RNAPII (Greco-Stewart, 2009; see section 1.9 of my thesis).

Moreover, my co-immunoprecipitation results suggest that PSF might facilitate RNAPII interacting with extreme tips of HDV genome and antigenome because firstly, base-pair co-variation analysis demonstrated that the secondary structures of these tips are conserved (Greco-Stewart *et al*., 2007); secondly, both PSF and RNAPII interact with these tips (Greco-Stewart *et al*., 2007; Greco-Stewart, 2009); thirdly, HDV RNA

transcription can start at these tips (Beard *et al*., 1996; Filipovska and Konarska, 2000; Abrahem and Pelchat, 2008); fourthly, mutations of these tips can prevent HDV RNA accumulation and the interaction of either PSF or RNAPII with HDV RNA (Beard *et al*., 1996; Abrahem and Pelchat, 2008; Greco-Stewart *et al*., 2007; Greco-Stewart, 2009). This hypothesis needs to be verified by further studies.

To further investigate the function of PSF in the transcription of HDV RNA, I performed an *in vitro* transcription assay using R199G RNA as a template. The formation of transcription product was not detectable when PSF-immunodepleted nuclear extract was used for transcription assay, but it was detectable when recombinant His6-PSF was added back to the above reaction system. This result suggests that PSF is necessary for the transcription of HDV RNA. Previously, our laboratory performed a PSF-knockdown experiment (Al-Ali, 2011; see section 1.9 of my thesis). The result showed that knockdown of PSF caused a significant reduction in HDV RNA accumulation in cells replicating HDV genome; therefore, PSF might be important for HDV RNA replication. However, because PSF is a multi-functional nuclear protein involved in many activities, such as pre-mRNA splicing, transcriptional regulation, RNA retention, DNA unwinding and pairing, and maintenance of pH homeostasis (Shav-Tal and Zipori, 2002), we cannot eliminate the possibility that knockdown of PSF might indirectly reduce HDV replication. My transcription assay experiment complements the PSF-knockdown experiment and suggests that PSF directly facilitates the transcription of HDV RNA.

In my transcription assay, approximately 0.36 μg/μl (36 μg / 100 μl) His6-PSF protein extract was required to add back in order to observe the band corresponding to

transcription product. This concentration is a little bit higher than 0.23 μg/μl (35 μg / 150 μl), the minimum concentration of His6-PSF protein extract required to recover the R199G-RNAPII interaction in above co-immunoprecipitation assay; however, both concentrations are in the same range. The same range of the two concentrations suggests that PSF might facilitate the transcription of HDV RNA through stimulating the interaction of RNAPII with R199G. It should be pointed out that the concentration of His6-PSF protein extract was the total protein concentration, and the Western blotting result showed that there were smaller prematurely truncated proteins other than full length His6-PSF in recombinant protein preparation. These truncated proteins might compete with His6-PSF during my experiments. Moreover, PSF is a very 'sticky' protein interacting with many nuclear proteins, such as p54nrb, PTB, Matrin 3, etc. (Shav-Tal and Zipori, 2002). Depletion of PSF could result in the loss of other proteins, which may be important for HDV RNA replication, for example, p54nrb, which interacts with R199G and RNAPII (Sikora *et al*., 2009; Emili *et al*., 2002). All the above factors are possible reasons for the high amount of His6-PSF protein extract required for complementing the loss of endogenous PSF after immunodepletion.

Additionally, PSF is an important component of paraspeckles (Fox and Lamond, 2010). This means paraspeckles have a relatively high local concentration of PSF, which favours the interaction between R199G and RNAPII. Moreover, previous immunofluorescence and laser confocal microscopy studies showed that HDAg-S co-localized with nuclear speckle protein SC35 in cells replicating HDV for at least 6 days after transfection (Bichko and Taylor, 1996). Consistently, SC35 interacts with R199G

(Abrahem and Pelchat, 2008). Furthermore, nuclear speckles contain a portion of RNAPII and coordinate the cycling of factors participating in the transcription and pre-mRNA processing (Spector and Lamond, 2011), and paraspeckles are located close to nuclear speckles (Fox and Lamond, 2010). Therefore, it is possible that HDV RNA hijacks PSF and RNAPII at some sites near paraspeckles and nuclear speckles. Immunofluorescence and RNA *in situ* hybridization studies should be performed to support this hypothesis.

I also investigated which domains of PSF are needed for the interaction with R199G or RNAPII, using a series of N-terminal c-myc-epitope tagged wild type and mutant PSF proteins. Firstly, my results showed that loss of RRM1 or both RRM1 and RRM2 affected the PSF-R199G interaction, which suggested that RRM1 or both RRM1 and RRM2 of PSF might be required for the interaction between PSF and R199G. RRM is the most abundant eukaryotic RNA binding domain, involved in all post-transcriptional processes (Maris *et al*., 2005). RRMs can be divided into canonical RRMs and non-canonical RRMs. The canonical RRMs have three conserved aromatic residues at specific positions, while the non-canonical RRMs do not contain these residues (Maris *et al*., 2005). The side chains of these conserved aromatic residues usually interact with RNA by stacking the bases of nucleic acids or contacting with sugar rings, according to the model of hnRNPA1 RRM2 (Maris *et al*., 2005). The RRM1 of PSF is a canonical RRM, containing three aromatic residues (Phe300, Phe334 and Phe336) (Passon *et al*., 2012). Therefore, these three residues may interact with HDV RNA. Furthermore, tandem RRMs often cooperate to recognize a longer nucleotide sequence and enhance the RNA binding affinity and specificity (Maris *et al*., 2005; Clery *et al*., 2008). Therefore, the

RRM2 of PSF might also cooperatively bind to R199G. Additionally, because eEF1A1, p54nrb, hnRNP-L, GAPDH and ASF/SF2 interact with R199G (Sikora *et al*., 2009), we can expect that RRMs of these proteins might interact with R199G, according to the model of my results, though more studies should be performed.

Secondly, my results showed that loss of RRM2 domain affected the PSF-RNAPII interaction, which suggested that RRM2 of PSF might be needed for interaction with RNAPII. This is consistent with the previous report indicating that RRM2 of PSF interacted with the purified recombinant CTD of RNAPII (Rosonina *et al*., 2005). Previous studies showed that this RRM2 is necessary for PSF to be located in nuclear speckles (Dye and Patton, 2001). This RRM2 can also interact with other proteins, such as four-and-a-half LIM-only protein 2 (FHL2) (Dye and Patton, 2001). Moreover, the RRM2 of PSF is a non-canonical RRM (Passon *et al*., 2012). Previous studies showed that a non-canonical RRM can interact with proteins only, or with both proteins and nucleic acids (Clery *et al*., 2008; Khoshnevis *et al*., 2010). Probably this characteristic gives the RRM2 of PSF the ability to interact with the CTD of RNAPII.

Thirdly, another important result in my thesis is that loss of N-terminal PRR domain affected the PSF-RNAPII interaction, which suggested that PRR of PSF might be needed for interaction with RNAPII. Additionally, a previous report also indicated that PSF binds to the CTD of RNAPII (Emili *et al*., 2002). Therefore, the PRR of PSF might be needed for interaction with the CTD of RNAPII. Moreover, the N-terminal PRR of PSF contains a lot of repetitive Pn, (XP)n, (XPY)n sequences, which are expected to adopt an extended and rigid polyproline II helix structure (Williamson, 1994). This kind of

structure can act as a ‘stiff stick arm’, non-stoichiometrically binding to the PRR of other proteins (Williamson, 1994). On the other hand, the CTD tail of RNAPII contains 52 repeats of the canonical sequence ‘YSPTSPS’. This highly proline-rich CTD can interact with PRRs of many other transcription factors (Williamson, 1994). This explains why the N-terminal PRR of PSF might be required for interaction with RNAPII.

My co-immunoprecipitation results also suggest that p54nrb and/or PSPC1 might interact with RNAPII and R199G in a similar way to PSF because PSF, p54nrb and PSPC1 share highly homological motifs (RRM1, RRM2, NOPS and coiled coil), and each of them has an N-terminal proline-rich region (Fox and Lamond, 2010). Further studies will be required to address these possibilities.

It should be pointed out that, in my thesis, the anti-RNAPII antibody used to study the interactions of RNAPII with R199G and PSF is the 8WG16 monoclonal antibody, which specifically recognizes the unphosphorylated Ser2 of YSPTSPS heptad of RNAPII CTD (Brookes and Pombo, 2009) corresponding to the pre-initiation and initiation forms of RNAPII. Therefore, we can deduce that R199G and PSF might interact with RNAPII CTD during pre-initiation and/or initiation of the RNAPII transcription of HDV RNA. All these results suggest that PSF is a transcription factor for the RNAPII transcription from a HDV RNA template (R199G), by facilitating RNAPII CTD binding to R199G during pre-initiation and/or initiation of transcription. However, we cannot exclude the possibility that PSF would interact with RNAPII and HDV RNA, and play a role during the elongation and termination of the transcription, because I did not do my experiments with an anti-RNAPII antibody against the hyperphosphorylated RNAPII CTD and a longer HDV

RNA template (R199G template is too short to investigate the elongation and termination).

In addition, the results of transcription assay using NE containing or not containing HDAg-S showed that HDAg-S was not able to significantly improve the R199G-dependent transcription under my experimental conditions. This is consistent with the previous study indicating that HDAg-S is not required for the initiation of the *in vitro* transcription of HDV RNA (Abrahem and Pelchat, 2008) but required for HDV RNA transcription elongation by RNAPII (Yamaguchi *et al*., 2001; Yamaguchi *et al*., 2007).

Chapter 5. Conclusion and future work

My results suggest that (1) RRM2 and N-terminal proline-rich region of PSF are required for the interaction of PSF with RNAPII CTD; (2) RRM1 or both RRM1 and RRM2 of PSF are required for the interaction of PSF with R199G; (3) PSF stimulates RNAPII-R199G interaction; (4) PSF facilitates the transcription of HDV RNA from R199G template. Moreover, PSF might simultaneously interact with R199G and the CTD of RNAPII (Greco-Stewart, 2009). Knockdown experiment showed that PSF directly or indirectly facilitates HDV accumulation *ex vivo* (Al-Ali, 2011). Therefore, by combining my results and all previous results, I can make a suggestion that PSF might act as a transcription factor for HDV RNA-dependent RNAPII transcription by linking the RNAPII CTD and R199G to facilitate the formation of a pre-initiation complex (Figure 5.1). In conclusion, my experimental results increase the knowledge of HDV RNA transcription and promoter recognition by RNAPII.

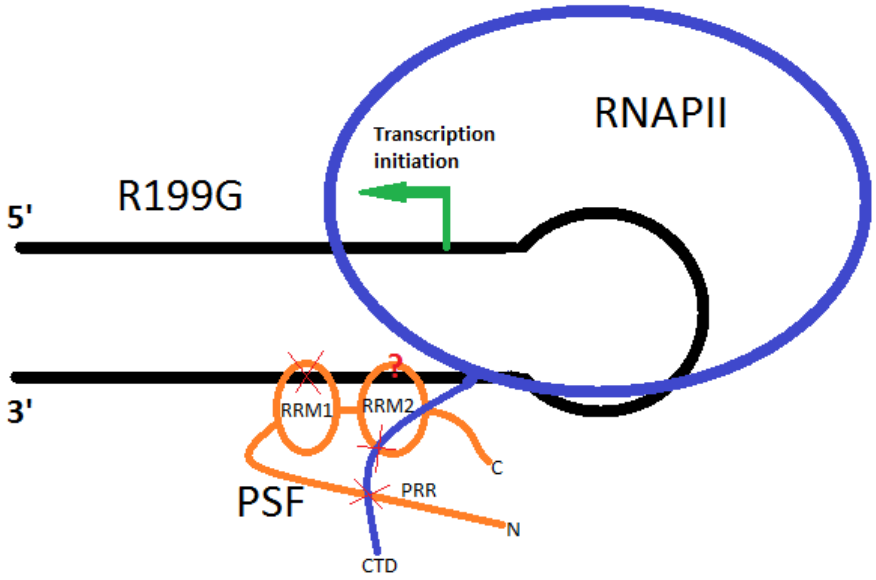

RNAPII
Transcription initiation
R199G
5'
3'
RRM1
RRM2
C
PSF
PRR
N
CTD

Figure 5.1 A proposed model of PSF acting as a transcription factor during the RNAPII transcription of HDV RNA.

PSF might act as a transcription factor during the RNAPII transcription of HDV RNA through linking the RNAPII CTD and R199G. N-terminal proline rich region and C-terminal RRM2 of PSF might interact with the CTD tail of RNAPII, while RRM1 might interact with R199G. Further experiments are required to solve whether RRM2 of PSF interacts with the HDV RNA.

My thesis only investigated the interaction and function of PSF in the transcription of HDV RNA. PSF often associates with P54nrb in regulating nuclear activities, for example, to retain highly A-to-I edited RNAs in paraspeckles, and to regulate gene expression (Shav-Tal and Zipori, 2002). Moreover, PSF and p54nrb interact with the CTD of RNAPII (Emili *et al*., 2002; Rosonina *et al*., 2005), and p54nrb was shown to interact with R199G (Sikora *et al*., 2009). Thus, we cannot exclude the possibility of p54nrb playing a role in the RNAPII transcription of HDV RNA by interacting with both R199G and RNAPII. To investigate whether p54nrb is able to facilitate R199G-RNAPII interaction, we can perform co-immunoprecipitations in a similar way as my experiments, to examine whether anti-RNAPII CTD (8WG16) antibody is able to co-immunoprecipitate R199G and RNAPII in the HEK293 or HeLa NE which is immunodepleted both endogenous p54nrb and PSF, and to examine whether the interaction between R199G and RNAPII is able to restore when adding back purified His6-tagged p54nrb protein. To investigate the function of p54nrb in the transcription of HDV RNA, we can follow the similar *in vitro* transcription assay protocol in my thesis, but using above p54nrb- and PSF-depleted NE and purified His6-tagged p54nrb, to check whether p54nrb is able to facilitate the transcription of HDV RNA from R199G template. This investigation will further enhance our knowledge of how HDV manipulate host transcription machinery for its own RNA transcription.

Appendix I:

Sequences of primers used to generate DNA templates for the synthesis HDV-derived RNA fragments and p11.60 (T7 promoter sequences are underlined; Y sequences are shown in red)

RNA species	Primer pair
R199G	5’-GGAATTCTAATACGACTCACTATAGGGACTGCTCGAGGATC TCTTCTCTCC-3’ 5’-CACTCCCCTCTCGGTGCTG-3’
Y-R199G	5’-GGAATTCTAATACGACTCACTATAGGGCATTCGATCTACAC CTGACACTGCTCGAGGATCTCTTCTCTCC-3’ 5’-CACTCCCCTCTCGGTGCTG-3’
P11.60	5’-GGAATTCTAATACGACTCACTATAGGG-3’ 5’-GGGCCGCTGAAATGCGGCGAAACTTTTGATGATATGAGTTT CGTCTCATTTCAGAGACCCTATAGTGAGTCGTATTA-3’
Y-Primer	5’-GGGCATTCGATCTACACCTG

References

Abrahem, A., & Pelchat, M. (2008). Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis delta virus RNA genome. Nucleic Acids Research, 36(16), 5201-5211.

Al-Ali, Youser (2011). MSc thesis.

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

Beard, M. R., MacNaughton, T. B., & Gowans, E. J. (1996). Identification and characterization of a hepatitis delta virus RNA transcriptional promoter. Journal of Virology, 70(8), 4986-4995.

Beck, J., & Nassal, M. (2007). Hepatitis B virus replication. World Journal of Gastroenterology : WJG, 13(1), 48-64.

Bichko, V. V., & Taylor, J. M. (1996). Redistribution of the delta antigens in cells replicating the genome of hepatitis delta virus. Journal of Virology, 70(11), 8064-8070.

Bonino, F., Heermann, K. H., Rizzetto, M., & Gerlich, W. H. (1986). Hepatitis delta virus: Protein composition of delta antigen and its hepatitis B virus-derived envelope. Journal of Virology, 58(3), 945-950.

Brookes, E., & Pombo, A. (2009). Modifications of RNA polymerase II are pivotal in regulating gene expression states. EMBO Reports, 10(11), 1213-1219.

Bushnell, D. A., Cramer, P., & Kornberg, R. D. (2002). Structural basis of transcription: Alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. Proceedings of the National Academy of Sciences of the United States of America, 99(3), 1218-1222.

Cao, D., Haussecker, D., Huang, Y., & Kay, M. A. (2009). Combined proteomic-RNAi screen for host factors involved in human hepatitis delta virus replication. RNA (New York, N.Y.), 15(11), 1971-1979.

Casey, J. L., & Gerin, J. L. (1995). Hepatitis D virus RNA editing: Specific modification of adenosine in the antigenomic RNA. Journal of Virology, 69(12), 7593-7600.

Chang, J., Gudima, S. O., Tarn, C., Nie, X., & Taylor, J. M. (2005). Development of a novel system to study hepatitis delta virus genome replication. Journal of Virology, 79(13), 8182-8188.

Chang, J., Nie, X., Chang, H. E., Han, Z., & Taylor, J. (2008). Transcription of hepatitis delta virus RNA by RNA polymerase II. Journal of Virology, 82(3), 1118-1127.

Chang, J., Nie, X., Gudima, S., & Taylor, J. (2006). Action of inhibitors on accumulation of processed hepatitis delta virus RNAs. Journal of Virology, 80(7), 3205-3214.

Chang, M. F., Chen, C. H., Lin, S. L., Chen, C. J., & Chang, S. C. (1995). Functional domains of delta antigens and viral RNA required for RNA packaging of hepatitis delta virus. Journal of Virology, 69(4), 2508-2514.

Chen, C. W., Tsay, Y. G., Wu, H. L., Lee, C. H., Chen, D. S., & Chen, P. J. (2002). The double-stranded RNA-activated kinase, PKR, can phosphorylate hepatitis D virus small delta antigen at functional serine and threonine residues. The Journal of Biological Chemistry, 277(36), 33058-33067.

Chen, P. J., Chang, F. L., Wang, C. J., Lin, C. J., Sung, S. Y., & Chen, D. S. (1992). Functional study of hepatitis delta virus large antigen in packaging and replication inhibition: Role of the amino-terminal leucine zipper. Journal of Virology, 66(5), 2853-2859.

Chou, H. C., Hsieh, T. Y., Sheu, G. T., & Lai, M. M. (1998). Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. Journal of Virology, 72(5), 3684-3690.

Clery, A., Blatter, M., & Allain, F. H. (2008). RNA recognition motifs: Boring? not quite. Current Opinion in Structural Biology, 18(3), 290-298.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., et al. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science (New York, N.Y.), 288(5466), 640-649.

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. The Biochemical Journal, 258(1), 165-169.

Darnell J., Lodish H., Baltimore D.(1986). Molecular cell biology (Scientific American Books, New York, N.Y), p 346.

Dong, B., Horowitz, D. S., Kobayashi, R., & Krainer, A. R. (1993). Purification and cDNA cloning of HeLa cell p54nrb, a nuclear protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and drosophila NONA/BJ6. Nucleic Acids Research, 21(17), 4085-4092.

Dye, B. T., & Patton, J. G. (2001). An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles. Experimental Cell Research, 263(1), 131-144.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P. W., Kobayashi, R., et al. (2002). Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA (New York, N.Y.), 8(9), 1102-1111.

Farci, P., Mandas, A., Coiana, A., Lai, M. E., Desmet, V., Van Eyken, P., et al. (1994). Treatment of chronic hepatitis D with interferon alfa-2a. The New England Journal of Medicine, 330(2), 88-94.

Ferre-D'Amare, A. R., Zhou, K., & Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. Nature, 395(6702), 567-574.

Filipovska, J., & Konarska, M. M. (2000). Specific HDV RNA-templated transcription by pol II in vitro. RNA (New York, N.Y.), 6(1), 41-54.

Fox, A. H., & Lamond, A. I. (2010). Paraspeckles. Cold Spring Harbor Perspectives in Biology, 2(7), a000687.

Fu, T. B., & Taylor, J. (1993). The RNAs of hepatitis delta virus are copied by RNA polymerase II in nuclear homogenates. Journal of Virology, 67(12), 6965-6972.

Ganem, D., & Varmus, H. E. (1987). The molecular biology of the hepatitis B viruses. Annual Review of Biochemistry, 56, 651-693.

Glenn, J. S., Watson, J. A., Havel, C. M., & White, J. M. (1992). Identification of a prenylation site in delta virus large antigen. Science (New York, N.Y.), 256(5061), 1331-1333.

Greco-Stewart, V.S. (2009). PhD thesis.

Greco-Stewart, V., & Pelchat, M. (2010). Interaction of host cellular proteins with components of the hepatitis delta virus. Viruses, 2(1), 189-212.

Greco-Stewart, V. S., Miron, P., Abrahem, A., & Pelchat, M. (2007). The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. Virology, 357(1), 68-78.

Greco-Stewart, V. S., Schissel, E., & Pelchat, M. (2009). The hepatitis delta virus RNA genome interacts with the human RNA polymerases I and III. Virology, 386(1), 12-15.

Greco-Stewart, V. S., Thibault, C. S., & Pelchat, M. (2006). Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis delta virus RNA. Virology, 356(1-2), 35-44.

Gudima, S., Chang, J., Moraleda, G., Azvolinsky, A., & Taylor, J. (2002). Parameters of human hepatitis delta virus genome replication: The quantity, quality, and intracellular distribution of viral proteins and RNA. Journal of Virology, 76(8), 3709-3719.

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. Journal of Virology, 73(8), 6533-6539.

Haussecker, D., Cao, D., Huang, Y., Parameswaran, P., Fire, A. Z., & Kay, M. A. (2008). Capped small RNAs and MOV10 in human hepatitis delta virus replication. Nature Structural & Molecular Biology, 15(7), 714-721.

Heidemann, M., Hintermair, C., Voss, K., & Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochimica Et Biophysica Acta, 1829(1), 55-62.

Hsieh, S. Y., Chao, M., Coates, L., & Taylor, J. (1990). Hepatitis delta virus genome replication: A polyadenylated mRNA for delta antigen. Journal of Virology, 64(7), 3192-3198.

Hsin, J. P., & Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & Development, 26(19), 2119-2137.

Huang, Z. S., & Wu, H. N. (1998). Identification and characterization of the RNA chaperone activity of hepatitis delta antigen peptides. The Journal of Biological Chemistry, 273(41), 26455-26461.

Hughes, S. A., Wedemeyer, H., & Harrison, P. M. (2011). Hepatitis delta virus. Lancet, 378(9785), 73-85.

Hwang, S. B., & Lai, M. M. (1993). Isoprenylation mediates direct protein-protein interactions between hepatitis large delta antigen and hepatitis B virus surface antigen. Journal of Virology, 67(12), 7659-7662.

Hwang, S. B., & Lai, M. M. (1993). A unique conformation at the carboxyl terminus of the small hepatitis delta antigen revealed by a specific monoclonal antibody. Virology, 193(2), 924-931.

Hwang, S. B., & Lai, M. M. (1994). Isoprenylation masks a conformational epitope and enhances trans-dominant inhibitory function of the large hepatitis delta antigen. Journal of Virology, 68(5), 2958-2964.

Kamitori, S., & Takusagawa, F. (1992). Crystal structure of the 2:1 complex between d(GAAGCTTC) and the anticancer drug actinomycin D. Journal of Molecular Biology, 225(2), 445-456.

Khoshnevis, S., Neumann, P., & Ficner, R. (2010). Crystal structure of the RNA recognition motif of yeast translation initiation factor eIF3b reveals differences to human eIF3b. PloS One, 5(9), 10.1371/journal.pone.0012784.

Klug, A. (2001). Structural biology. A marvellous machine for making messages. Science (New York, N.Y.), 292(5523), 1844-1846.

Kuo, M. Y., Goldberg, J., Coates, L., Mason, W., Gerin, J., & Taylor, J. (1988a). Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: Sequence, structure, and applications. Journal of Virology, 62(6), 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. Journal of Virology, 62(12), 4439-4444.

Lai, M. M. (1995). The molecular biology of hepatitis delta virus. Annual Review of Biochemistry, 64, 259-286.

Lai, M. M. (2005). RNA replication without RNA-dependent RNA polymerase: Surprises from hepatitis delta virus. Journal of Virology, 79(13), 7951-7958.

Le Gal, F., Gault, E., Ripault, M. P., Serpaggi, J., Trinchet, J. C., Gordien, E., et al. (2006). Eighth major clade for hepatitis delta virus. Emerging Infectious Diseases, 12(9), 1447-1450.

Lee, C. H., Chang, S. C., Wu, C. H., & Chang, M. F. (2001). A novel chromosome region maintenance 1-independent nuclear export signal of the large form of hepatitis delta antigen that is required for the viral assembly. The Journal of Biological Chemistry, 276(11), 8142-8148.

Lee, C. Z., Chen, P. J., Lai, M. M., & Chen, D. S. (1994). Isoprenylation of large hepatitis delta antigen is necessary but not sufficient for hepatitis delta virus assembly. Virology, 199(1), 169-175.

Lee, C. Z., Lin, J. H., Chao, M., McKnight, K., & Lai, M. M. (1993). RNA-binding activity of hepatitis delta antigen involves two arginine-rich motifs and is required for hepatitis delta virus RNA replication. Journal of Virology, 67(4), 2221-2227.

Lehmann, E., Brueckner, F., & Cramer, P. (2007). Molecular basis of RNA-dependent RNA polymerase II activity. Nature, 450(7168), 445-449.

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. Journal of Virology, 80(13), 6478-6486.

Lin, S. S., Chang, S. C., Wang, Y. H., Sun, C. Y., & Chang, M. F. (2000). Specific interaction between the hepatitis delta virus RNA and glyceraldehyde 3-phosphate dehydrogenase: An enhancement on ribozyme catalysis. Virology, 271(1), 46-57.

MacNaughton, T. B., Gowans, E. J., McNamara, S. P., & 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(1), 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. Journal of Virology, 76(8), 3920-3927.

Maris, C., Dominguez, C., & Allain, F. H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS Journal, 272(9), 2118-2131.

Meyer, P. A., Ye, P., Suh, M. H., Zhang, M., & Fu, J. (2009). Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data. The Journal of Biological Chemistry, 284(19), 12933-12939.

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. Molecular and Cellular Biology, 20(16), 6030-6039.

Moraleda, G., & Taylor, J. (2001). Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. Journal of Virology, 75(21), 10161-10169.

Nechaev, S., & Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimica Et Biophysica Acta, 1809(1), 34-45.

Nie, X., Chang, J., & Taylor, J. M. (2004). Alternative processing of hepatitis delta virus antigenomic RNA transcripts. Journal of Virology, 78(9), 4517-4524.

Otto, J. C., & Casey, P. J. (1996). The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells. The Journal of Biological Chemistry, 271(9), 4569-4572.

Pascarella, S., & Negro, F. (2011). Hepatitis D virus: An update. Liver International : Official Journal of the International Association for the Study of the Liver, 31(1), 7-21.

Passon, D. M., Lee, M., Rackham, O., Stanley, W. A., Sadowska, A., Filipovska, A., et al. (2012). Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. Proceedings of the National Academy of Sciences of the United States of America, 109(13), 4846-4850.

Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., & Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. Genes & Development, 7(3), 393-406.

Pelchat, M., & Perreault, J. P. (2004). Binding site of escherichia coli RNA polymerase to an RNA promoter. Biochemical and Biophysical Research Communications, 319(2), 636-642.

Rizzetto, M., Canese, M. G., Arico, S., Crivelli, O., Trepo, C., Bonino, F., et al. (1977). Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. Gut, 18(12), 997-1003.

Rizzetto, M., Canese, M. G., Gerin, J. L., London, W. T., Sly, D. L., & Purcell, R. H. (1980). Transmission of the hepatitis B virus-associated delta antigen to chimpanzees. The Journal of Infectious Diseases, 141(5), 590-602.

Rosonina, E., Ip, J. Y., Calarco, J. A., Bakowski, M. A., Emili, A., McCracken, S., et al. (2005). Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. Molecular and Cellular Biology, 25(15), 6734-6746.

Rozzelle, J. E.,Jr, Wang, J. G., Wagner, D. S., Erickson, B. W., & Lemon, S. M. (1995). Self-association of a synthetic peptide from the N terminus of the hepatitis delta virus protein into an immunoreactive alpha-helical multimer. Proceedings of the National Academy of Sciences of the United States of America, 92(2), 382-386.

Ryu, W. S., Netter, H. J., Bayer, M., & Taylor, J. (1993). Ribonucleoprotein complexes of hepatitis delta virus. Journal of Virology, 67(6), 3281-3287.

Samuel, D., Zignego, A. L., Reynes, M., Feray, C., Arulnaden, J. L., David, M. F., et al. (1995). Long-term clinical and virological outcome after liver transplantation for cirrhosis caused by chronic delta hepatitis. Hepatology (Baltimore, Md.), 21(2), 333-339.

Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G., & Taylor, J. (1988). Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. Journal of Virology, 62(8), 2674-2679.

Shav-Tal, Y., Cohen, M., Lapter, S., Dye, B., Patton, J. G., Vandekerckhove, J., et al. (2001). Nuclear relocalization of the pre-mRNA splicing factor PSF during apoptosis involves hyperphosphorylation, masking of antigenic epitopes, and changes in protein interactions. Molecular Biology of the Cell, 12(8), 2328-2340.

Shav-Tal, Y., Lee, B., Bar-Haim, S., Vandekerckhove, J., & Zipori, D. (2000). Enhanced proteolysis of pre-mRNA splicing factors in myeloid cells. Experimental Hematology, 28(9), 1029-1038.

Shav-Tal, Y., & Zipori, D. (2002). PSF and p54(nrb)/NonO--multi-functional nuclear proteins. FEBS Letters, 531(2), 109-114.

Sikora, Dorota (2012). PhD thesis.

Sikora, D., Greco-Stewart, V. S., Miron, P., & Pelchat, M. (2009). The hepatitis delta virus RNA genome interacts with eEF1A1, p54(nrb), hnRNP-L, GAPDH and ASF/SF2. Virology, 390(1), 71-78.

Sikora, D., Zhang, D., Bojic, T., Beeharry, Y., Tanara, A., & Pelchat, M. (2013). Identification of a binding site for ASF/SF2 on an RNA fragment derived from the hepatitis delta virus genome. PloS One, 8(1), e54832.

Sobell, H. M., & Jain, S. C. (1972). Stereochemistry of actinomycin binding to DNA. II. detailed molecular model of actinomycin-DNA complex and its implications. Journal of Molecular Biology, 68(1), 21-34.

Spector, D. L., & Lamond, A. I. (2011). Nuclear speckles. Cold Spring Harbor Perspectives in Biology, 3(2), 10.1101/cshperspect.a000646.

Srisawat, C., & Engelke, D. R. (2002). RNA affinity tags for purification of RNAs and ribonucleoprotein complexes. Methods (San Diego, Calif.), 26(2), 156-161.

Sureau, C. (2006). The role of the HBV envelope proteins in the HDV replication cycle. Current Topics in Microbiology and Immunology, 307, 113-131.

Sureau, C., Guerra, B., & Lanford, R. E. (1993). Role of the large hepatitis B virus envelope protein in infectivity of the hepatitis delta virion. Journal of Virology, 67(1), 366-372.

Sureau, C., Guerra, B., & Lee, H. (1994). The middle hepatitis B virus envelope protein is not necessary for infectivity of hepatitis delta virus. Journal of Virology, 68(6), 4063-4066.

Taylor, J. M. (2003). Replication of human hepatitis delta virus: Recent developments. Trends in Microbiology, 11(4), 185-190.

Taylor, J. M. (2006). Structure and replication of hepatitis delta virus RNA. Current Topics in Microbiology and Immunology, 307, 1-23.

Tseng, C. H., Jeng, K. S., & Lai, M. M. (2008). Transcription of subgenomic mRNA of hepatitis delta virus requires a modified hepatitis delta antigen that is distinct from antigenomic RNA synthesis. Journal of Virology, 82(19), 9409-9416.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & Development, 12(3), 343-356.

Wang, C. C., Chang, T. C., Lin, C. W., Tsui, H. L., Chu, P. B., Chen, B. S., et al. (2003). Nucleic acid binding properties of the nucleic acid chaperone domain of hepatitis delta antigen. Nucleic Acids Research, 31(22), 6481-6492.

Wang, H. W., Chen, P. J., Lee, C. Z., Wu, H. L., & Chen, D. S. (1994). Packaging of hepatitis delta virus RNA via the RNA-binding domain of hepatitis delta antigens: Different roles for the small and large delta antigens. Journal of Virology, 68(10), 6363-6371.

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(1), 119-131.

Wang, K. S., Choo, Q. L., Weiner, A. J., Ou, J. H., Najarian, R. C., Thayer, R. M., et al. (1986). Structure, sequence and expression of the hepatitis delta (delta) viral genome. Nature, 323(6088), 508-514.

Weinmann, R., & Roeder, R. G. (1974). Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. Proceedings of the National Academy of Sciences of the United States of America, 71(5), 1790-1794.

Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. The Biochemical Journal, 297 (Pt 2), 249-260.

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. Journal of Virology, 71(7), 5408-5414.

Xia, Y. P., & Lai, M. M. (1992). Oligomerization of hepatitis delta antigen is required for both the trans-activating and trans-dominant inhibitory activities of the delta antigen. Journal of Virology, 66(11), 6641-6648.

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., et al. (2001). Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science (New York, N.Y.), 293(5527), 124-127.

Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S., & Handa, H. (2007). Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. Genes to Cells : Devoted to Molecular & Cellular Mechanisms, 12(7), 863-875.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., et al. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell, 97(1), 41-51.

Zuccola, H. J., Rozzelle, J. E., Lemon, S. M., Erickson, B. W., & Hogle, J. M. (1998). Structural basis of the oligomerization of hepatitis delta antigen. Structure (London, England : 1993), 6(7), 821-830.

Contribution of collaborators:

All of the laboratory experiments were performed by Da Jiang Zhang.

Curriculum Vitae:

DA JIANG ZHANG

Education

Master of Science, Biochemistry, University of Ottawa, Canada. Thesis Supervisor: Dr. Martin Pelchat, January 2011 – 2014

Bachelor of Science, Biochemistry, Concordia University, Montreal, Canada, 2001- 2004

Chemical Engineering (undergraduate), Soochow University, Suzhou, China, 1989 - 1993

Conference Poster and Presentation

Involvement of PSF in HDV genome replication, Biochemistry Microbiology and Immunology Annual Graduate Seminars, University of Ottawa. Feb., 2012

Characterization of the transcription complex involving PSF, HDV RNA and RNAP II, RiboClub 12th Annual Meeting, Sherbrooke, QC, Sept., 2011

Involvement of PSF in HDV replication, Biochemistry Microbiology and Immunology Annual Graduate Poster Sessions, University of Ottawa. 2011

Publication

Sikora, D., Zhang, D., Bojic, T., Beeharry, Y., Tanara,, A., Pelchat, M. Identification of a Binding Site for ASF/SF2 on an RNA Fragment Derived from the Hepatitis delta Virus Genome PLOS ONE (2013) 8:e54832

Bojić, T., Beeharry,Y., Zhang. D.J., Pelchat M. Tomato RNA polymerase II interacts with the rod-like conformation of the left terminal domain of the potato spindle tuber viroid positive RNA genome J Gen Virol. (2012)

Professional Experience

Laboratory Chemist, January 2013 - March 2013, Food Directorate, Health Product and Food Branch, Health Canada

QC Analyst, June2007 - October 2010, DRAXIS Pharma Inc., Montréal, Québec, Canada

Laboratory Assistant, August 2004 - June2007, LABOVAL INC., Montréal, Québec, Canada

Assistant Engineer, 1997 - 1999, Suzhou Chung-Hwa Chemical & Pharmaceutical Industrial Co., Ltd., Jiangsu, China

Assistant Engineer, 1995 - 1997, Kunshan ROTAM Reddy Pharmaceutical Co., Ltd.

Lab Technician, 1993 - 1994, Soochow University Fine Chemical Factory in Suzhou New District, Jiangsu, China