ROLE OF RNA GENOME STRUCTURE AND PARASPECKLE PROTEINS IN HEPATITIS *DELTA* VIRUS REPLICATION

Yasnee Beeharry

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the PhD degree in Biochemistry

Thesis Supervisor: Dr Martin Pelchat

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

University of Ottawa

ABSTRACT

The Hepatitis *Delta* Virus (HDV) is an RNA pathogen that uses the host DNA-dependent RNA polymerase II (RNAP II) to replicate. Previous studies identified the right terminal domain of genomic polarity (R199G) of HDV RNA as an RNAP II promoter, but the features required for HDV RNA to be used as an RNA promoter were unknown. In order to identify the structural features of an HDV RNA promoter, I analyzed 473,139 sequences representing 2,351 new R199G variants generated by high-throughput sequencing of a viral population replicating in 293 cells. To complement this analysis, I also analyzed the same region from HDV sequences isolated from various hosts. Base pair covariation analysis indicates a strong selection for the rod-like conformation. Several selected RNA motifs were identified, including a GC-rich stem, a CUC/GAG motif and a uridine at the initiation site of transcription. In addition, a polarization of purine/pyrimidine content was identified, which might represent a motif favourable for the binding of the host Polypyrimidine tract-binding protein-associated-splicing-factor (PSF), p54 and Paraspeckle Protein 1 (PSP1). Previously, it was shown that R199G binds both RNAP II and PSF, that PSF increased the HDV levels during *in vitro* transcription and that p54 binds R199G. In the present study, I showed that PSP1 also associates with HDV RNA and I investigated whether these proteins are required for HDV replication. My results show that knockdown of PSF, p54 and PSP1 resulted in a decrease of HDV accumulation. These proteins are highly concentrated in paraspeckles, which are nuclear structures involved in storage of transcripts generated by RNAP II. I found that upon viral replication in 293 cells, PSP1 appeared as bigger foci present outside of the nucleus, while PSF and p54 foci remained in the nucleus. NEAT1 is a long non-coding RNA essential for the formation of paraspeckles. Upon HDV replication, I found an increase of the intensity and size of NEAT1 foci that correlates with an increase of NEAT1 transcripts.

Altogether, these data suggest that HDV replication results in an alteration of the paraspeckles structures, providing foundation for further investigation of the paraspeckles role in HDV cycle. Overall, the present study addresses the importance of the HDV RNA structure and of the host paraspeckle proteins for HDV replication.

ACKNOWLEDGEMENTS

I would like to thank Dr Martin Pelchat for his mentorship, his enthusiasm for science (from tiny RNAs to the solar system), for the many learning opportunities I had during my PhD and for the time he invested in teaching bio-informatics. I am sincerely thankful to my thesis advisory committee members, Dr Alain Stinzi, Dr Jean-François Couture and Dr Derrick Gibbings, as well as Dr Earl Brown for their guidance and their input. I am grateful to Dr Jocelyn Côté's laboratory and CBIA Core for their advice and help for immunofluorescence and microscopy techniques.

I am grateful to previous and current members of the laboratory for their support and good humour. Thank you Lynda Rocheleau for your kindness and for your help with the bio-informatics. To DJ and Gabrielle, thank you for always being willing to help and your sense of humour. I am indebted to you Dorota, for your invaluable support during the long days of experimental optimization, for the scientific advice, the daily help and for being such a great friend.

To the hard-working and passionate people I met in this research community, I will not enumerate you all because the list would be too long, but I address you a deep thank you for the insightful scientific discussions, help and for the good laughs. I have to mention the very exceptional people that are my friends Mohammed, Helina, Céline and Nicole, that were always by my side through my long PhD journey, with the magic of your conversations, laughs and your support, you filled my graduate student life with joy. To my dearest friends Dorota, Alexis, Helina and Julien, I am grateful to you for the proofreading of my thesis.

To my precious family and friends, here and overseas, I have always been blessed with your unconditional love, support and understanding when I was always busy with work all these years. Even if I spent most of my hours and days in another world, working in the laboratory or on my computer, thanks for the talks that kept me always connected to the world outside of the lab.

A Rubin, Reedima et mes beaux-parents, merci de toujours veiller sur moi, l'amour et les perpétuels encouragements. Merci maman Joanne pour tous les bons petits plats et de m'avoir conduit pendant cette dernière année où j'étais blessée au genou.

A toi mon amour, merci pour ta patience et ton inébranlable positivité. Un mari comme toi est une telle force motrice pour affronter les obstacles en tout genre et toujours persévérer pour de bonnes causes.

"Ce n'est qu'au prix d'une ardente patience que nous pourrons conquérir la cité splendide qui donnera la lumière, la justice et la dignité à tous les hommes. Ainsi la poésie n'aura pas chanté en vain." (Rimbaud)

Merci de m'avoir (au sens propre et figuré) soutenue et de toujours me rappeler que tout est une question d'équilibre.

"De temps en temps il est bon d'arrêter notre quête du Bonheur et d'être tout simplement heureux."(Apollinaire)

A vous mes très chers parents, je vous aime. Merci de m'avoir donné l'opportunité de poursuivre ces longues études et de m'avoir motivée sans relâche. Maman, tu es mon exemple de force morale, merci pour avoir partagé toutes les peines et joies durant mon doctorat et d'avoir toujours été de bon conseil. A mon formidable papa qui a toujours été là pour moi, j'admire ton dur labeur et tes grands accomplissements, tu es ma source d'inspiration.

Table of Contents

List of Abbreviations

ADAR	Adenosine deaminase acting on ribonucleic acid
B23	Nucleophosmin
BRE	B Recognition Element
BSA	Bovine Serum Albumine
CBF	Core Binding Factor
C/EBP	CAAT Enhancer Binding Protein
CTD	C Terminal Domain
COS-7	Cercopithecus aethiops monkey kidney cells (CV-1) with SV40 genetic material
DAPI	4',6-Diamidino-2-Phenylindole (double stranded DNA staining)
DBD	DNA binding domain
DBHS	Drosophila Behaviour Human Splicing family of proteins
DPE	Downstream Promoter Element z
ddH_2O	Double Distilled Water
DIPA	*Delta*-Interacting Protein A
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DdRp	DNA-Dependant RNA Polymerase
DRB	D-ribofuranosylbenzimidazole
DSIF	DRB Sensitivity Inducing Factor

EBV	Epstein Barr Virus
FISH	Fluorescent In Situ Hybridization
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
HBV	Hepatitis B Virus
HDAg-S	Hepatitis *Delta* Antigen-Small
HDAg-L	Hepatitis *Delta* Antigen-Large
HDRNP	HDV Ribonucleoproteins
HDV	Hepatitis *Delta* Virus
HeLa	Henrietta Lacks cells
HIV	Human Immunodeficiency Virus
IGFRE	Insulin-like Growth Response Element
INR	Initiator element
INS	cis-acting instability element
IAV	Influenza A Virus
Lnc	Long Non Coding RNA
mRNA	messenger Ribonucleic Acid
MTE	Motif Ten Element
NEAT1	Nuclear Enriched Associated Transcript1
NELF	Negative Elongation Factor
NES	Nuclear Export Signal
NLS	Nuclear Localization Signal
p54 /NONO	Non-Pou domain containing octamer binding protein

NOPS	NONA/ParaSpeckle domain
ORF	Open Reading Frame
PIC	Preinitiation Complex
PIR	Protein Interaction Domains
PABP	PolyA Binding Protein
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PKR	Double-stranded RNA-activated Protein Kinase R
PLMVd	Peach Latent Mosaic Viroid
PML	Promyelocytic Leukemia Protein Bodies
PSTVd	Potato Spindle Tube Viroid
PSF	Polypyrimidine tract-binding protein associated splicing factor
qPCR	Quantitative Polymerase Chain Reaction
R199G	Right terminal Genomic segment of 199 nucleotides
RIP	Ribonucleic Acid Immunoprecipitation
RNA	Ribonucleic Acid
RdRp	RNA dependant RNA Polymerase
RNAP	RNA Polymerase
RRM	RNA Recognition Motifs
RT	Reverse Transcription
SELEX	Systematic evolution of ligands by exponential enrichment

SFPQ	Splicing Factor Proline/Glutamine Rich
SHAPE	Selective 2'OH acylation analyzed by primer extension
SDS	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SNP	Single Nucleotide Polymorphism
TAF	Transcription Activation Function
TBP	TATA Binding Protein
TET	Tetracycline
TF	Transcription Factor
VINC	Virus Inducible Non-coding RNA
VL30	Retro-element Virus-Like 30S
Y12/ Sm	Smith antigen
293-Ag	293 cells expressing the HDAg-S
293-HDV	293 cells allowing the replication of HDV RNA

List of Figures

List of Tables

Chapter1: INTRODUCTION

1.1 Epidemiology and pathogenesis

Hepatitis is derived from the Greek words “Hepar” (liver) and “is” (inflammation; Riva et al., 2011; http://www.etymonline.com/). The Hepatitis *Delta* Virus (HDV) causes one of the most fulminant forms of hepatitis. HDV is a satellite virus dependent on Hepatitis B Virus (HBV) to infect human cells. The Rizzetto team first discovered its antigen (HDAg) in a sample of patients infected by HBV (Rizzetto et al., 1980b). The *delta* antigen was later associated with HDV and this virus was identified as a pathogen transmissible to patients infected with HBV (HBAg ; Bonino et al., 1984; Rizzetto et al., 1977).

Worldwide, it is estimated that around 20 million people suffer from HDV (Wedemeyer and Manns, 2010). Epidemics have mainly been reported in the Amazon Basin, the Mediterranean Basin and Central Africa (www.who.int). The countries most severely affected are Venezuela, Brazil, Romania, Central Republic of Africa and Kenya (www.who.int). HDV requires a helper virus since it does not encode any envelope proteins (Bonino et al., 1984; Sureau, 2006). In the case of a co-infection with HBV, less than 5% of patients develop a chronic infection of HDV. In patients already infected with HBV, a super-infection with HDV leads to a chronic HDV infection in 90% of the cases, which accelerates the development of fulminant liver disease and also increases their susceptibility to develop hepatocarcinoma by three-fold (Abbas et al., 2015; Fattovich et al., 1987; Fattovich et al., 2000; Romeo et al., 2009; Wedemeyer and Manns, 2010). To date, there is no treatment for people suffering from chronic infections. Currently, these patients are given non-specific treatment such as *alpha*-interferon injections or rely on a liver transplantation (reviewed by Wedemeyer and Manns, 2010). It is suggested that immunization against HBV provides protection against HDV (Gaeta et al., 2000; Huo et al., 2004; Rizzetto et al., 1980a). However, the 350 million people already chronically infected with HBV are at risk of an

HDV infection (www.who.int). HDV RNA levels differ according to the patient. In case of acute infection, it can reach more than 10^{12} RNA containing particles/mL of infected blood (www.who.int ; Ponzetto et al., 1991). A study of 300 patients indicated that the degree of the pathogenicity and the likeliness to develop hepatocarcinoma is dictated by the magnitude of the HDV infection (Romeo et al., 2009).

1.2 Classification, phylogeny and evolution

HDV does not belong to any viral family and was classified as its own genus, the Deltavirus (Fauquet and Fargette, 2005; Wang et al., 1986; Fields B, 2006). The HDV and viroids, a plant pathogen, are both Ribonucleic Acids (RNA) pathogens that are smaller than viruses, and both have a highly structured RNA genome; based on these common characteristics, they have been classified as the "subviral" agents (Davies et al., 1974; Diener, 1974; Rocheleau and Pelchat, 2006).

Originally, HDV was classified into three genotypes based on pathogenicity (Casey et al., 1993; Imazeki et al., 1991; Wang et al., 1987). The first genotype includes most of the sequences that have been studied; they form subtypes depending on their geographic origin such as the sequences from North America, Europe, the Middle East, the South Pacific and Asia (Wang et al., 1987). The second genotype is constituted of a single HDV isolate, from Japan (Imazeki et al., 1991). The third genotype is represented by sequences from South America (Casey et al., 1993). Subsequently, HDV was classified into eight clades, based on the genetic distances of the HDV antigen coding sequence (Deny, 2006; Le Gal et al., 2006). The diversity of HDV clades is also linked to the geographic origin (Deny, 2006). It has also been proposed that the similarities between HDV sequences of the different clades might be linked to the eight genotypes of HBV (Deny, 2006).

1.3 Virion structure and viral antigen modifications

The 34 nm enveloped infectious particle is composed of HBV envelope proteins, the HDV genome, the small antigen (HDAg-S) and the large antigen (HDAg-L; Fields, 2006). The nucleocapsid is around 19 nm in diameter and contains the HDAg and genomic RNA (Fields, 2006 ; Lai, 1995). A density fractionation assay suggested that in proportion, there is one molecule of genomic RNA for seventy molecules of HDAg (Ryu et al., 1993). However, more recent quantitative assays indicated that this molar ratio is approximately 1:200 (Dingle et al., 1998; Gudima et al., 2002). The genomic strand (300 000 copies per cell) is present in larger amounts than the antigenomic strand (30 000 copies per cell; Chen et al., 1986). Both genomic and antigenomic polarities contain a ribozyme, which is a very conserved RNA sequence capable of self-cleavage (fig.1.1; Wu et al., 1989). Northern blots showed that HDV has a circular genome since bands that migrated slower than the linear form of the genome were detected (Chen et al., 1986). Electron microscopy pictures showed its extended and unbranched genome (Wang et al., 1986). The approximately 1700-nucleotide genome is a circular negative single-stranded RNA that self-folds at 74% and forms a rod-like structure (fig.1.1; Kuo et al., 1988a; Wang et al., 1986). The genome has only one open reading frame (ORF), which encodes the two HDAg proteins (Chen et al., 1986; Lai, 2005).

The HDAg-S, of 24 kDa and 195 amino acids, is synthesized during the first steps of the viral replication cycle. Following the accumulation of viral genomes, a post-transcriptional editing of HDV antigenome by the human Adenosine Deaminase acting on Ribonucleic Acid 1 (ADAR1) allows for the deamination of an adenosine to an inosine, which results in a read-through of the stop codon of HDAg-S, and the addition of 19 amino acids (Jayan and Casey, 2005; Sato et al., 2004; Wong and Lazinski, 2002). This results in the expression of HDAg-L, of 27 kDa, which is involved in the viral encapsidation. HDAg-L

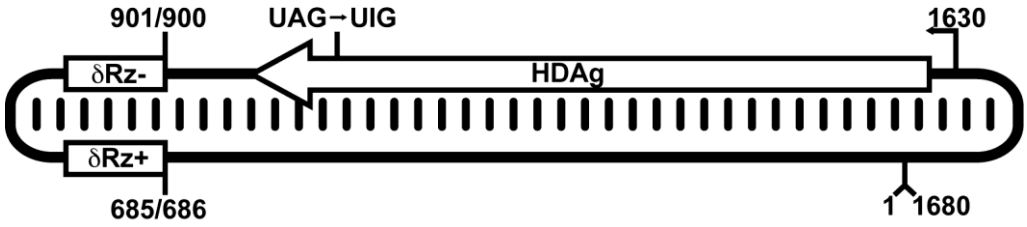

901/900
UAG→UIG
1630
δRz-
HDAg
δRz+
685/686
1
1680

Figure 1.1 Representation of HDV rod-like structure RNA genome.

The HDV genome is a rod-like RNA structure. The two ribozymes are indicated by the boxes on the left terminal side of the genome. The ORF of HDAg is indicated by the white arrow. The arrow at position 1630 indicates the start site of transcription. The UAG amber codon stop of HDAg-S is edited by ADAR in an UIG codon, which allows the synthesis of the HDAg-L (taken from Beeharry and Pelchat, 2011) .

interacts with HBV surface antigens and has a dominant negative effect on the production of the HDAg-S (Chang et al., 1995; Chao et al., 1990; Kuo et al., 1989; Lai et al., 1991; Lee et al., 1995; Ryu et al., 1992; Sureau et al., 1992).

The HDAg-S undergoes phosphorylation, acetylation, methylation, sumoylation, and the HDAg-L isoprenylation/farnesylation ; these post-translational modifications modulate both HDV transcription and replication, and also influence both HDAg nuclear localizations (Chang et al., 1988; Choi et al., 2002; Glenn et al., 1992; Li et al., 2004; Mu et al., 2001; Mu et al., 2004; Mu et al., 1999; Otto and Casey, 1996; Tseng et al., 2010). For instance, the methylation of arginine 13 retains the HDAg-S in the nucleus. Three modifications of the HDAg-S were demonstrated to be crucial for HDV replication: the methylation of arginine 13, the acetylation of lysine 72 and the phosphorylation of serine 177 (Li et al., 2004; Mu et al., 2001; Mu et al., 2004; Tseng et al., 2010). Both HDAg-S have a canonical Nuclear Localization Signal (NLS) enriched in basic amino-acids between amino-acids 66-75 (Xia et al., 1992; Alves et al., 2008). The NLS allows targeting HDV ribonucleoproteins (HDRNP) to the nucleus by using karyopherin (importin) 2 *alpha* (Chou et al., 1998). The HDAg-L has a Nuclear Export Signal (NES; Lee et al., 2001).

Since HDV uses HBV envelope proteins during infection, it has been proposed that HDV uses the same entry receptors as HBV (Littlejohn et al., 2016). It still remains controversial which liver cell receptor is used for HBV entry (Littlejohn et al., 2016). Several recent reports have suggested the NCTP sodium taurocholate cotransporting peptide as a cellular receptor candidate for HBV and HDV infection (Littlejohn et al., 2016).

1.4 Symmetrical rolling circle replication

Since HDV is only composed of an RNA genome and encodes two antigens, it has to rely heavily on its host cell components to ensure its replication and therefore it is a unique example of the extent to which an animal virus can hijack the cell (Beeharry and Pelchat, 2011). HDV replicates via a symmetrical rolling circle mechanism (fig.1.2). The circular genomic RNA strand is used as a template to produce the multimeric antigenomic strand (Kuo et al., 1988b; Lazinski and Taylor, 1995; Taylor, 2009). The antigenomic multimers are cleaved by the ribozyme self-cleaving motifs encoded on the left terminal antigenomic region from nucleotide 847 to 966 (Chao et al., 1990; Lazinski and Taylor, 1994; Taylor, 2009). The resulting linear monomers are ligated, which results in the circularization of the antigenomes (Lazinski and Taylor, 1995; Taylor, 2009). In a similar manner, the antigenomic template is used for the production of the genomic polarity and the genomic multimers are cleaved by the genomic ribozymes situated from nucleotides 659 to 772. The genomic strand is also used as a template to synthesize the messenger ribonucleic acid (mRNA) encoding the HDAg (Gudima et al., 2000; Nie et al., 2004). The HDV mRNA (600 copies per cell) is capped and polyadenylated, similarly to mRNA transcribed by the human RNA Polymerase II (RNAP II; Gudima et al., 2000; Nie et al., 2004; Taylor, 2009). Early studies showed that HDV replication was sensitive to low levels of *alpha*-amanitin (10^{-8} to 10^{-9} M), known to inhibit RNAP II mediated transcription (Goodman et al., 1984; Muhlbach and Sanger, 1979). This was surprising since it was thought that RNAP II could only use Deoxyribonucleic Acid (DNA) templates. A gain of function approach confirmed these findings, with an *alpha*-amanitin resistant allele of RNAP II, conferring a partial recovery of the levels of HDV RNA

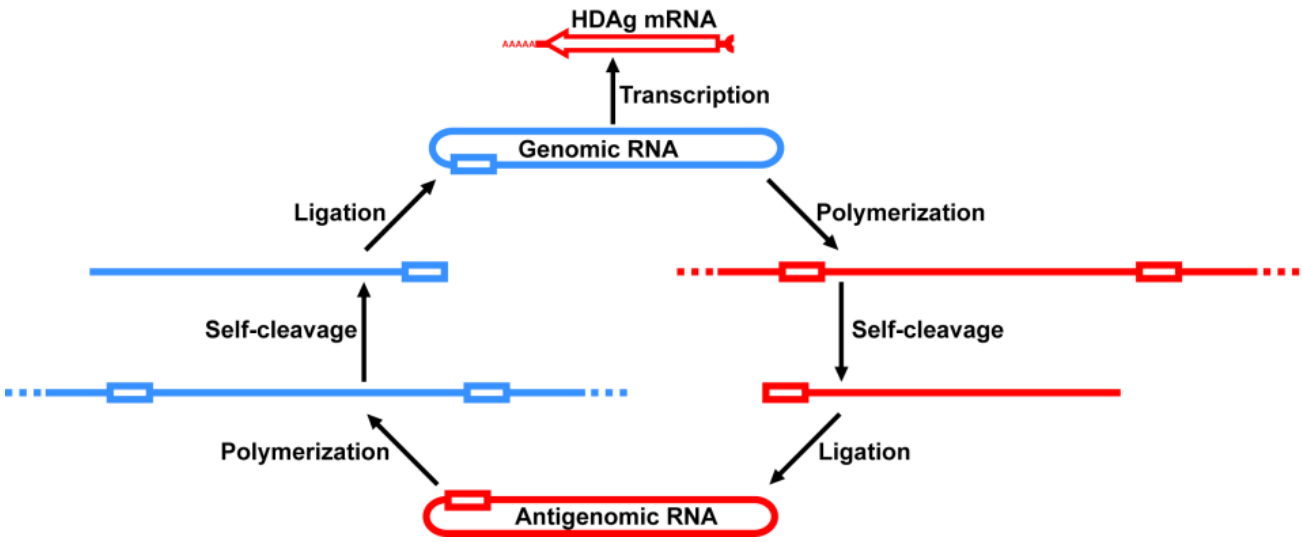

HDAg mRNA
Transcription
Genomic RNA
Ligation
Polymerization
Self-cleavage
Self-cleavage
Polymerization
Ligation
Antigenomic RNA

Figure 1.2 Symmetrical rolling replication cycle of HDV.

HDV replication cycle: the symmetrical rolling circle mechanism. The genomic strand (blue) is used as a template for the synthesis of the viral mRNA and antigenomic multimers (red). The multimers are self-cleaved by the ribozymes (rectangles). A ligation step allows the production of circular monomeric antigenomic RNA. The latter is then used as a template in the second part of the cycle to produce the genomic circular monomer RNA, in the same manner as in the first part of the cycle (taken from Beeharry and Pelchat, 2011).

products (Bushnell et al., 2002). It was later confirmed that HDV used RNAP II to replicate its RNA genome (Greco-Stewart et al., 2007).

1.5 Models used to study HDV replication

1.5.1 Study of HDV replication in cellular culture

After the viral budding and the viral entry steps, the HDV replication takes place in hepatic cells (Rizzetto et al., 1977). One major limit to study HDV replication is the cellular systems available. Primary liver cells of human or woodchuck are the ideal models to study HDV infection, but these cells are very restrictive and their availability for use in research is constraining and highly controlled (Bhogal et al., 2011; Dandri et al., 2001). HDV is a satellite virus able to replicate independently in various cell types, but relies on the HBV capsid proteins for the viral entry step (Bonino et al., 1984; Sureau, 2006). Therefore, one hypothesis that has been proposed to explain that HDV tropism is limited to the liver is that it relies on HBV proteins, which themselves interact with receptors of liver cells (Kuo et al., 1989; Netter et al., 1993; Taylor and Pelchat, 2010).

While an entire host might be necessary to produce the HDV/HBV infection and characterize the viral pathogenesis, cellular systems allow for understanding HDV RNA ability to divert its host machinery in order to replicate. In addition, HDV can be used as a tool to better characterize the host cellular pathways. For instance, the John Taylor laboratory developed a system that allows for controlling the time of induction of the HDV replication in 293 cells (Chang et al., 2005). At 16 hours after induction with tetracycline (TET), $1x10^6$ molecules of HDAg, 20 000 molecules of HDV RNA and 320 000 molecules of RNAP II per cell have been measured (Chang et al., 2005; Chang et al., 2008). This molecular ratio of

HDAg and HDV RNA is similar to that observed in the woodchuck model (Gudima et al., 2002; Mota et al., 2009).

1.5.2 Pathogenesis in cellular culture

The studies of the pathogenic effects of HDAg or HDV replication in cellular culture are very conflicting. A study on the effect of a high expression of HDAg-S in HeLa and HepG2 cells indicated that it caused cytotoxicity in cellular culture (Cole et al., 1991). In insect cells, it was shown that the expression of HDAg-S caused a cell cycle arrest at the mitotic phase, while the expression of HDAg in avian cells resulted in apoptosis (Chang et al., 2000; Hwang and Park, 1999). Nonetheless, another study in human HeLa and monkey CV1 cells showed that it was not the HDAg, but it was only HDV RNA replication that resulted in cytopathic effects on the cells (Wang et al., 2001). However, the expression of HDAg-S slightly affected the ability of the cells to reach confluency (Wang et al., 2001). The laboratory of J. Taylor compared the phenotype of 293 cells expressing only HDAg-S (herein referred as the 293-Ag), or allowing for the replication of HDV RNA (293-HDV; Chang et al., 2005). While the expression of high levels of HDAg-S over two years did not cause any major phenotypic changes in the 293 cells nor affect their ability to grow, the induction of HDV replication caused a cell cycle arrest at the G1/G0 phase in the 293-HDV cells and after 72 hours most of the cells lost their adherence properties (Chang et al., 2005). In addition, HDV heavily relies on its host for its replication and if the host gene expression is changed, this might result in pathogenesis (Taylor J, Chapter 39 in Monga and Cagle, 2010).

1.6 Evolution rate of HDV

While eukaryotes are known to have a mutation rate of 10^{-8} per site per generation, the error-rate of DNA viruses is between 10^{-7} to 10^{-6} mutation per site and RNA viruses are

known to have the highest error-rate in nature with a mutation rate of 10^{-5} to 10^{-4} mutation per site (Domingo et al., 2012). The high polymerase error-rate of RNA results in the generation of a quasi-species (Domingo et al., 2012). However, while RNA viruses encode their own viral polymerase, HDV uses RNAP II to replicate.

Like HDV, viroid replication uses its host polymerase(s). Viroids are the smallest infectious RNA agents known and infect plants. They are composed of a circular genome of approximately 350 nucleotides. They do not encode any proteins, and therefore rely heavily on their host for replication using the plant polymerase. They are classified in two families, the *Avsunviroidae* and the *Pospiviroidae*. The members of the latter have a rod like secondary structure RNA similar to HDV genome. In a similar way to HDV, the viroids are able to use DNA-dependent RNA polymerase redirected as RNA-dependent RNA polymerase (RdRp). However, viroid replication occurs in a different environment and viroids have to move through the plant (Goodman et al., 1984). Using a member of the *Avsunviroidae*, it was shown that viroids have the highest error-rate in nature, with $2.5x10^{-3}$ mutations/sequence/cycle (Gago et al., 2009). However, this result was obtained by one study consisting of an analysis of 300 sequences of a hammerhead viroid (Chrysanthemum chlorotic mottle viroid) from an *in vivo* infection in its chrysanthemum plant host. More recently, the J. P. Perreault team inoculated a Peach Latent Mosaic Viroid (PLMVd) in a peach tree and six month post-inoculation they isolated the RNA to perform deep-sequencing; they calculated that there was at least a 20% variation of the sequence of the viroid population isolated compared to the initial sequence inoculated (Glouzon et al., 2014).

The human RNAP II has been found to have an error rate of approximately 10^{-5} mutation/site during the transcription on a DNA template (Cramer, 2004; Ninio, 1991). It remains to be elucidated what the characteristics of RNAP II are while being hijacked by

HDV. Very few studies concerning the error-rate made by the RNAP II during the replication of HDV are documented (Casey et al., 1993; Chang et al., 2005; Lee et al., 1992; Netter et al., 1995). The analysis of sequences from patients chronically infected by HDV allowed to calculate a mutation of $3x10^{-2}$ to $3x10^{-3}$ substitution/nucleotide/year, which fluctuated according to the infection period (Lee et al., 1992). Subsequent analysis of a region of approximately 350 nucleotides of the antigen region modified by ADAR led to the conclusion that there are 70% nucleotide changes that are caused by ADAR editing. It was found that 90% of the mutations were transition U>C and A>G (Casey et al., 1993; Netter et al., 1995). Compared to the other regions of the genome, the fragment located at the right terminal region was found to be among the most conserved regions (Casey et al., 1993; Netter et al., 1995). The 293-HDV system was used in another study where the mutation rate over the whole HDV genome was found to be 2.1% changes/nucleotide/year (Chang et al., 2005). Chang and colleagues reported that the rod-like structure was conserved at 100%, but they indicated that they only analyzed at best nine clones per nucleotide. Indeed, all these studies on HDV error-rate during replication were based on cloning experiments to isolate several sequences of the HDV population, restricting the studies to a maximum of 100 sequences. In addition, since it was shown that HDAg-S modifies the error-rate of RNAP II when transcribing from a DNA template, HDAg-S could be contributing to the observed error-rate (Yamaguchi et al., 2007).

1.7 Involvement of the RNA Polymerase II in replication and transcription

1.7.1 RNAPII pre-initiation complex and DNA promoter features in the human cell

One of the dogmas of Molecular Biology is that RNAP II uses a DNA template to synthesize the mRNA (Sims et al., 2004; Smale and Kadonaga, 2003). RNAP II is a complex

enzyme divided in twelve main subunits and is a major protein in the human host cell since it is responsible for the transcription of mRNAs (Sims et al., 2004; Smale and Kadonaga, 2003). The DNA/RNA complex forms in the catalytic domain of RNAP II, which is composed of two large subunits, RBP1 and RBP2, of respectively 220 kDa and 140 kDa (Sims et al., 2004; Smale and Kadonaga, 2003). The C Terminal Domain (CTD) of the RBP1 is constituted of 52 repeats of the YSPTSPS consensus peptide (Corden et al., 1985). The phosphorylation of Serine 2 and the Serine 5 of the YSPTSPS heptapeptide allows for the elongation step of the transcription to proceed (Cadena and Dahmus, 1987; Komarnitsky et al., 2000; Meinhart et al., 2005).

The first step of the transcription preinitiation is the binding of the TATA Binding Protein (TBP), a subunit of TFIID (Butler and Kadonaga, 2002). The sequence elements present within the DNA promoters may vary but there are DNA promoter elements that are most commonly found on an RNAP II template. The upstream core promoter elements are situated before the transcription initiation site. The TBP is recruited to a TATA box (TATA[A/T]AA[G/A]) at position -31 to -25 upstream of the site of initiation and allows for the recruitment of the pre-initiation complex (PIC) of RNAP II. When the TATA box is not present, there is often a recognition element (BRE) or a downstream core promoter element. The general transcription factors TFIIA and TFIIB are then recruited to the TFIID complex (Orphanides et al., 1996). TFIID may bind the motif ten element (MTE), a downstream core promoter element situated at position +18 to +32: C[G/A]A[A/G]C[G/C][C/A/G]AACG[G/C][A/G]G[A/T][C/T][G/A/C] (Butler and Kadonaga, 2002). When the TATA box is absent, the Transcription Associated Function 9 (TAF 9) and TAF6 from TFIID bind the downstream Promoter Element (DPE) [A/G]G[A/T][C/T][G/A/C] located at positions +28 to +32 and promote the recruitment of

TFIID. TFIIB binds to the BRE ([G/C][G/C][G/A]CGCC) at position -37 to -32. TAF1 and TAF2 bind to the Initiator element (INR; [C/T] [C/T] [A+1]N[T/A][C/T] [C/T]). RNAP II and TFIIF are then recruited to the DNA promoter, followed by TFIIE and TFIIH and TFIIJ (Butler and Kadonaga, 2002; Orphanides et al., 1996; Zawel and Reinberg, 1993). Additional TFs are then recruited and the mediator/SRB complex may interact with the proximal promoter core elements, that are sometimes present in the region approximately around positions -200 to +250: the CAAT binding protein (CBF), NFI and the CAAT Enhancer Binding Protein (C/EBP) bind the CAAT box [CCAAT] and Sp1 binds the GC box [GGGCGG] (Bjorklund and Gustafsson, 2005; Butler and Kadonaga, 2002; Orphanides et al., 1996; Ossipow et al., 1999).

1.7.2 RNAP II used by HDV

Previously in the laboratory, RNA Immunoprecipitation (RIP) with an antibody directed against the CTD of RNAP II provided confirmation that RNAP II binds to the terminal regions of both polarities of the HDV genome in HeLa cells (Greco-Stewart et al., 2007). An RNA affinity chromatography and *in vitro* transcription assay developed in the laboratory showed that the same RNAP II PIC used in the human cell is used by HDV on the right terminal genomic segment of 199 nucleotides (R199G; Abrahem and Pelchat, 2008). In addition, the TBP is also recruited on the HDV promoter and facilitates the recruitment of the PIC, analogously to what occurs on a DNA template (Abrahem and Pelchat, 2008). The re-initiation complex, comprising TFIIA, TFIID and TFIIE, might stay on the RNA promoter after the initiation complex, similarly to what occurs on a DNA template (Yudkovsky et al., 2000; Abrahem and Pelchat, 2008).

During the progression of transcription, the polymerase stalls due to the encumbrance of the template stem (Yamaguchi et al., 2007). The binding of the HDAg-S to the Rpb1 and Rbp2 subunits of RNAP II allows a conformational change of the polymerase and forces the progression of the transcription (Chao et al., 1990; Yamaguchi et al., 2001; Yamaguchi et al., 2002; Yamaguchi et al., 2007). Based on this, it was proposed that HDAg-S modulates the transcription rate and increases the error-rate during HDV replication (Yamaguchi et al., 2007). The cocrystallization of an HDV RNA antigenomic scaffold with RNAP II showed that the RNA template was found in the same site usually occupied by the DNA template (Lehmann et al., 2007). **While this crystal structure gave insight into the mechanism of RNA-templated RNAP II transcription, the structure of the RNA template was not solved and was the subject of our investigation.**

1.7.3 RNAP II RNA promoter features

The RNAP II and HDAg-S are crucial for HDV replication and the HDV RNA promoter characteristics are likely important for the use of RNAP II by HDV. The two HDV terminal regions bind RNAP II and were identified as two potential RNA promoters (Filipovska and Konarska, 2000; Greco-Stewart, 2009; Greco-Stewart et al., 2007). R199G was shown to be a promoter region allowing initiation of transcription by RNAP II *in vitro* (Beard et al., 1996). It has been proposed by the J. Taylor team that R199G could be a promoter for the synthesis of the antigenome (Gudima et al., 1999; Gudima et al., 2000). Transcription initiation occurred at the uracil at position 1630 on genomic HDV RNA (fig. 1.1), using the same preinitiation complex used by RNAP II for a DNA template (Abrahem and Pelchat, 2008; Gudima et al., 1999). The TATA-binding protein interacts with R199G (Abrahem and Pelchat, 2008). The deletion or addition of nucleotides in the 3 nucleotide

bulge containing the initiation site resulted in an abrupt decrease of HDV genome accumulation (fig. 1.3; Greco-Stewart et al., 2006; Gudima et al., 1999). Mutation of the nucleotides composing the right terminal loop did not have a noticeable impact on the HDV RNA accumulation while insertion, deletion or segment inversion of the nucleotides composing the double-stranded region had a drastic effect (Wu et al., 1997). In the case of large insertions perturbing the rod-like structure, reversion to wild-type was observed, showing the importance of HDV rod-like structure. I have summarized the state of knowledge of the requirements of HDV R199G promoter to use RNAP II in fig. 1.3 and Appendix I. Recent experiments in the laboratory allowed refining the HDV right terminal region that interacts with RNAP II and mutations that abrogated the interaction of R199G with RNAP II were identified (Greco-Stewart, 2009; Greco-Stewart et al., 2006). However, these experiments were performed *in vitro* and the essential secondary and tertiary structural features of the HDV RNA promoter allowing for the recognition by RNAP II remain to be solved.

1.8 Localization of the HDAg-S and the HDV genomes in the host cell and interaction with the nuclear structures

1.8.1 Localization of HDAg

There have been many conflicting reports concerning the exact sites of the HDV genome and HDV antigen cellular localization (Bell et al., 2000; Bichko and Taylor, 1996; Cullen et al., 1995; Cunha et al., 1998; Dourakis et al., 1991; Gowans et al., 1988; Han et al., 2009; Huang et al., 2001; Lai et al., 1991; Lazinski and Taylor, 1993; Lee et al., 1998; Macnaughton et al., 1990; Negro et al., 1991; Tavanez et al., 2002; Taylor et al., 1987; Wu et al., 1992). The nuclear sites of HDV transcription and replication remain unclear. HDAg-S is involved in the replication of the genome and it has been proposed that it allows for the

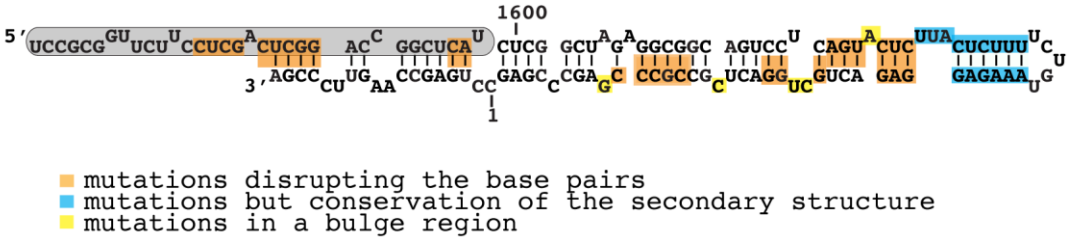

1600
1
mutations disrupting the base pairs
mutations but conservation of the secondary structure
mutations in a bulge region

Figure 1.3 Summary of positions important for HDV replication identified by previous *in vitro* mutagenesis assays.

For the region at the tip of the stem (region between positions 1632 to 1648), mutations that changes the sequence (blue), but allows for preserving the secondary structure resulted in a decrease of interaction of HDV with RNAP II (Greco-Stewart, 2009) or lower levels of HDV RNA accumulated in cells (Wu et al., 1997). An alteration of the bulge "UUA" (blue; positions 1629 to 1631) located at the tip of the sequence also resulted in a decrease of HDV RNA accumulated in cells (Gudima et al., 1999). Sequence deletions or mutations of the bulges (yellow) and/or the base pairs (orange) in the regions located between positions 1585-1628 and 1649-18 results in a decrease of HDV RNA accumulated in cells (Beard et al., 1996; Liao et al., 2012; Wang et al., 1997).

progression through RNAP II pausing (Yamaguchi et al., 2001). Previously, different laboratories performed immunostaining of HDAg in Huh7 and HepG2 cells, which were transfected with DNA plasmids allowing the production of HDV genomes. They concluded that the HDAg-S is directed to the nucleolus and interacts with B23, a nucleolus marker, during early HDV replication (Cunha et al., 1998; Huang et al., 2001; Lee et al., 1998). Without the expression of HDV genome, both the HDAg-S and HDAg-L accumulate in the nucleolus (Cunha et al., 1998). However, later on, using a 293 cellular system (described in section 1.5.3 and in fig. 2.2 of Chapter 2), John Taylor's team found that the change between the nucleolar or nucleoplasmic localization of HDAg depends on the time elapsed since the seeding of the cells. For the 293-Ag cells, in an early stage, the HDAg-S is in the nucleolus but later on, after 16 to 24h following the seeding of the cells, the HDAg-S localizes to the nucleoplasm (Han et al., 2009). Another study also showed that when solely the HDAg is expressed, it is present in the nucleus (Tavanez et al., 2002).

J. Taylor's team identified a concentration of HDAg located just beside the nuclear speckle structures (Bichko and Taylor, 1996). They showed that in the earlier steps of the viral cycle, during replication, the HDAg-S was localized in the speckles which is not in accordance with the data of Cunha et al and the Gerd G Maul team, where no colocalization of HDAg with the speckles marker SC-35 is observed in HEp-2 and Huh7-D12 cells (Bell et al., 2000; Bichko and Taylor, 1996; Cunha et al., 1998). However, in the later steps of HDV viral cycle, Taylor et al. also observed that the HDAg-S were localized next to the speckles, which is in accordance with the results of Bell P et al. and Cunha et al. (Bell et al., 2000; Bichko and Taylor, 1996; Cunha et al., 1998). In Huh7-D12 cellular system, which has cDNA allowing the expression of HDV RNA integrated in their chromosome, the HDV RNA

genomes colocalized with HDAg in the nucleoplasm and also as brighter foci in proximity to the nuclear speckles (Cunha et al., 1998). This is in accordance with observations that upon HDV replication in 293 cells, the HDAg-S localizes to the nucleoplasm (Han et al., 2009).

1.8.2 Study of the link between the HDV foci and HDV transcription

These HDAg foci do not seem to be sites of HDV transcription (Cunha et al., 1998; Han et al., 2009). When the cells were treated with *alpha*-amanitin, at a dose known to inhibit HDV replication, these nuclear HDAg remained unchanged and HDV RNA foci were not disrupted (Cunha et al., 1998; Han et al., 2009). In addition, in Huh7 cells containing HDV RNA mutants that are defective for replication, the HDAg-S localization was also nucleoplasmic (Han et al., 2009). Therefore, the authors suggested that this nucleoplasmic localization of HDAg might be linked to the HDV RNA accumulation, but they could not conclude that HDAg localization would remain the same in a cell where HDV is actively replicating (Han et al., 2009). Moreover, by using a labeled nucleotide, bromo-dUTP, *de novo* synthesis of the genome was monitored and no correlation between the HDV foci and the transcription sites was identified (Cunha et al., 1998). Furthermore, these foci composed of HDAg and HDV RNA did not contain significant players of the transcriptional machinery such as RNAP II, PolyA Binding Protein (PABP), PolyA polymerase, nor Cleavage and Polyadenylation Specificity Factor (Cunha et al., 1998). In addition, the HDV genome is edited by the human enzyme ADAR and this allows the expression of HDAg-L (Cunha et al., 1998). However, in this study, there was no significant difference in the nucleoplasmic distribution of ADAR in cells containing HDV genomes (Cunha et al., 1998). Based on these findings, the authors concluded that the nuclear foci containing HDAg and genomes were likely not the sites of replication or transcription, but probably just viral RNA-protein

complexes (Cunha et al., 1998). These findings refute the hypothesis of the J Taylor team that the HDAg foci localized next to the speckles might be the sites of HDV transcription (Bichko and Taylor, 1996). Since both HDV genomic RNA and RNAP II are diffused in the nucleoplasm during HDV replication, it seems that the transcription is not concentrated in foci (Cunha et al., 1998).

1.8.3 Localization of genomic and antigenomic HDV RNAs

Overall, using different systems, two teams showed that HDV RNA is present in the nucleus during HDV replication (Bell et al., 2000; Cunha et al., 1998; Tavanez et al., 2002). The team of M. Carmo-Fonseca developed an assay to perform an *in situ* hybridization against HDV genome (Cunha et al., 1998). They used Huh7 cells stably transfected by the plasmid pSLV(D3) that allows expressing a trimer of HDV RNA (Cunha et al., 1998). They performed an *in situ* hybridization assay using the plasmid pSLV(D3) as a probe in these cells expressing HDV genomes (Cunha et al., 1998). However, one can argue that a limitation of this study is that it does not imitate the HDV rolling circle replication mechanism, which relies only on RNA intermediates. More recently, the team of M. Carmo-Fonseca performed an *in situ* hybridization with the same method previously employed, using pSLV(D3) as a probe (Tavanez et al., 2002). One limit of the protocol is that it still includes a heat-denaturing step of the sample to denature the HDV double-stranded region, which might degrade or denature the RNA and change its properties and localization. They used the plasmid pDL542 to only express the HDV RNA, without expressing the HDAg and found that HDV RNA expressed alone had a cytoplasmic localization (Tavanez et al., 2002). However, the plasmid pDL542 generates unstable RNA circles defective for replication, which differs from the normal HDV RNA symmetrical rolling circle mechanism. They also

observed a similar cytoplasm localization upon transfection of the plasmid pDL481, that generates the antigenome (Tavanez et al., 2002). However, when HeLa cells were transfected with pDL542 and pDL481, they found the RNA to be localized in the nucleus (Tavanez et al., 2002). The authors attributed this difference in HDV RNA localization to the different cellular system used (Tavanez et al., 2002). When both the HDAg and the HDV RNA were expressed, the HDV RNA localized in the nucleus (Tavanez et al., 2002). HDAg seems to redirect the HDV RNA to the nucleus, likely due to the fact that HDAg has an RNA Recognition Motif (RRM) as well as an NLS (Tavanez et al., 2002). Tavanez et al. also looked for the localization of the HDRNPs by inhibiting *de novo* protein synthesis of the cell and then using immunofluorescence and Fluorescent In Situ Hybridization (FISH) to detect the simultaneous colocalization of HDAg and HDV RNA when both the HDAg and the HDV RNA were expressed (Tavanez et al., 2002). The HDRNPs were found in nucleoplasmic foci and in the cytoplasm (Tavanez et al., 2002). Further work will allow for determining why HDV RNA expressed alone is directed in the cytoplasm and how the HDRNPs are exported in the cytoplasm. Additional studies will be required to ascertain the exact localization of HDV genome and antigen during the replication cycle and to localize where the HDV replication and transcription takes place. Their work suggested that the RNA genome might be responsible for the export of HDRNPs to the cytoplasm (Tavanez et al., 2002). Since the HDRNPs do not seem to accumulate in the cytoplasm following passive diffusion, the authors speculated that one possibility might be that the binding of the HDV RNA to the HDAgs triggers the exposure of a NES of the HDAgs; another possibility is that the HDV RNA itself has a cytoplasmic localization signal.

In their study, P. Bell et al. observed foci located close to SC-35 composed of the HDAg-L and HDV antigenomes. The HDAg-S and HDAg-L are both detected by a rabbit

antibody against HDAg-S while a specific antibody immunized from guinea pigs against the 19 terminal amino-acids allow for detection of the HDAg-L. Smaller foci of HDAg-L located close to the Promyelocytic Leukemia Protein Bodies (PML) were found to hijack PML bodies proteins and these aggregates were later merged to the bigger HDAg-L foci. They found that the genomic RNA together with the HDAg-S was diffuse in the nucleus but excluded from the nucleolus compartment. For this RNA *in situ* hybridization experiment, the Gerd G Maul team used probes that allowed differentiating genomic and antigenomic strands, and also a probe recognizing both polarities to better characterize the localization of the genome during the HDV replication cycle (Bell et al., 2000). They transfected HEp-2 cells with the plasmid pDL538 that allows for the expression of genomic HDV RNAs or with the plasmid pDL456 that allows for the expression of antigenomic HDV RNAs (Bell et al., 2000). The transfection of the pDL538 allows the synthesis of HDV RNA of antigenomic polarity from the genomic RNA, similarly to what occurs during HDV replication cycle.

In summary, from the data obtained from the FISH assays performed by the two teams, it seems that the genomic RNA is present in the nucleoplasm, while the antigenomic RNA is present in the PML bodies and in proximity of the speckles (Bell et al., 2000; Cunha et al., 1998; Tavanez et al., 2002). Much work remains to be done to further clarify the exact localization of HDV genomes during replication cycle.

1.9. Host proteins linked to HDV cycle

1.9.1 Host proteins affected by HDV viral cycle

Large-scale analysis of the proteins affected by HDV have previously been performed by the C. Cunha team (Mota et al., 2009; Mota et al., 2008). They found that the expression of 32 proteins was changed in Huh7 transfected with cDNAs expressing HDV RNAs and

both HDAgs separately (Mota et al., 2008). In a Huh7 cell line stably transfected with a cDNA allowing the expression of HDV and both HDAgs from a same cDNA (Huh-D12), using the 2-DE technique, they identified 15 proteins that are down-regulated and a reverse-transcription followed by a Quantitative Polymerase Chain Reaction (qPCR) confirmed the down-regulation of the histone H1-binding protein and the triosephosphate isomerase (Mota et al., 2009). Among the eight proteins that were found to be up-regulated, Western Blot analysis confirmed the up-regulation of La protein and lamin A/C (Mota et al., 2009). However, the cell growth did not seem to be affected (Mota et al., 2009; Mota et al., 2008). Simultaneously, the M. A. Kay team constructed a 293 cell line that stably produces a FLAG-HDAg-S and identified more than 100 proteins associated with the FLAG-HDAg-S (Cao et al., 2009). Among 65 proteins tested, they identified that the knockdown of 15 resulted in a down-regulation of HDV levels and the knockdown of 3 resulted in an up-regulation of HDV levels (Cao et al., 2009). Later on, a micro-array analysis of the 293-Ag system identified a significant increase (at least two fold compared to the 293 cells) of 72 genes and a decrease of 22 genes (Chapter 39, Taylor J in Monga and Cagle, 2010). For the 293-HDV cells, a significant increase (at least two fold) of 662 genes and a decrease of 295 genes was measured. Independently, the team of C. Cunha analyzed the variation in cellular gene expression during HDV replication using the same 293-HDV system, and identified that the expression of 89 genes was changed (Mendes et al., 2013). Altogether, these data show that HDV replication considerably changes the transcriptional expression of its host cell. However, for the assays using a system that does not produce HDV RNA, the data should be taken with caution, since even if these proteins may interact with HDAg, this interaction might be modified or abrogated in presence of the HDV genome.

1.9.2 Host proteins involved in HDV cycle

Other studies further investigated the role of the host proteins in HDV replication cycle (reviewed by Beeharry and Pelchat, 2011). The binding of the Histone H1e to HDAg-S was found to be important for the HDV RNA accumulation (Lee and Sheu, 2008). Surprisingly, it was discovered that humans have a homolog protein of HDAg, now named the *Delta*-Interacting Protein A (DIPA; Brazas and Ganem, 1996). The overexpression of DIPA results in an inhibition of HDV replication and DIPA interacts with HDAg through the coiled-coil domain. DIPA also represses cellular genes (Bezy et al., 2005; Du et al., 2006). A later screening identified several proteins that interact with HDAg and they further characterized the protein MOV10, an RNA helicase that localizes to the P bodies (Haussecker et al., 2008). The knockdown of MOV10 resulted in a decrease of HDV accumulation by three times (Haussecker et al., 2008). HDAg interacts with nucleolin and nucleophosmin B23 (Huang et al., 2001; Lee et al., 1998). Another factor, Yin Yan 1 interacts with nucleolin, nucleophosmin, RNAP II and HDAg (Huang et al., 2008; Inouye and Seto, 1994; Lee and Lee, 1994; Usheva and Shenk, 1996). The Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) interacts with the U/C rich region at positions 379 to 414 of HDV antigenome and enhances the ribozymes activity (Lin et al., 2000).

ADAR1 is responsible for the editing of the HDV stop codon and HDAg-S represses the action of ADAR 1 acting on HDV RNA (Wong and Lazinski, 2002). The complex composed of the transcription factor (TF) Yin Yan 1 and the acetyltransferases CBP and p300 promotes HDV replication (Huang et al., 2008). This allowed identifying that the acetylation of HDAg-S stimulated HDV replication (Huang et al., 2008). The post-translational state of HDAg might dictate its interactants (Huang et al., 2008). Double-stranded RNA-activated Protein Kinase R (PKR) binds all the different forms of HDV RNA

and also the HDAg-S (Chen et al., 2002; Robertson et al., 1996). Nonetheless, the assays performed in a reticulocyte translation system suggested that the HDV RNA binding to PKR does not result in a shut-off of the translation; rather, it was proposed that PKR might have a role in HDAg-S phosphorylation (Manche et al., 1992; Robertson et al., 1996). More recently, it was demonstrated that an HDV ribozyme with a semi-globular structure, formed by a dimer of short sequences, would activate PKR (Heinicke and Bevilacqua, 2012). As the entire HDV genome does not activate PKR, unlike this dimer of HDV ribozyme, this highlights the importance of the RNA structure of the entire HDV genome in the HDV cycle, in order to impede the triggering of cellular pathways downstream of PKR activation and therefore to escape the innate immune system (Heinicke and Bevilacqua, 2012).

While the involvement of RNAP II in HDV replication was identified by the observation that HDV RNA accumulation was sensitive to 10^{-8} to 10^{-9} M of *alpha*-amanitin, it was also observed that the antigenomic RNAs are resistant to this dose (Macnaughton et al., 2002; Modahl et al., 2000). This observation, together with assays showing that the HDAg-S binds to the TF SL1, suggests an involvement of RNAP I for the synthesis of the antigenomes (Li Y J, 2006). The Negative Elongation Factor (NELF) and the DRB Sensitivity Inducing Factor (DSIF) are involved in the RNAP II pausing during the cellular transcription (Yamaguchi et al., 2002). HDAg has a high sequence similarity with the subunit A of NELF. The binding of HDAg to RNAP II dislodges NELF and DSIF (Yamaguchi et al., 2001; Yamaguchi et al., 2002; Yamaguchi et al., 2007). It is likely that this displacement of the transcriptional repressors would force the progression of transcription, and it has been proposed that HDAg-S could modify the transcription rate of RNAP II (Yamaguchi et al., 2007). As for the majority of the interactants of HDAg and HDV RNA, even if the interaction was determined, their function in the HDV cycle has not been elucidated yet.

1.10 PSF and p54 involvement in HDV replication and transcription

1.10.1 Host proteins interacting with HDV right terminal promoter

In addition to RNAP II, *in vitro* immunoprecipitation assays indicate that R199G also interacts with RNAP I and RNAP III (Abrahem and Pelchat, 2008; Greco-Stewart, 2009). Future *in vitro* transcription assays will establish whether R199G can be used as templates by these polymerases. To identify what host proteins interact with R199G and might be involved in transcription, the laboratory performed an Ultra Violet-crosslinking of R199G with nuclear extract followed by mass spectrometry analysis of the bound proteins (Sikora et al., 2009). It was previously shown that R199G interacts and uses RNAP II, and as it was expected, RNAP II core units were present in the complex (Greco-Stewart et al., 2006). The other proteins identified were the Eukaryotic translation factor 1A1, the nuclear mitotic apparatus protein 1, the F-box and leucine-rich repeat protein 17, the ANKS6 protein, the RhoGEF kinase isoform 3, the Polypyrimidine tract-binding protein associated splicing factor (PSF) and Non Pou domain containing octamer binding protein (p54). Furthermore, the laboratory confirmed *in vitro* that R199G had the ability to co-immunoprecipitate with GAPDH, Alternative Splicing Factor, Heterogeneous Nuclear Ribonucleoprotein-L, p54 and PSF (Greco-Stewart, 2009; Sikora, 2012).

1.10.2 PSF and p54 interact with the right terminal region of HDV RNA

My laboratory has further investigated whether PSF was involved in HDV replication. Immunoprecipitations and electrophoretic mobility shift assays showed that the two terminal loop regions of HDV genome were essential and sufficient for the interaction of the HDV RNA with RNAP II and PSF (Greco-Stewart et al., 2006). Previous findings in the laboratory indicate that the same region of HDV, R199G, co-immunoprecipitates with PSF

(Greco-Stewart et al., 2006; Zhang DJ, 2012). Since mutations that disrupt R199G secondary structure abrogate the interaction with PSF and RNAP II, it was suggested that the same secondary structure may be required for the interaction of HDV RNA with RNAP II and PSF (Greco-Stewart et al., 2006). In addition, *in vitro* transcription assays demonstrated that RNAP II is required for HDV replication, and that the depletion of PSF from the nuclear extract resulted in a decrease in transcription product accumulation (Abrahem and Pelchat, 2008; Sikora, 2012). The transcription efficiency was rescued by adding back the PSF protein to the nuclear extract immunodepleted of PSF (Zhang, 2012).

R199G binds directly to the TBP subunit of TFIID (Abrahem, 2008). A fragment of R199G from nucleotides 1618 to 1655, named R38G was found to be the minimal fragment required to interact with both RNAP II and PSF (Greco-Stewart et al., 2006; Greco-Stewart, 2009; Greco-Stewart et al., 2007). Given that R38G binds RNAP II and PSF, and that PSF was demonstrated to be able to interact with the CTD of RNAP II, one hypothesis that was previously investigated was whether the initial interaction of PSF with R199G would facilitate the recruitment of RNAP II to the HDV promoter, and thus act as a transcriptional activator (Greco-Stewart, 2009). PSF interacts with the HDV RNA via its RNA binding domain (RRM), but also has a DNA binding domain (DBD; Zhang, 2013; Yarosh et al., 2015). PSF could then act to enhance the recruitment of RNAP II from cellular DNA promoters to the HDV RNA promoter (Greco-Stewart, 2009). To test this hypothesis, competition experiments with RNA segments that have a high affinity for PSF RRM were performed, and then the RNAP II was immunoprecipitated and the amount of RNAP II bound to R199G was quantified (Greco-Stewart, 2009). The Systematic evolution of ligands by exponential enrichment (SELEX) technique allowed for obtaining two RNAs with high affinity for PSF (Peng et al., 2002). These two RNAs and another trimeric consensus repeat

RNA with high affinity for PSF were used in excess to block PSF RNA binding sites (Greco-Stewart, 2009; Peng et al., 2002). These assays showed that in the presence of RNA competitors for PSF binding sites, there was a decrease of R199G interaction with RNAP II. These findings support the hypothesis that PSF might act as a link between R199G and RNAP II. Altogether, these results support the hypothesis that PSF is required for HDV to replicate its RNA genome. **However, in addition to these *in vitro* experiments, further investigation into whether PSF is necessary for HDV genome accumulation in cells is required.**

PSF often dimerizes with its molecular partner p54. Previously in the laboratory, p54 has also been shown to interact with R199G by co-immunoprecipitation and by fluorescence binding assays (Sikora et al., 2009; unpublished data). This led me to speculate that p54 might also be required for HDV to replicate its genome by using RNAP II. In my work, I investigated whether PSF and p54 were required for HDV replication in cells.

1.11 Structure and roles of the DBHS proteins PSF, p54 and PSP1

1.11.1 PSF, p54 and PSP1 protein structure

The amino acids sequence homology between p54, PSF and PSP1 is around 50% (Fox et al., 2005). Similarly to PSF and p54, PSP1 also belongs to the Drosophilia Behaviour Human Splicing family of proteins (DBHS; Yarosh et al., 2015). In the drosophila, there is only one member of this protein family, suggesting a functional diversification among these three paralogous mammalian proteins (Fox et al., 2005; Fox AH, 2010; Passon et al., 2012). One structural characteristic of these proteins is that they have a pair of RRMs, followed by a 52-amino acid sequence motif designed as the NONA/Paraspeckle domain (NOPS), and a

coiled coil region of 100 amino acids separating the NOPS region from the N-terminal domain (Major et al., 2015; Passon et al., 2011; Yarosh et al., 2015).

PSF is a protein that was first identified in association with the polypyrimidine tract-binding proteins (Patton et al., 1993; Yarosh et al., 2015). PSF is composed of approximately 707 amino acids and its molecular weight is 76 kDa (Passon et al., 2011). PSF is also referred to as SFPQ, referring to its role as a Splicing Factor, and its Proline and Glutamine Rich domains. It also has a glycine rich domain at its N terminal end, two NLS at its C terminal end, two RRMs and a DBD. PSF is subject to methylation, phosphorylation and sumoylation of its N-terminal and RRM1 domains (Buxade et al., 2008; Yarosh et al., 2015; Zhong et al., 2006).

p54 was initially identified in co-purification with PSF by the Krainer team (Zhang et al., 1993). p54 has 471 amino acids (Dong et al., 1993). During mitosis, both PSF and p54 are phosphorylated (Proteau et al., 2005). When considering a region of 320 amino acids around the RRM region, PSF and p54 share 71% sequence identity (Dong et al., 1993). The first RRM domain is conserved and has typical RRM consensus: four aromatic amino-acids at specific positions, that are likely the sites of interaction with RNAs (Passon et al., 2012). Conversely, in the second RRM, three of these amino acids are replaced by a Threonine, Lysine or Isoleucine, and it remains to be understood whether this RRM does not bind RNA or whether it is the site of a non-conventional type of interaction with RNA (Passon et al., 2012).

PSP1 was first identified in a proteomic assay to identify the nucleolus components (Andersen et al., 2002). PSP1 is a protein of 523 amino-acids and of 59 kDa, present in lesser cellular quantities than p54 and PSF (Fox et al., 2005; Yarosh et al., 2015). There are two alternatively spliced isoforms of PSP1: PSP1 *alpha* and PSP1 *beta*. The PSP1 *beta* protein,

that lacks the amino acids 388 to 523, is also present in nuclear paraspeckle structures. PSP1 *alpha* is present in larger quantities in HeLa cells, while PSP1 *beta* presence varies in mouse and human tissues (Fox et al., 2002; Myojin et al., 2004). It is expressed in different tissues, including the liver and the kidney. Both isoforms of PSP1 are phosphorylated by players of the DNA damage response pathway (Matsuoka et al., 2007). PSP1 contains two theoretical NLS sequences (Myojin et al., 2004; Passon et al., 2012). The first NLS is in the coiled-coil domain and the second NLS is in the C-terminal region (Major et al., 2015; Myojin et al., 2004).

DBHS proteins often form homodimers or heterodimers by interacting through the region constituted by the second RRM domain along with the coiled coil structure (Fox et al., 2005; Passon et al., 2012). PSP1 and p54 interact via the NOPS. The C. H. Bond team found that the truncated PSP1 (amino acid 61 to 320) and p54 (amino acid 53 to 312) were soluble and the team achieved the crystallization of the core region of the p54/PSP1 heterodimer (PDB ID code 3SDE; Passon et al., 2011; Passon et al., 2012). This model of PSP1/p54 heterodimer core region shows an antiparallel coiled-coil domain present on the right region, and that forces the two RRM2 domains of p54 and PSP1 to rearrange on each side of a 20 Å channel (Passon et al., 2011; Passon et al., 2012). This rearrangement is quite unusual since no RNA is present in the tunnel formed by the two RRM2. The authors compare this structure to proteins that act together on structured DNA and hypothesize that this rearrangement of the two RRM2 together might be important for the heterodimer to act on structured RNAs (Passon et al., 2012). The protein-protein interaction interface is hydrophobic and involves the four protein domains expanding on a large region of the protein (Passon et al., 2012). The authors suggest that these proteins are constrained to form transient dimers (Passon et al., 2012). In addition, PSP1 truncated at position 358 could target

paraspeckles but this was not the case for the crystallized form of the truncated PSP1. However, these PSP1 truncations did not abrogate PSP1 dimerization with p54. Therefore, the authors propose that the additional anti-parallel coiled-coil region located at the C terminal end of the heterodimer is a prerequisite for the establishment of the paraspeckle structure (Passon et al., 2012).

1.11.2 Functions and localization of PSF, p54 and PSP1 in the cell

The DBHS proteins are dynamic and the phosphorylation state and the nuclear localization of PSF and p54 is linked to their diverse roles (Shav-Tal and Zipori, 2002; Yarosh et al., 2015; Appendix II). They are involved in the control of DNA replication, DNA double-strand break repair pathway and radioresistance pathways (Akhmedov et al., 1995; Akhmedov and Lopez, 2000; Bladen et al., 2005; Ha et al., 2011; Rajesh et al., 2010; Gao et al., 2014; Gao et al., 2016; Li et al., 2014; Mahaney et al., 2009; Matsuoka et al., 2007; Straub et al., 1998; Udayakumar et al., 2003; Wu et al., 2010). In addition to the regulation of replication, interference with the expression of the DBHS proteins leads to a destabilization of the cellular cycle (Gao et al., 2016; Mahaney et al., 2009; Rajesh et al., 2010; Riva et al., 2011). Moreover, PSF and p54 were first identified as splicing factors (Patton et al., 1993; Zhang et al., 1993). In addition to their role in the processing of spliced RNAs, the DBHS proteins are highly involved in other RNA metabolism pathways, (Gozani et al., 1994; Kameoka et al., 2004; Peng et al., 2002). They also have a role in transcriptional regulation (Emili et al., 2002; Song et al., 2005). In particular, they have been identified in many hormonal transcriptional pathways (Dong et al., 2005; Dong et al., 2007; Ishitani et al., 2003; Mathur et al., 2001; Shav-Tal and Zipori, 2002; Kuwahara et al., 2006; Urban et al., 2000; Urban et al., 2002). The DBHS proteins participate in the cellular immunity, for instance

through the regulation of cytokine transcripts (Buxade et al., 2008). They participate in the regulation of RNA stability and of RNA transport (Kaneko et al., 2007; Major et al., 2015; Chen and Carmichael, 2009; Mao et al., 2011a; Nakagawa and Hirose, 2012; Sasaki and Hirose, 2009; Zhang and Carmichael, 2001; Hundley and Bass, 2010; Lu and Sewer, 2015). PSP1 role in the cell is not well understood and still under investigation (Fox et al., 2002; Gao et al., 2016; Myojin et al., 2004).

1.12 Link between HDV replication and the paraspeckle structure

1.12.1 The paraspeckles structure

PSF, p54 and PSP1 are found ubiquitously in the nucleoplasm and in the nucleolar caps but are present in higher concentrations in the paraspeckles (Bond and Fox, 2009; Fox et al., 2005; Shav-Tal and Zipori, 2002). Consistently, the fluorescent immunostaining pattern of these proteins is a staining throughout the nucleoplasm and brighter foci which are the paraspeckles (Fox et al., 2005). PSP1 was the first paraspeckle marker used (Fox et al., 2005). Paraspeckles are the most recently discovered nuclear structures, having been discovered around 15 years ago (Fox et al., 2002). They are situated at the periphery of the speckle structures, other granular nuclear structures which are sites enriched with splicing factors (Spector and Lamond, 2011). The paraspeckles are sausage-like in shape and of very small size, approximately 0.25 to 1 micron (Cardinale et al., 2007; Fox et al., 2005; Fox et al., 2002; Hutchinson et al., 2007; Souquere et al., 2010). Usually, the number of paraspeckle foci per cell ranges between 2 to 20 (Fox et al., 2002). They were identified in different histological sections, primary cells and transformed cells (Fox et al., 2002). Electron microscopy analysis has described paraspeckles as interchromatine granules-associated

zones, which are fibrillar structures separate from the speckles granular structures (Cardinale et al., 2007). The paraspeckle structures are sensitive to RNase treatment (Fox et al., 2005).

1.12.2 The long non-coding RNA NEAT1

The RRMs of the DHBS proteins are important for paraspeckle targeting (Fox et al., 2005). The paraspeckle structure forms around an aggregation of several DBHS proteins and the long non-coding RNA Nuclear Enriched Associated Transcript 1 (NEAT1), of approximately 4kb (Sasaki and Hirose, 2009). NEAT1, also known as Virus Inducible Non-coding RNA (VINC), was identified by the Rangarajan team in the brains of mice infected by the Japanese Encephalitis Virus and rabies. NEAT1 is also known as MEN *epsilon/beta* since it is transcribed from the multiple endocrine neoplasia locus, on the human chromosome 11 (Hutchinson et al., 2007). NEAT1 has 2 isoforms, NEAT1_1 of 3.7 kb and and NEAT1_2 of 23 kb (Sasaki and Hirose, 2009; Hutchinson et al., 2007). NEAT1 is an RNA of considerable size and could possibly fold on itself to adopt secondary structures such as hairpin loops. Another possibility would be that it forms a netting scaffold for the paraspeckle components (Clemson et al., 2009; Clemson et al., 2011).

1.12.3 Maintenance of paraspeckle

Although PSP1 was the first paraspeckle marker used, it is not required for the formation of paraspeckles. A constant flow of PSP1 proteins is exchanged between the paraspeckles and the nucleolus structures (Fox et al., 2002). Upon chemical inhibition of the RNAP II activity, PSP1 and PSF markers relocalize to perinucleolar caps (Fox et al., 2002; Fox et al., 2005; Sasaki and Hirose, 2009). The NOP domain of PSP1 is necessary for its relocalization to the perinucleolar caps (Fox et al., 2005). The presence of the RRM domains is dispensable for the perinucleolar localization, but is correlated to an increased amount of

p54 binding to PSP1 (Fox et al., 2005). Upon removal of the RNAP II transcriptional inhibitor D-ribofuranosylbenzimidazole (DRB), PSP1 relocalizes to the paraspeckles (Fox et al., 2005; Yamaguchi et al., 1998; Sasaki and Hirose, 2009). Both the second RRM that retains its RNA binding function and its NOPS domain are required for its localization to the paraspeckles (Fox et al., 2005).

More recently, the direct interaction of importin *alpha* 2 with PSP1 was linked to a change in PSP1 foci. Overexpression or knockdown of the nuclear receptor showed that its levels were directly positively correlated to the size and number of PSP1 foci per cell, and also to the number of cells that contained PSP1 foci (Major et al., 2015). However, the changes in importin *alpha* 2 did not affect the number or size of PSF foci (Major et al., 2015). The authors proposed that importin *alpha* 2 facilitates the transport of PSP1 from the cytoplasm to the nucleus. Furthermore, the authors suggested that the scaffold of the paraspeckles does not depend on PSP1 flow, but might be controlled by other RNA transcripts or other paraspeckle players (Major et al., 2015).

Fluorescence immunostaining showed that most DBHS protein markers, such as PSP1, are diffused at different locations through the nucleoplasm, but NEAT1 is present only in paraspeckles (Clemson et al., 2009). Most of the bright fluorescent foci of PSP1 also colocalize with NEAT1 foci of faint or bright staining; this suggests the important role of NEAT1 for paraspeckles. Furthermore, in cells knocked down for NEAT1_2, upon removal of DRB, the paraspeckle proteins fail to relocalize as paraspeckle foci (Sasaki and Hirose, 2009). A high dose of the transcriptional inhibitor actinomycin D results in a redistribution of NEAT1_2 diffusely through the nucleoplasm and a disappearance of the paraspeckle foci and also a decrease of NEAT1_2 transcripts to 60% levels compared to control cells (Sasaki and Hirose, 2009). This decrease of NEAT1_2 transcripts could be due to the inhibition of RNAP

II activity or to the physical separation of NEAT1_2 and the paraspeckle proteins (Sasaki and Hirose, 2009). Altogether, these results show that the NEAT1_2 isoform is required for the maintenance of the paraspeckles (Sasaki and Hirose, 2009).

The silencing of p54 and PSF decreases the levels of NEAT1_2 transcripts by 20%, but this has no effect on the NEAT1_1 transcript (Sasaki and Hirose, 2009). In addition, the silencing of p54 and PSF leads to their diffuse relocalization in the nucleoplasm and a disassembly of the paraspeckles. Surprisingly, the silencing of PSP1 does not change the levels of NEAT1 transcripts nor disrupt the paraspeckle foci (Sasaki and Hirose, 2009). Altogether, these data suggest that PSF and p54, but not PSP1, sustain the accumulation of NEAT1_2 transcripts (Sasaki and Hirose, 2009). In addition, both *in vitro* and *in vivo* RIP show that NEAT1_2 interacts with PSF and p54. Several teams have proposed a model where the primary components of the paraspeckle assembly would be NEAT1_2, PSF and p54, followed by secondary players such as PSP1 and NEAT1_1 (Naganuma and Hirose, 2013; Nakagawa and Hirose, 2012). Later on, two other paraspeckle components with RRMs were identified: PSP2 and CFI(m)68 (Auboeuf et al., 2004; Dettwiler et al., 2004; Iwasaki et al., 2001). The tumour suppressor Appendixin 10 is also present in the paraspeckles (Quiskamp et al., 2014). Recent proteomic screen led to the discovery of 32 new paraspeckle proteins (Nakagawa and Hirose, 2012; Yamazaki and Hirose, 2015). The paraspeckle proteins can be separated in at least two categories: the components essential for the integrity of the paraspeckle, such as PSF and p54, and the secondary component proteins (Nakagawa and Hirose, 2012; Yamazaki and Hirose, 2015).

1.12.4 Initiation and formation of paraspeckles

NEAT1 is transcribed by RNAP II and present in large quantities in the cell (Clemson et al., 2009). Upon a knockdown of NEAT1, the paraspeckles are disassembled, but the PSF, p54 and PSP1 levels do not change. Instead they appear to be diffuse in the nucleoplasm (Sasaky Y T F., 2009). This is consistent with the initial observations that the paraspeckle structures are abrogated upon an RNAse treatment (Fox et al., 2005). Furthermore, upon silencing NEAT1, the number of paraspeckles decreases drastically from an average of 13 paraspeckle foci to 1 or 2 remaining foci (Clemson et al., 2009). In addition, during the cell cycle, NEAT1 transcription and assembly close to the NEAT1 locus is the first event towards the building of paraspeckles and is followed by the recruitment of DBHS proteins to the foci, an hour after the initiation of RNA transcription (Clemson et al., 2009). In accordance, an overexpression of NEAT1 leads to an increase of paraspeckle numbers but an overexpression of PSP1 does not affect the number of paraspeckles. NEAT1 interacts with PSP1 through its RRM domain (Clemson et al., 2009). In addition, the binding of p54 to the protein interaction domain (PIR) of PSP1 has a positive cooperative binding effect on the ability of PSP1 to bind NEAT1 (Clemson et al., 2009). Altogether, these findings show that NEAT1 is essential for the formation of paraspeckles.

1.12.5 The biological and cellular role of paraspeckles

Paraspeckles store a large amount of A-to-I edited RNAs (Fox and Lamond, 2010). A proposed role for paraspeckles is that they participate in the regulation of gene expression (Prasanth et al., 2005). The hyper-edited CTN RNA interacts with p54 (Prasanth et al., 2005; Zhang and Carmichael, 2001). CTN is a long non-coding RNA transcribed from the same gene that also allows for the transcription of the membrane protein MCAT2 mRNA (Prasanth et al., 2005). Upon cellular stress, the levels of CTN non-coding RNAs are reduced since

CTN RNAs are cleaved and converted into MCAT2 mRNAs (Prasanth et al., 2005). Therefore, the authors proposed that this could be an example of how the paraspeckles would provide a mode to store the spliced and processed CTN RNA, and that upon a stress stimulus, that would allow for a more rapid cytoplasmic export of the mRNA and translation (Prasanth et al., 2005). The sequestration of the CTN RNA could be a way of regulation of the transcription. Several other roles have been proposed for paraspeckles: a crossroad of regulation of RNA-processing or a place of storage of proteins, such as transcriptional repressors (Prasanth et al., 2005). This latter role is in accordance with the findings that PSF might regulate oncogene transcription (Song et al., 2005). The paraspeckles may also be involved in the response to cellular stress (Clemson et al., 2009; Hirose et al., 2014). It was reported that NEAT1 formed paraspeckle foci even upon stress shock and interferon stimulation (Clemson et al., 2009). It was also found that a cellular stress induces NEAT1 transcription (Hirose et al., 2014).

Other teams have studied the role of paraspeckles upon viral infection, such as Human Immunodeficiency Virus (HIV), Epstein Barr Virus (EBV) and Influenza A Virus (IAV), which help better understand the cellular role of paraspeckles (Bond and Fox, 2009; Imamura et al., 2014; Mao et al., 2011b; Zolotukhin et al., 2003). Interestingly, upon IAV infection, the paraspeckle foci were bigger. Indeed, the transcriptional repressor PSF was recruited to paraspeckles, which resulted in an increase of the transcription of NEAT1 and IL8 (Imamura et al., 2014). Since PSF seems to be involved in HDV replication, I investigated what would be the consequences on the cellular paraspeckles.

1.13 Rationale, hypothesis and objectives

HDV is a satellite virus that heavily hijacks human host proteins to replicate. HDV uses the human RNAP II, in an unconventional manner, to transcribe and replicate its RNA genome. Previous *in vitro* experiments in the laboratory indicated that the R199G structure could be important for its interaction with both RNAP II and PSF. However, the RNA structure that binds RNAP II and PSF has not been fully elucidated. PSF shares similar structural features with p54 and PSP1. R199G also binds p54. No research has been undertaken to determine whether PSP1 binds HDV RNA, and whether p54 and/or PSP1 are necessary for HDV replication cycle. Interestingly, PSF, p54 and PSP1 are highly concentrated in the nuclear paraspeckle domains. It is known that other viruses or stress stimuli can result in a disruption, a bulge or a reduction of the size and number of paraspeckle foci. It remains unexplored whether there is a link between HDV replication and the paraspeckles.

Through my thesis, I investigated these questions based on the hypothesis:
A specific HDV RNA structure is recognized and used by RNAP II, PSF, p54 and PSP1 and the HDV replication disrupts the paraspeckle structures.

The objectives of my thesis were:

I. To characterize the conserved features located at the right terminal domain of HDV RNA

II. To establish whether PSF, p54 and PSP1 are required for HDV replication

III. To investigate whether HDV replication disrupts the paraspeckles

This knowledge will allow further elucidation of the importance of the HDV RNA promoter structure to hijack its host proteins PSF, p54 and PSP1 and to replicate its RNA genome via RNAP II. Therefore, this study will contribute to the knowledge on HDV replication. In addition, the study of the link between HDV replication and the paraspeckles will add to the wealth of knowledge on the cellular subversion of the host cell by HDV, and improve the understanding of the role of paraspeckles in the cell.

Chapter 2: MATERIALS AND METHODS

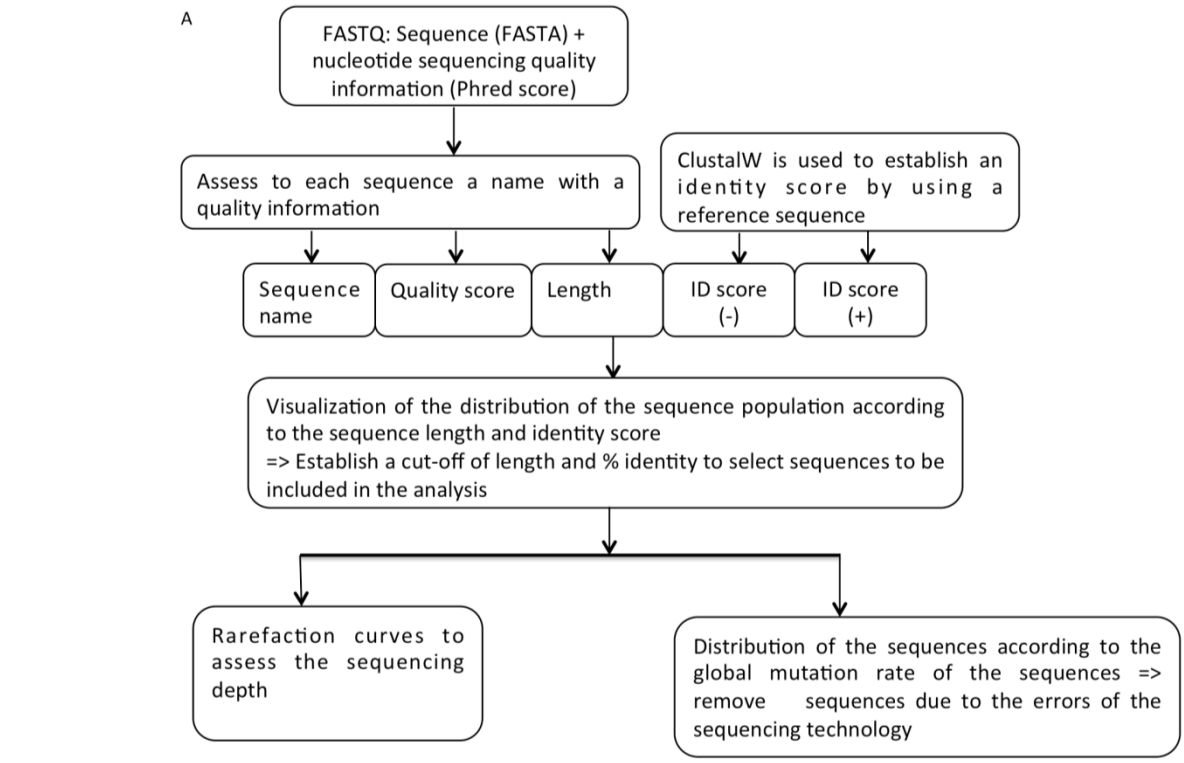
2.1 Bio-informatics methods

2.1.1 Variant sequences analysis

Sequences of the HDV variants were extracted from the Subviral RNA Database (Rocheleau and Pelchat, 2006). The alignment was performed sequentially. First, seven sequences (GI X77627, U19598, AJ584844, AB118833, AJ584848, AJ584844, and AJ584847) representative of clades 1 and 2 were aligned. Then, based on this alignment and the model of the R199G secondary structure obtained by *in vitro* nuclease mapping, a covariance model was established with the Infernal 1.0.2 software (Beard et al., 1996; Nawrocki et al., 2009). With this model generated, the rest of the sequences were aligned with Infernal to generate a second covariance model. However, the sequences of clade 3 could not be aligned at this step since they were too divergent from the rest of the sequences. A last covariance model was used in order to combine the sequences of clade 3 to the alignment. The R package R4RNA v.0.1.3 "R package for RNA visualization and analysis" was used to obtain a graphical output of the proportion of covariation across the alignment (Lai et al., 2012). In house R and PERL scripts allowed for analyzing the variability of the nucleotide per site across the alignment of sequence, the nucleotide composition and the covariation per site (the composition in nucleotides for a base pair).

2.1.2 Viral population sequences analysis

The pipeline summarizing the main steps to sort and analyze the sequences is summarized in fig. 2.1. The PERL scripts that allowed sorting and analyzing the sequences were written by Dr Martin Pelchat and Mrs Lynda Rocheleau. The R Project for statistical computing was used to generate the graphical outputs (R Core Team, 2013). Briefly, each sequence was sorted according to the composition in nucleotides, the length and the sequencing quality


A
FASTQ: Sequence (FASTA) + nucleotide sequencing quality information (Phred score)
Assess to each sequence a name with a quality information
ClustalW is used to establish an identity score by using a reference sequence
Sequence name
Quality score
Length
ID score (-)
ID score (+)
Visualization of the distribution of the sequence population according to the sequence length and identity score
=> Establish a cut-off of length and % identity to select sequences to be included in the analysis
Rarefaction curves to assess the sequencing depth
Distribution of the sequences according to the global mutation rate of the sequences => remove sequences due to the errors of the sequencing technology

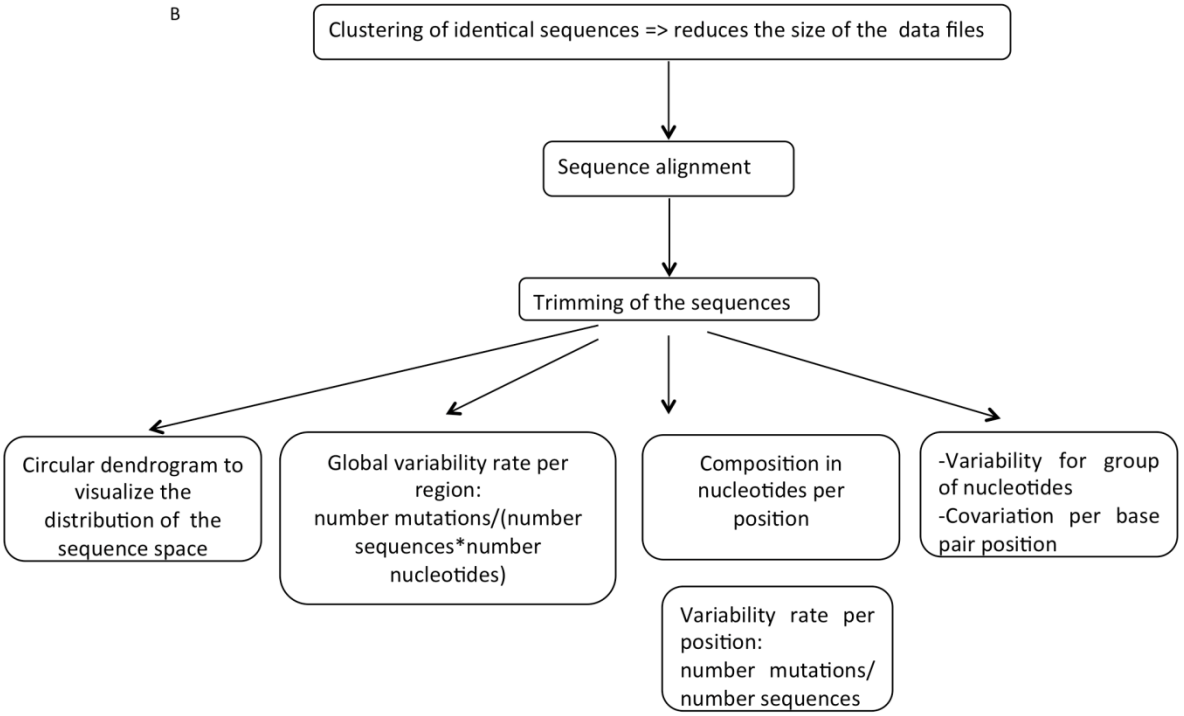

B
Clustering of identical sequences => reduces the size of the data files
Sequence alignment
Trimming of the sequences
Circular dendrogram to visualize the distribution of the sequence space
Global variability rate per region:
number mutations/(number sequences*number nucleotides)
Composition in nucleotides per position
-Variability for group of nucleotides
-Covariation per base pair position
Variability rate per position:
number mutations/ number sequences

Figure 2.1 Overview of the bioinformatics strategy to analyze the viral population sequences

(A) Sorting and filtering of the sequences. The sequences and their Phred score information were stored in a mySQL database. Using the clustal tool, an identity score to the Hepatitis *Delta* genome was computed for each sequence (Camacho C, 2008; Thompson, 1994). By taking in account the identity, the length, and the error rate of the control sample, a cut-off was established to choose the sequences to be used in the sequence alignment by Mosaik program (Lee et al., 2014). This same cut-off was used to select sequences to be further analyzed. **(B) Analysis of the sequences alignment.** The identical sequences were clustered and the sequences were aligned. The ends of the sequences were trimmed. In house PERL and R scripts allowed for obtaining the composition of nucleotides and base-pairs at each position of the alignment. From this data, the global variability and the variability rate per position were calculated. An R script was used to generate a circular dendrogram to represent the RNA sequence space.

score. The percentage of identity of each sequence to two reference HDV sequences of genomic and antigenomic polarities was calculated using Clustal 2.1 (Camacho C, 2008; Thompson, 1994) and this identity score was stored in a MySQL database (Widenius and Axmark, 2002). A PERL script was used to include in each sequence name the information of the sequencing quality score and the identity score for each polarity. A PERL script allowed for calculating the frequency and length of the sequences. The information of the length of the sequence, the identity score and the frequency of the sequences was represented as a 2D dot-plot, which allowed for establishing a cut-off to select the sequences to be used in the alignment. The sequences were filtered to remove sequences with a length (<160 nucleotides) or identity score (<60%). The sequences filtered were aligned with Mosaik 1.3.0 with the parameters –hs 15 (hash size), -mm 0.05 (mismatches) and –act 55 (minimum length of base pairs), as recommended for the 454 Titanium sequencing technology in the documentation of Mosaik (Lee et al., 2014). The sequences were trimmed from 160 to 140 nucleotides. In house PERL scripts allowed for calculating the variability of the sequences while taking in account the occurrence of the sequences. First, we generated a new fasta file that clustered the identical sequences and indicated the frequency of the sequence in its name. We obtained different fasta files that contain sequences only present at a given number of occurrences. A boxplot allowed for visualizing the correlation between the occurrence of identical sequences and the global variability of the sequences, which allowed for establishing a cut-off to select the clusters to be further analyzed. The phylogenetic distances between the sequences was calculated with the R package APE with the function "newick2phylog" and using a neighbor-joining method; the graphical output of the circular dendogram was obtained with a modified radial.phylog R script of APE4 (Thompson, 1994 ; Paradis, 2004; Dray and Dufour, 2007). An in house PERL script allowed obtaining the

frequency of nucleotides for each position. The frequency of the consensus nucleotide per position was represented as a bar plot and this allowed detecting hot spots of variability. The threshold that defines the background variability was calculated by using the R Parody package, which contains statistical methods of four outlier detection ("GESD","boxplot", "medmad" and "shorth" ; Carey V.J.). In order to generate a secondary structure model, Mfold web server version 3.5 was used (Zuker, 2003). A PERL script was used to obtain the nucleotide composition and frequency per position, which allowed for measuring the base pair covariation for the double-stranded and the bulge regions.

2.2 Molecular and Cellular Biology techniques

2.2.1 Cell Culture

A cellular system allowing the replication of HDV was kindly provided by John Taylor (Chang et al., 2005; fig. 2.2). 293 cells were stably transfected with a plasmid encoding the HDAg-S under the control of TET and an HDV genome deleted of two nucleotides in its unique *Eco*RI site, and therefore deleterious for the HDAg production. 293 cells were grown and HDV replication was induced as described previously (Chang et al., 2005). Briefly, 293-Ag and 293-HDV were grown at 37°C with 5% CO2 in Dulbecco's Modified Eagle's Medium (DMEM) with 10% calf serum (CS) added with 200 µg/mL hygromycine and 5 µg/mL blasticidin. The viral replication was induced upon addition of 1mg/mL of TET and two days later the total RNA was extracted with TRIzol (Invitrogen) according to the manufacturer's recommendations.

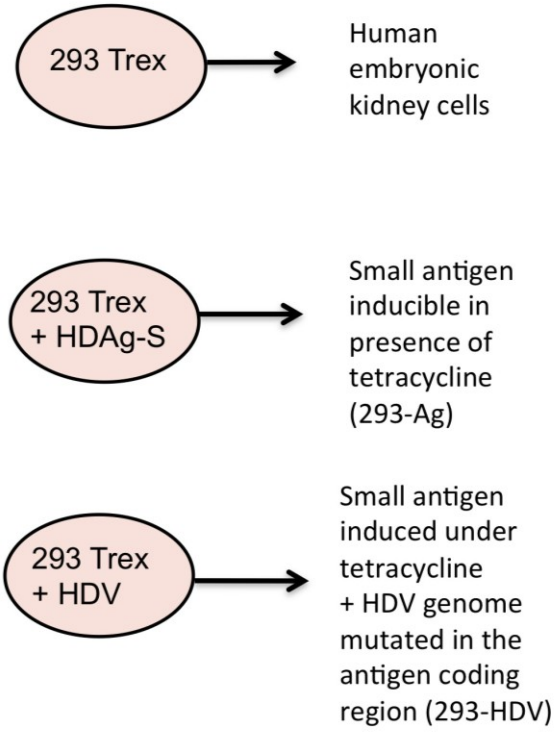

293 Trex
Human embryonic kidney cells
293 Trex + HDAg-S
Small antigen inducible in presence of tetracycline (293-Ag)
293 Trex + HDV
Small antigen induced under tetracycline + HDV genome mutated in the antigen coding region (293-HDV)

Figure 2.2 Overview of the cellular system allowing HDV replication.

293 cells Fl-In T-REx were stably transfected with a plasmid encoding the HDAg-S under the control of a TET promoter (293-Ag). This cell line was transfected with a genomic HDV RNA with a mutation in the HDAg-S coding region, and is therefore unable to synthesize its own HDAg-S (293-HDV). The addition of TET allows the production of the HDAg-S and therefore allows inducing HDV replication.

2.2.2 Synthesis of HDV RNA

To generate the RNA used as a control for the high throughput sequencing, the pHDVd2 plasmid was used. The pHDVd2 plasmid is a derivative of the pBluescriptKS$^+$ cloning vector (Stratagene) which contains a dimer of HDV cDNA (Kuo et al., 1988a). *Escherichia coli* XL-1 Blue $CaCl_2$ competent cells (Invitrogen) were transformed with pHDVd2 by the heat-shock method. The bacteria were thawed on ice for 10 min. An aliquot (50 µl) of competent cell was mixed with approximately 5 ng of pHDVd2. Immediately after an incubation at 42°C for 30s, the tube was placed on ice for 2 min. 950 µl of LB-medium with 100 µg/µL ampicillin was added and this suspension was incubated with shaking at 225-250 rpm at 37°C for 1 h. After incubation, 100 µl were plated on an agar plate containing LB at 100 µg/mL ampicillin and incubated at 37°C overnight. Single colonies were grown in LB medium with shaking at 225-250 rpm at 37°C overnight. After growth of the bacteria, the plasmid was isolated using QIAprep Spin Miniprep Kit (Qiagen). The R199G region was amplified by Polymerase Chain Reaction (PCR) and the forward primer contained a T7 promoter region at its 5' terminal end (Appendix III). The RNA was synthesized in an *in vitro* run-off transcription assay with 1 µg of the generated DNA template and 200 units of T7 RNA polymerase (New England Biolabs) in a transcription buffer (80mM Tris-HCl, pH 7.9, 40mM dithiothreitol, 20mM $MgCl_2$, 2mM spermidine and 1.25mM of each NTP), for 4 hours at 37 ^oC. Then, the reaction was subject to DNase treatment with 1 unit of DNase enzyme (Promega) at 37 ^oC for 30 min. The newly synthesized RNA was separated on a 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). Then the RNA corresponding to a size of 199 nucleotides was visualized by UV shadowing, the band excised and the RNA eluted in 1mL 500mM ammonium acetate, 0.1% SDS and 10mM

EDTA overnight on a rotator at 4 °C. The RNA was then precipitated in 1mL ethanol containing 0.3 M NaOAc at pH 5. The purified RNA was resuspended in 100 µl double distilled water (ddH_2O), desalted through a Sephadex G-50 column, precipitated again with ethanol containing 0.3 M NaOAc, washed with ethanol 70% and finally resuspended in 20 µl ddH_2O and stored at -80°C. An aliquot of the RNA was used to determine the RNA concentration by UV spectrophotometry at 260 nm.

2.2.3 Reverse transcription, Polymerase Chain Reaction (PCR)

The Reverse Transcription (RT) was carried out with the iScript cDNA synthesis kit (Biorad), following the manufacturer recommendations. The random primers were used to synthesize the total cDNA with 1 µg of RNA in a 20 µL total volume reaction. PCR were performed with the high-fidelity Deep-Vent polymerase according to the manufacturer's recommendations (NEB). All the primers used are indicated in Appendix III. The cycles used were as follow: 5 min at 25°C, 30 min at 42°C, 5 min at 85°C. For the deep-sequencing, the PCR fragment identity was confirmed by Sanger sequencing (Stem Core Facilities, Ottawa Research Institute). The final sample was a pool of 1 µg of the viral population DNA and 10 ng of the control, which was sent for high-throughput sequencing with the Roche 454 GS FLX Titanium technologies (McGill sequencing facilities, Genome Quebec). The control and viral population sequences are accessible on NCBI [SRA: SRR765851, SRR765852].

2.2.4 Quantitative PCR

After TRIzol extraction of the total RNA, the RNA concentration was assessed by spectrophotometry at 260 nm, the ribosomal RNA quality was assessed on an agarose gel and the RNA was reverse-transcribed (as described in 2.2.3). The efficiency curves of the primers

used for the qPCR were generated in order to ensure that they had similar amplification efficiencies. The qPCR reaction mixes were prepared with the IQ Syber Green Supermix kit (Biorad), following the manufacturer recommendations. For a total volume reaction of 10 µL, 0.1 µL of total cDNA was used. The primers used are presented in Appendix III. In order to assess that there was no DNA contamination from the RNA extract used for RT, as a negative control, a PCR was performed on the RNA extract. The qPCR was performed with Chromo 4 (Biorad) with the Opticon software (Biorad) with the following cycles: 1 cycle of 95 ^{o}C for 3 min, 40 cycles of 95 ^{o}C for 30 s, 55^{o}C for 30 s and 72^{o}C for 30 s, and the calculation of the melting curves from 55 ^{o}C to 100 ^{o}C, with a constant increase of temperature of 0.2^{o}C. The melting curves peaks allowed for assessing that only the specific genes were amplified by the primers during the qPCR and that there were no primer dimers contamination. The Livak $2^{-\Delta\Delta Ct}$ method was used and the gene expression was normalized to the house keeping gene *Beta*-2 microglobulin and compared to the gene expression in cells non-replicating HDV (Livak KJ and Schmittgen TD, 2001). Each experiment was performed in technical triplicates and in three independent biological assays. The mean, standard deviation and the unpaired two-tailed Student's t-test were calculated in Excel.

2.2.5 Knockdown of proteins

First, 0.5x10^{6} 293 Trex cells, 293-Ag cells and 293-HDV cells were seeded in 6 wells plates with DMEM 200 µg/mL hygromycin and 5 µg/mL blasticidin and the day after, when they reached 60% confluency, they were transfected with the siRNAs. The knockdown of PSF, p54 and PSP1 was performed using commercially available (Santa Cruz) siRNA directed against PSF (#sc-37007), p54 (#sc-38163) and PSP1 (#sc-76279). The manufacturer indicated that each siRNA was a pool of 3 specific siRNAs between 19 to 25 nucleotides. As

a negative control, I transfected a "scrambled" siRNA (#sc-37007) which, according to the manufacturer, is an siRNA of 20-25 nucleotides not specific to cellular mRNA targets. As two other negative controls, I also included a sample with no treatement and a mock (H_2O) transfected sample. For the transfection, I used the Lipofectamine 2000 reagent and followed the instructions for the transfection of an RNA (Invitrogen). A ratio of 80 picomoles siRNA: 5 µL lipofectamine: 100 µL DMEM was used. Twenty-four hours post transfection, the expression of HDAg-S was induced by addition of TET at 1µg/mL in the 293-Ag and 293-HDV cells respectively. Twenty-four hours post-induction, the total cellular lysate were centrifuged at 3000 rpm for 5 min and separated in two samples. For one sample, the total RNA was extracted using TRIzol (Invitrogen) and resuspended in 20 µL ddH_2O. The RNA was used for RT-PCR and RT-qPCR with primers to amplify NEAT1 and HDV (Appendix III). For the other sample, the total proteins were resuspended in Laemmli 4X buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125M pH 6.8 Tris-HCl), heated for 5 min at 95 ^{o}C and cooled on ice. The knockdown of the proteins was confirmed by Western Blotting.

2.2.6 RNA-immunoprecipitation

This protocol was adapted from the report of Niranjanakumari S et al, with a few modifications (Niranjanakumari et al., 2002). First, $1x10^6$ 293-HDV cells were seeded and after twenty-four hours the HDV replication was induced by addition of 1 µg/mL TET. After twenty-four hours, the cells were lysed in Phosphate-Buffered Saline (PBS) by pipetting up and down and 2 mL of cells out of 10 mL were aliquoted in order to perform an RNA extraction and a RT-PCR that allows for ensuring that the HDV replication was induced. For the rest of the cells, 1% of formaldehyde was added to cross-link the cellular RNA and the

proteins. After 10 min of incubation at room temperature, 0.25 M of glycine, pH 7 was added to stop the reaction. After 5 min of incubation at room temperature, the cells were centrifugated at 3000 rpm for 4 min and washed twice with cold PBS, and then resuspended in 2 mL of RIPA buffer (50 mM Tris–HCl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl) supplemented with 1 % protease inhibitor cocktail for mammalian extracts (Sigma-Aldrich) and 1 µL RNAse inhibitor (Biobasic, 10 mg/mL). The cells were placed on ice and were mechanically lysed by pipetting, followed by an agitation step for 30 min at 4°C. The cellular lysate was cleared with a centrifugation of 20 000 rpm for 20 min at 4°C. Then, the supernatant was sonicated and the lysates stored at -80 °C. The co-immunoprecipitation was performed with the Protein G immunoprecipitation kit (Sigma-Aldrich). 5 µg of antibody was added to the immunoprecipitation spin columns containing 50 µL of prewashed dynabeads. The antibodies used were targeted against: PSF (#B92, Abcam), PSP1 (#SAB4200067, Sigma-Aldrich) and *Beta*-Actin (mouse monoclonal, #6276, Abcam). PSF was used as a positive control and *Beta*-Actin as a negative control for the binding of R199G and NEAT1 RNA. The antibody-dynabeads complex was incubated on a rotator overnight at 4 °C and then washed in PBST buffer (PBS with 0.02% Tween). 500 µL of cellular lysate was added to the dynabeads-antibody complex and incubated for 1 h at 4 °C, followed by 4 washing steps in 1X RIPA buffer. The dynabeads-antibody-antigen complex was eluted in 200 µL of a protein storage buffer (50 mM Tris-Cl, 0.5 mM EDTA, 10 mM DTT, 1 % SDS, pH 7.5). In order to reverse the crosslinking reaction, the samples were heated at 70°C for 45 min. The sample was separated in two fractions of 100 µL, one for RNA extraction and one for protein extraction. The protein lysate was resuspended in Laemmli buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M pH 6.8 Tris-HCl) and

heated at 95°C for 5 min. After migration of the protein on a Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE; 10%), the Sypro-Ruby staining (Sigma-Aldrich) or Coomassie staining (1% Brillant Blue, 50% methanol, 10% acetic acid, 40% H_2O) were performed to visualize the total proteins, or a Western Blot was carried out to detect the specific proteins. The RNA was extracted with 500 µL of TRIzol (Invitrogen), ethanol-precipitated and resuspended in 20 µL ddH_2O. The HDV and NEAT1 RNA levels were assessed by RT followed by PCR and qPCR (Appendix III).

2.2.7 Western-Blot

In order to measure the protein concentrations of some assays, Bradford protein assays were performed following the manufacturer recommendations (Biorad). The protein samples from cell lysates were mixed with Laemmli 4X buffer and heated for 5 min at 95°C, cooled on ice, and if required stored at -20 °C. The protein lysates were migrated on polyacrylamide gel electrophoresis SDS PAGE (10%) in running buffer (25mM Tris-HCl, 200mM Glycine, 0.1% SDS). For some gels, the total proteins were visualized with a Coomassie staining to ensure a proper migration of the proteins. The gel was destained overnight in 5% methanol, 7.5% acetic acid, 87.5% H_2O. For other gels, a Western Blot was carried out. The proteins were subject to an overnight transfer to a PVDF membrane at 4 °C, at 40 volts in a transfer buffer (48mM Tris, 39mM glycine, 20% methanol, 0.037% SDS). After the transfer, the membrane was stained with Ponceau according to the commercial indications (Sigma-Aldrich) to visualize the total proteins and ensure that the protein transfer was homogenous and that similar levels of total proteins was loaded in all lanes. After destaining with NaOH 0.1 M, the membrane was then thoroughly washed with water and TBS buffer (200 mM Tris, 5 M NaCl, pH 7.5). The PVDF membrane was blocked for 1 hour at room temperature in 5%

Bovine Serum Albumin (BSA) in 1X Tris-buffered saline Tween (TBST ; 200 mM Tris, 5 M NaCl, pH 7.5, 0.1% (v/v) Tween) on a rocking shaker and washed three times for 10 min in TBST. The membrane was incubated with the primary antibody diluted in TBST containing 3% BSA. The dilutions were the following: PSF (mouse monoclonal, #B92, Abcam) 1/500, p54 (rabbit monoclonal, #05-950 Upstate) 1/300, PSP1 (rabbit polyclonal, #SAB4200067, Sigma-Aldrich) 1/300, Beta-actin 1/500 (mouse monoclonal, #6276, Abcam). The Beta-actin was used as a loading control. After three washes with TBST, the membranes were incubated with the appropriate secondary antibody at a dilution 1/20000 in TBST with 3% BSA on a shaker. The secondary antibodies used were the rabbit anti-mouse IgG HRP (polyclonal, # ab6728, Abcam) and goat anti-rabbit IgG HRP (polyclonal, #ab6721, Abcam). After three washes in TBST and one wash in TBS, the blots were visualized using ECL reagent according to the manufacturer's recommendations (Thermo Scientific #32106). The membrane was then exposed to a photosensitive film and the film was scanned. The densitometric measurement was performed for each representative digitized blots using Image-J software. Each experiment was performed in three independent assays.

2.2.8 Fluorescence Immunostaining

Cells were plated in cell culture dishes with glass bottom (Ibidi). Fourty-eight hours after the induction of HDV replication with TET at 1µg/mL, an immunostaining was performed. The cells were fixed with 4% paraformalhehyde in 1X PBS, washed three times in PBS containing 1% BSA and permeabilized 15 min at room temperature (RT) with 0.5% Triton X-100 at RT. The cells were incubated with primary antibodies diluted in PBS containing 1% BSA at RT for 1 h. The different dilutions used for the primary antibodies were: mouse polyclonal Y12 (kindly provided by the laboratory of Dr Jocelyn Côté) diluted 1:100, mouse

monoclonal PSF (#B92, Abcam) diluted 1:300, rabbit monoclonal p54 (#05-950 Upstate) diluted 1:300, mouse monoclonal nucleolin/B23 (#B0556, Sigma-Aldrich) diluted 1:300, rabbit polyclonal antibody directed against the C-terminal domain PSP1 (rabbit polyclonal, #SAB4200067, Sigma-Aldrich) diluted 1:300, mouse monoclonal PABP (#SC32318, Cell Signaling) diluted 1:200. After three washes in PBS containing 1% BSA, the cells were washed with 0.1% Triton X-100 in PBS and incubated with the appropriate secondary antibodies diluted at 1:300 in PBS containing 1% BSA, in the dark at RT for 1 h. The secondary antibodies used were as following: polyclonal goat anti-mouse coupled to Alexa 488 (#A11001, Life Technologies), polyclonal anti rabbit coupled to Alexa 488 (# A11008, Life Technologies), polyclonal goat anti-mouse coupled to Alexa 594 (A-11005, Life Technologies) and polyclonal goat anti-rabbit coupled to Alexa 594 (# A11012, Life Technologies). After a washing with 0.1% Triton X-100 in PBS, the cells were mounted in Vectashield containing DAPI (4', 6-Diamidino-2-Phenylindole) staining. The cells were visualized using AxioImager inverted Z.1 microscope (Carl Zeiss Canada) with the objectives PLAN APOCHROMAT 40X/0.95 and PLAN APOCHROMAT 63X/1.4 OIL, and analyzed with the Axiocam software. The immunofluorescence pictures shown are the combination of at least 25 Z stacks, to better represent the entire layer of the cell. Each experiment was performed in three independent assays.

2.2.9 *In situ* hybridization

A day after plating 293-HDV cells, I added TET [1µg/mL] to induce HDV replication. Fourty-eight hours after the induction of HDV replication, I performed an RNA fluorescence in situ hybridization using a probe against NEAT1 (#SMF-2036-1Stellaris, Biosearch Technologies). The protocol of the manufacturer for the simultaneous FISH and

Immunofluorescence protocol was followed with minor modifications. Cells were plated in cell culture dishes with glass bottom (Ibidi). Fourty-eight hours after the induction of HDV replication, the cells were washed twice in PBS and fixed with 4% paraformalhehyde in 1X PBS, washed three times in PBS containing 1% BSA and permeabilized 5 min at room temperature with 0.1% Triton X-100 for 5 min at room temperature. After a washing step in PBS, the cells were incubated with a polyclonal primary rabbit antibody directed against the C-terminal domain of PSP1 diluted 1:300 in PBS containing 1% BSA at room temperature for 1 h. After three washes in PBS containing 1% BSA, the cells were washed with 0.1% Triton X-100 in PBS and incubated with the secondary antibody anti-rabbit coupled to Alexa 488 (Life Technologies) diluted at 1:300 in PBS containing 1% BSA, in in the dark at RT for 1 h. After three washes in PBS, the cells were incubated in the fixation buffer (3.7% formaldehyde in 1X PBS) for 10 min at RT. After two washes in PBS, the cells were incubated in the wash buffer (10% formamide in 2X SSC) for 5 min at RT. Then, the cells were incubated overnight, at 37 ^{o}C, in the dark, with 200 µL hybridization buffer (100 mg/mL dextran sulfate, 10 % formamide in 2X SSC) containing 25 µM of NEAT1 probe. Then, the cells were washed in the hybridization buffer for 30 min in the dark at 37 ^{o}C and they were mounted in Vectashield containing DAPI. The cells were visualized using AxioImager inverted Z.1 microscope (Carl Zeiss) with the objectives 40 and 63, and analyzed with the Axiocam software. The immunofluorescence pictures shown are the combination of at least 25 Z stacks, to better represent the entire layer of the cell. Each experiment was performed in three independent assays.

2.2.10 Analysis of microcopy pictures

The pictures were visualized, acquired and analyzed using AxioVision Rel 4.8 software. Linear adjustements were made in Axiovision software to remove the background after acquisition of the pictures. Twenty-five cells of a representative field of cells replicating HDV were analyzed. I took at least 25 Z stacks (different focus at regular intervals across the section) during the acquisition of the pictures. The images were deconvolved using AxioVision Rel 4.8 to remove the background and readjust the fluorescent signal that is out of focus. My method was adapted from the reading of a previous report on analysis of fluorescent microscopy pictures of M A Williams and colleagues (Williams, 2013). The extended focus module of Axiovision Rel 4.8 allowed for calculating a sharp image from several Z stack images (taken at different focus across the sample). The colocalization of PSF or p54 with PSP1 was performed and quantified manually, by considering only the brighter small dots colocalizing. In order to highlight the brighter foci of the whole image, I used the Levels tool from Photoshop CC 2014 to reduce background signal and increase higher signal. I have also added a high pass filter in order to reduce blurriness and sharpen the edges. The number of spots per cell were counted using counter tool module from Photoshop. For quantification of colocalization by the software, a representative area of 25 cell was analyzed by the colocalization analysis extended module of Axiovision Rel 4.8 and the ratio of the two fluorophores corresponding to PABP and PSP1 was represented. The correlation of the distribution of the two fluorophores was calculated and the Pearson coefficient and Manders coefficient were generated. For the NEAT1 and PSP1 analysis, the two corresponding colour channels were overlaid. In order to decrease the uneven background and highlight the brighter foci, the image layer and high-pass filters were added. Then all the layers were selected and flattened into one layer. The counting of NEAT1 foci per cell was done by using the Count tool of Photoshop. All the foci counts were entered in Excel and the R software

was used to generate box plots. Average on 25 cells, n=3 independent experiments. In order to measure the area of the NEAT1 foci in Photoshop, a scale was set up. The ruler was used to measure the scale and the known linear distance was entered in µm. This length was automatically converted in a number of pixels by the software. The blue channel (DAPI) was hidden in the channel window and the red channel (NEAT1) was then grey scaled. The outlines of the individual NEAT1 foci were selected with the lasso tool. Under the analysis menu, the record function of Photoshop was pressed and the area of the selected foci in square micrometers was displayed in the measurement window. All the values were entered in an Excel table and the R software was used to generate a graphical output, the boxplot of the quantification of the area of NEAT1 foci. The mean and the median grey value of the area of the outline areas were also displayed by the record measurements. A student t-test was performed to calculate the p-value in R.

CHAPTER 3 Analysis of sequences of the right terminal region of HDV RNA

Data included in sections 3.1 was published in Virology (2014) (Appendix VI)

Author Contributions

The experiments were performed by Yasnee Beeharry, the bio-informatic scripts were written by Mrs Lynda Rocheleau and Dr Martin Pelchat, the Infernal alignment and the generation of the arc diagram (fig. 3.2.1) were performed by Dr Martin Pelchat.

Chapter 3.1 -Analysis of the sequences from an HDV population

3.1.1 Deep-sequencing of the right terminal region of HDV RNA derived from a viral population replicating in 293 cells

Previously, *in vitro* immunoprecipitation assays were performed in the laboratory to study the features of R199G that are necessary for its interaction with the RNAP II (Greco-Stewart et al., 2006). Mutational assays that disrupted the secondary structure led to an abrogation of the interaction with RNAP II. Altogether, these data suggested the importance of its RNA secondary structure, rather than its RNA primary sequence. During my PhD work, I investigated the conservation of the fragment of 199 nucleotides located at the right terminal region of the genomic HDV RNA in a viral population actively replicating in a cellular system. This work was published in Virology in 2014 (Beeharry et al., 2014, Appendix VI). Previous attempts to develop a cellular system to study the infection of HDV in a stable manner have remained unsuccessful. However, the J. Taylor team developed a system that allows for inducing a stable HDV replication (Chang et al., 2005). Briefly, 293 celluls were stably transfected with a plasmid allowing for the synthesis of the HDAg-S under the control of a TET-on promoter. This cell line was stably transfected with a HDV RNA genome deleted for two nucleotides in a unique *Eco*RI site located within the HDAg coding region, and therefore unable to produce the HDAg. There is still a very basal level of HDV replication without TET, and the addition of TET induces a strong amplification of this pool of HDV sequences (Chang et al., 2005). This cellular system was a good model to study the selective pressure favoring HDV sequences able to replicate.

In order to investigate the RNA promoter features conserved among a whole viral population, I grew this cell line for more than a year to allow the accumulation of mutations and I added TET to allow for the amplification of HDV genomes able to replicate in this cell

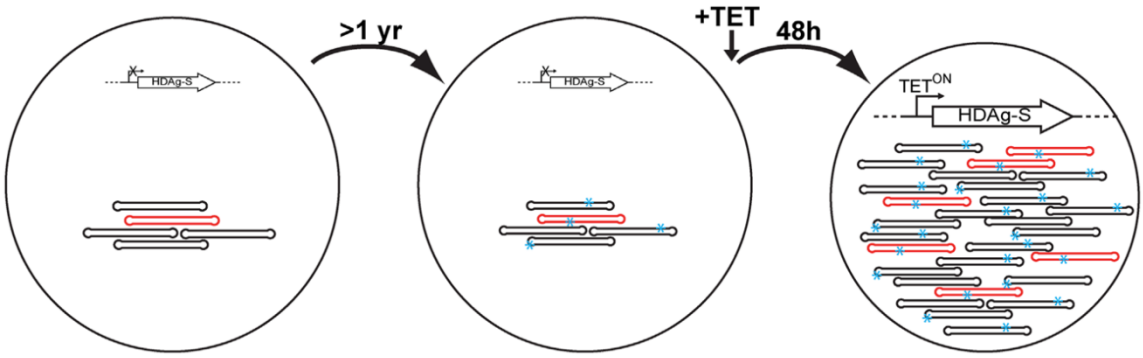

>1 yr
+TET
48h
HDAg-S
HDAg-S
TET^ON
HDAg-S

Figure 3.1.1 Overview of the experimental steps to generate the HDV population.

A 293 cell line has been stably transfected with an HDV RNA genome containing a frame-shift deletion in the HDAg ORF and a plasmid containing the HDAg-S gene under the control of a TET inducible promoter. This cell line has been maintained for more than a year, allowing HDV replication at a basal level and allowing the accumulation of mutations. Addition of TET allowed HDV RNA production and amplification of functional or even ameliorated HDV genomes (taken from from Beeharry et al., 2014).

line (fig. 3.1.1). After 2 days, total RNA was extracted by TRIzol. The viral population RNA was reverse-transcribed using degenerated primers and amplified by PCR using primers designed to specifically amplify the R199G region, which includes the initiation site of transcription, and the beginning of HDAg ORF (Abrahem and Pelchat, 2008; Greco-Stewart et al., 2006). The PCR primers also contained a ligated adapter for deep sequencing by 454. I checked the quality of the amplified fragment on an agarose gel and detected only one band of the expected size (data not shown). The identity of the sequence was confirmed by Sanger sequencing (Stem Core Centre, Ottawa). To control for errors due to the reverse-transcriptase, the PCR and the deep sequencing process, I performed the same experiments with an RNA species made by *in vitro* transcription using an RNA template of an HDV sequence of the same region. This control sequence corresponds to the RNA that was initially transfected in the 293-HDV system used to generate the HDV population. Since it likely shares sequence similarities with the viral population sequences, it is a good control to establish a cut-off to distinguish experimental errors from mutations due to the polymerase during the viral replication. Both populations tagged with a bar-code were multiplexed, with a ratio control:viral population of 1:100, and sent for deep-sequencing by the Roche 454 technology (Genome Quebec sequencing facilities, Mc Gill). I obtained 2510 sequences for the control and 747 158 sequences for the viral population.

3.1.2 Filtering of sequences obtained by deep-sequencing

In order to discriminate between the sequences resulting from experimental errors and the sequences corresponding to HDV, I established a cut-off to only select the HDV sequences of high quality and of a sufficient length for further analysis. Samples from the total reads from each experiment, control and 1 year passage, were used to calculate each dot

on the graph. By using an HDV reference sequence, an identity score was computed to each sequence. The sequences were organized in a database according to their Phred quality score, the sequence length and the identity score. The graph of fig. 3.1.2 shows the correlation of the frequency (upper graphs), identity score and length of each sequence (lower graphs). For both the control and the viral population, most of the sequences formed a population of the expected size, 200 nucleotides and with at least 75% identity to the reference sequence of genomic polarity (dots in blue, fig. 3.1.2). The shorter sequences had low identity scores and most corresponded to chimeras, half composed of HDV and half of unidentified sequence, as previously observed during deep-sequencing (Gorzer et al., 2010). The sequences of antigenomic polarity had a low identity score (red dots, fig. 3.1.2.B), as expected since the primers used during the PCR were specific of the genomic polarity of R199G. We carried out the analysis with the dominant population of sequences, that had an identity score superior to 60% and of at least 160 nucleotides to obtain the whole promoter region (blue population, fig. 3.1.2.B). A total of 490,183 sequences (65.6%) for the viral population and 2,070 (82.4%) sequences for the control were selected for further analysis. Of note, while looking at the totality of the selected sequences, for 0.02% of the viral population, a deletion of a stretch of nucleotides was found in the region of initiation of transcription, upstream from the loop, from position 1643 to position 1631 (Appendix IV). This deletion was not found in the control population.

In order to ensure that enough sequencing depth was reached, I calculated rarefaction curves that represent the number of variants while increasing the number of reads (fig. 3.1.3). For both populations, the curve reaches a plateau showing that most of the variants are represented (fig. 3.1.3). More variants were present for the viral population than for the control population, suggesting a larger sequence space for the viral population.

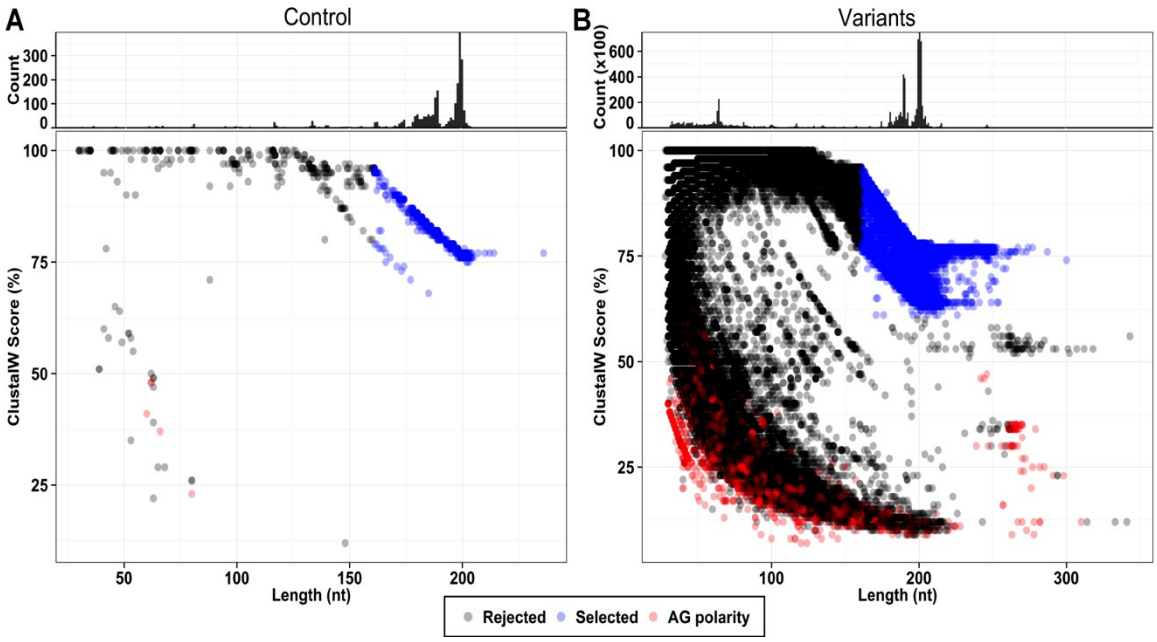
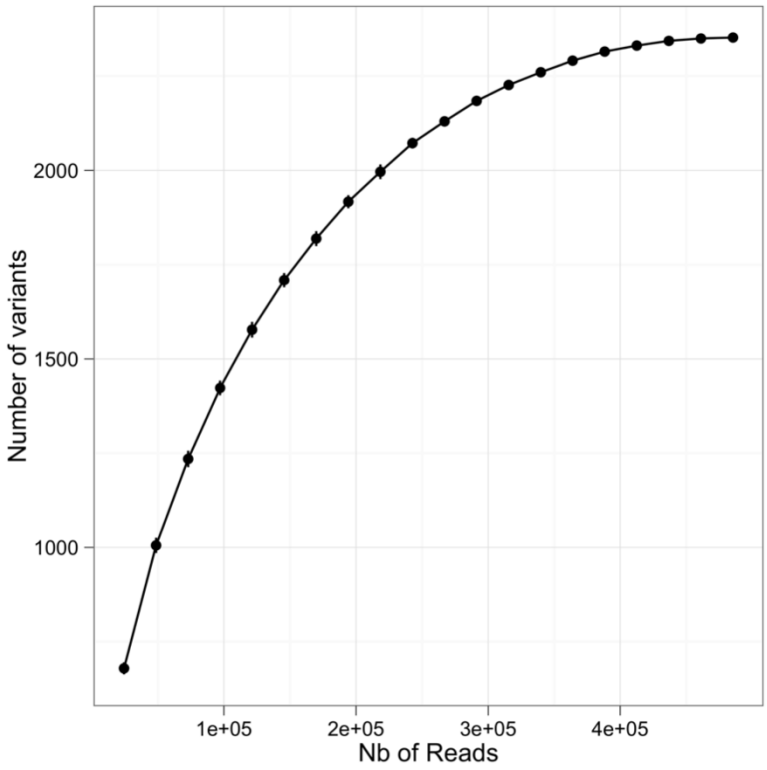

A
Control
B
Variants
Count
Count (x100)
ClustalW Score (%)
Length (nt)
Rejected
Selected
AG polarity

Figure 3.1.2 Filtering of the reads obtained by deep- sequencing.

Filtering of the reads obtained by deep-sequencing using the 454 technology according to the sequence length and percentage of identity to the reference sequence, for both the control (**A**) and the sequences amplified from 293-HDV cells (**B**). Top parts represent the number of sequences sorted according to their lengths. Bottom parts represent the lengths and the percentage of identity of each read to the reference sequence, as calculated by ClustalW (Thompson et al., 1994). Black and red dots indicate sequences with higher identities to the genomic and antigenomic polarity of the reference sequence, respectively. Sequences selected for further analysis are represented by blue dots. These sequences are at least 160 nucleotides long and are 60% identical to the reference sequence (taken from Beeharry et al., 2014)).


Number of variants
2000
1500
1000
1e+05
2e+05
3e+05
4e+05
Nb of Reads

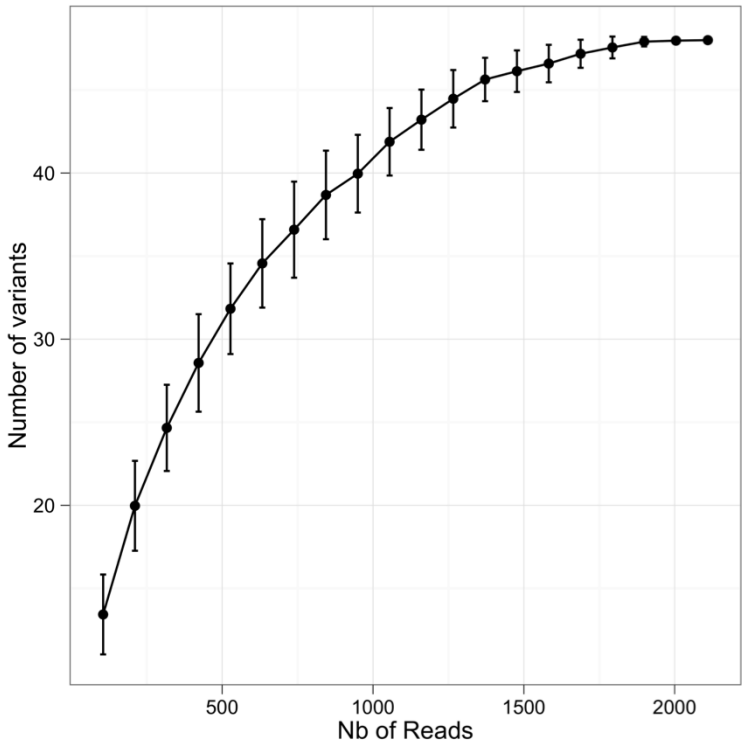

Number of variants
40
30
20
500
1000
1500
2000
Nb of Reads

Figure 3.1.3 Rarefaction curves to assess the sequencing depth.

(Upper) Viral population (Lower) Control population The number of HDV variants is represented as a function of the number of reads. The rarefaction curves of both populations reach a plateau, indicating that the totality of the variants are represented and that the sequencing depth is adequate. The calculation of the curve was reiterated 100 times.

The 490,183 filtered sequences for the viral population were clustered in identical sequences and were aligned with Mosaik, with a sequence composing the viral population used as a reference sequence (Lee et al., 2014). The sequences of the control population were aligned to the reference sequence obtained by Sanger sequencing. For certain positions of the alignment I found gaps, which might represent nucleotide insertions in the sequences. However, because there was no method to differentiate them from the experimental errors, I did not take into account insertions since this would cause an overestimation of the error-rate. To calculate the sequence variability of the R199G region, I only analyzed single nucleotide polymorphism (SNP) and deletions.

3.1.3 Occurrence of the sequence and global mutation rate

The composition in nucleotides at each position of the alignment was investigated. In order to calculate the global mutation rate, the total number of SNPs was divided by the total number of nucleotides (length of the sequences multiplied by the total number of the sequences). However, the unique sequences (occurring only once in the population of sequences) are likely due to experimental errors, during the reverse transcription or the deep-sequencing, and are not representative of the selected variation (Beerenwinkel et al., 2012). Therefore, this analysis was also repeated while filtering the sequences by their number of occurrences (fig. 3.1.4; Table 3.1.1). The variability of the control population is representative of the background variability, inherent to the experimental processes. When considering sequences that occur at least 3 times, only 7 positions of the alignment of the control sequences vary and these mutations were not present in the viral population. Furthermore, when considering sequences present at least 3 times, the number of sequences

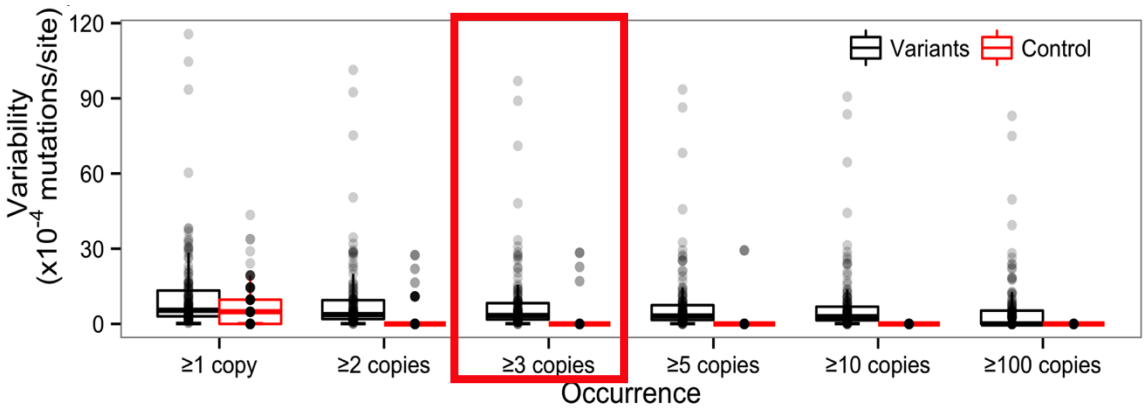

Variability
(x10-4 mutations/site)
120
90
60
30
0
Variants
Control
≥1 copy
≥2 copies
≥3 copies
≥5 copies
≥10 copies
≥100 copies
Occurrence

Figure 3.1.4 Reduction of the background nucleotide variability by removal of sequences with low occurrence.

The composition of nucleotides per position of the alignment was calculated. The global mutation rate (y axis) was calculated by using sequences occurring at least 1, 2, 3, 5, 10 and 100 times (x axis). The results are represented as a box-plot. Each dot represents the global variability of a read. Black and red boxes (region between the first and the third quartile of the populations) indicate the nucleotide variability for the variants and the control populations, respectively. When considering the sequences occurring at least 3 times (indicated by a bold red box), the sequence variability of the control population is greatly reduced, while the sequence variability of the viral population is not greatly changed. Therefore, only sequences that occurred at least 3 times were selected for further analysis since this allowed to reduce the background nucleotide variability (taken from Beeharry et al., 2014).

Position	Secondary Structure	Sequence	A	C	G	U	Purine	Pyrimidine	Transversion	Transition	Variability ($x10^{-4}$ mutations/site)
1566	.	U	21	32	3	473080	24	473112	24	32	1.25
1567	.	C	20	472591	4	524	24	473115	24	524	11.58
1568	.	C	30	472830	0	235	30	473065	30	235	6.53
1569	.	G	136	4	472937	48	473073	52	52	136	4.27
1570	.	C	24	472924	5	186	29	473110	29	186	4.54
1571	.	G	278	9	472777	75	473055	84	84	278	7.65
1572	.	G	360	14	472698	50	473058	64	64	360	9.32
1573	.	U	135	730	0	472274	135	473004	135	730	18.28
1574	.	U	15	366	6	472544	21	472910	21	366	12.58
1575	.	C	19	473066	16	5	35	473071	35	5	1.54
1576	.	U	11	128	4	472996	15	473124	15	128	3.02
1577	.	U	9	80	0	472919	9	472999	9	80	4.65
1578	.	C	30	473055	7	47	37	473102	37	47	1.78
1579	.	C	47	472953	15	54	62	473007	62	54	3.93
1580	.	U	42	83	0	473008	42	473091	42	83	2.77
1581	.	C	41	473029	0	61	41	473090	41	61	2.32
1582	.	G	269	3	472742	78	473011	81	81	269	8.39
1583	.	A	472924	13	162	15	473086	28	28	162	4.54
1584	.	C	26	473040	6	62	32	473102	32	62	2.09
1585	(	U	30	26	5	473075	35	473101	35	26	1.35
1586	(	C	35	473020	14	66	49	473086	49	66	2.52
1587	(	G	225	4	472864	46	473089	50	50	225	5.81
1588	(	G	194	3	472870	62	473064	65	65	194	5.69
1589	(	A	472965	0	87	11	473052	11	11	87	3.68
1590	(	C	10	473014	3	112	13	473126	13	112	2.64
1591	.	C	36	472980	0	119	36	473099	36	119	3.36
1592	(	G	191	15	472905	28	473096	43	43	191	4.95
1593	(	G	1236	24	471786	43	473022	67	67	1236	28.60
1594	(	C	44	472937	3	148	47	473085	47	148	4.27
1595	(	U	71	262	25	472757	96	473019	96	262	8.07
1596	(	C	14	473085	5	35	19	473120	19	35	1.14
1597	(	A	468553	0	4583	0	473136	0	0	4583	96.93
1598	.	U	0	238	3	472895	3	473133	3	238	5.16
1599	(	C	23	473053	0	60	23	473113	23	60	1.82
1600	(	U	42	51	3	473043	45	473094	45	51	2.03
1601	(	C	47	472965	7	120	54	473085	54	120	3.68
1602	(	G	114	0	472979	43	473093	43	43	114	3.38
1603	(	G	607	5	472484	40	473091	45	45	607	13.84

1604	(	C	14	473073	0	52	14	473125	14	52	1.39
1605	(	U	51	301	4	472729	55	473030	55	301	8.67
1606	.	A	468928	3	4069	11	472997	14	14	4069	89.00
1607	(	G	52	0	473037	46	473089	46	46	52	2.16
1608	.	A	472969	6	130	18	473099	24	24	130	3.59
1609	(	G	84	0	472997	58	473081	58	58	84	3.00
1610	(	G	81	0	473016	39	473097	39	39	81	2.60
1611	(	C	22	472930	0	177	22	473107	22	177	4.42
1612	(	G	405	0	472687	47	473092	47	47	405	9.55
1613	(	G	1127	27	471932	48	473059	75	75	1127	25.51
1614	(	C	14	472714	0	411	14	473125	14	411	8.98
1615	(	A	472402	4	716	17	473118	21	21	716	15.58
1616	(	G	82	6	473022	19	473104	25	25	82	2.47
1617	(	U	12	128	0	472994	12	473122	12	128	3.06
1618	(	C	20	473101	0	18	20	473119	20	18	0.80
1619	(	C	24	473074	3	38	27	473112	27	38	1.37
1620	.	U	27	63	9	473040	36	473103	36	63	2.09
1621	(	C	13	473088	4	34	17	473122	17	34	1.08
1622	(	A	473092	4	36	4	473128	8	8	36	0.99
1623	(	G	55	5	473058	18	473113	23	23	55	1.71
1624	(	U	321	60	0	472724	321	472784	321	60	8.77
1625	.	A	472077	7	1016	16	473093	23	23	1016	22.45
1626	(	C	15	473074	5	21	20	473095	20	21	1.37
1627	(	U	3	28	0	473088	3	473116	3	28	1.08
1628	(	C	18	473073	0	12	18	473085	18	12	1.39
1629	.	U	10	351	5	472749	15	473100	15	351	8.24
1630	.	U	107	349	3	472492	110	472841	110	349	13.67
1631	.	A	470863	0	2180	74	473043	74	74	2180	48.10
1632	(	C	21	473023	6	38	27	473061	27	38	2.45
1633	(	U	4	26	0	473074	4	473100	4	26	1.37
1634	(	C	13	473096	0	0	13	473096	13	0	0.91
1635	(	U	0	6	0	473133	0	473139	0	6	0.13
1636	(	U	0	10	0	473099	0	473109	0	10	0.85
1637	(	U	0	0	0	473084	0	473084	0	0	1.16
1639	.	C	27	473051	6	4	33	473055	33	4	1.86
1640	.	U	4	26	3	473096	7	473122	7	26	0.91
1641	.	G	1495	9	471560	21	473055	30	30	1495	33.37
1642	.	U	328	488	5	472167	333	472655	333	488	20.54
1643	)	A	473063	0	76	0	473139	0	0	76	1.61
1644	)	A	473087	0	49	3	473136	3	3	49	1.10

1645	)	A	472808	0	259	0	473067	0	0	259	7.00
1646	)	G	27	0	473032	36	473059	36	36	27	2.26
1647	)	A	473076	0	39	3	473115	3	3	39	1.33
1648	)	G	150	9	472951	25	473101	34	34	150	3.97
1649	)	G	90	0	472979	12	473069	12	12	90	3.38
1650	)	A	473056	0	66	10	473122	10	10	66	1.75
1651	)	G	36	0	473029	68	473065	68	68	36	2.32
1652	)	A	473046	0	20	27	473066	27	27	20	1.97
1653	)	C	9	473055	9	47	18	473102	18	47	1.78
1654	)	U	6	71	13	473049	19	473120	19	71	1.90
1655	)	G	113	8	472981	31	473094	39	39	113	3.34
1656	.	C	26	472663	14	208	40	472871	40	208	10.06
1657	.	U	35	123	9	472934	44	473057	44	123	4.33
1658	)	G	148	41	472596	278	472744	319	319	148	11.48
1659	)	G	66	0	473033	10	473099	10	10	66	2.24
1660	)	A	472886	172	74	7	472960	179	179	74	5.35
1661	)	C	8	473048	5	54	13	473102	13	54	1.92
1662	)	U	32	45	6	473056	38	473101	38	45	1.75
1663	.	C	44	472694	43	347	87	473041	87	347	9.41
1664	)	G	7	1032	472085	0	472092	1032	1032	7	22.28
1665	)	C	0	473107	0	21	0	473128	0	21	0.68
1666	)	C	0	473139	0	0	0	473139	0	0	0.00
1667	)	G	652	0	472439	32	473091	32	32	652	14.79
1668	)	C	11	473018	0	110	11	473128	11	110	2.56
1669	)	C	25	473043	5	66	30	473109	30	66	2.03
1670	)	C	0	473075	0	14	0	473089	0	14	1.35
1671	.	G	464	3	472568	49	473032	52	52	464	12.07
1672	)	A	473078	4	42	4	473120	8	8	42	1.29
1673	)	G	167	0	472908	50	473075	50	50	167	4.88
1674	.	C	26	471839	10	1264	36	473103	36	1264	27.48
1675	)	C	11	473093	0	16	11	473109	11	16	0.97
1676	)	C	8	473093	0	38	8	473131	8	38	0.97
1677	)	G	125	0	472963	44	473088	44	44	125	3.72
1678	)	A	473024	4	79	26	473103	30	30	79	2.43
1679	)	G	183	10	472900	35	473083	45	45	183	5.05
1	.	C	61	471884	17	1177	78	473061	78	1177	26.52
2	.	C	4	473109	0	26	4	473135	4	26	0.63
3	)	U	34	3310	11	469777	45	473087	45	3310	71.06
4	)	G	66	0	473035	38	473101	38	38	66	2.20
5	)	A	472774	3	347	6	473121	9	9	347	7.71

6	)	G	154	0	472952	33	473106	33	33	154	3.95
7	)	C	19	472500	4	611	23	473111	23	611	13.51
8	)	C	0	473134	0	5	0	473139	0	5	0.11
9	.	A	473063	0	63	7	473126	7	7	63	1.61
10	.	A	473013	16	77	14	473090	30	30	77	2.66
11	)	G	58	0	473042	27	473100	27	27	58	2.05
12	)	U	9	1263	0	471854	9	473117	9	1263	27.16
13	.	U	34	124	0	472942	34	473066	34	124	4.16
14	.	C	50	472856	6	224	56	473080	56	224	5.98
15	)	C	0	473136	0	0	0	473136	0	0	0.06
16	)	C	43	472701	0	163	43	472864	43	163	9.26
17	)	G	144	0	472861	119	473005	119	119	144	5.88
18	)	A	471679	5	1417	20	473096	25	25	1417	30.86

Table 3.1 Statistics on the occurrence of variants of the right terminal domain of genomic HDV RNA obtained from high-throughput sequencing of a population replicating in 293 cells (taken from Beeharry Y et al, Virology, 2014).

The position 1597 has the highest variabily, $96.93.10^{-4}$ mutations/site. The variability is elevated for positions 1593, 1606, 1613, 1631, 1641, 1674, 1, 3, 12 and 18. The positions 1666, 8 and 15 are totally conserved for the populations with sequences present at least in 3 copies. The U residue at location 1638 of fig. 1.3 was not used in the analysis due to high variability caused by the homopolymer effects during high-throughput 454 sequencing (Huse et al., 2007).

of the viral population was not drastically reduced since only 3.48% of the sequences were removed. For sequences present in at least 3 copies, the mutation rate of the control was calculated to be $1.2x10^{-4}$ mutation/site and the one of the viral population was eight times higher, at $8.1x10^{-4}$ mutation/site. When considering the sequence occurrences, it was striking that one sequence occurs dominantly in the population: the sequence corresponding to the variant initially transfected in this cellular system represented 76.3% of the population or 360 863 sequences (Table 3.2; Chang et al., 2005). However, there were still 23.7% of different sequences composing this viral population (Table 3.2).

3.1.4 HDV exists as a heterogeneous population

In order to represent the genetic distances between the different R199G sequences, a neighbour-joining tree was generated and plotted as a circular dendrogram (fig. 3.1.5). The size of a cluster in the tree is proportional to the number of identical sequences composing this cluster (fig. 3.1.5). The tree was rooted on the sequence that occurs dominantly. The same scale was used for both the control and the viral populations. I observed that the sequence space is larger for the viral population than for the control: the viral population sequences form a heterogeneous population.

Number of occurence	Number of different reads	Total number of reads	Percentage of reads	Percentage of reads (occurrence >2)
1	13424	13424	2.74	-
2	1810	3620	0.74	-
3	692	2076	0.42	0.44
4	362	1448	0.3	0.31
5	208	1040	0.21	0.22
6	146	876	0.18	0.19
7	112	784	0.16	0.17
8	88	704	0.14	0.15
9	78	702	0.14	0.15
10	56	560	0.11	0.12
11	46	506	0.1	0.11
12	35	420	0.09	0.09
13	37	481	0.1	0.1
14	26	364	0.07	0.08
15	25	375	0.08	0.08
16	29	464	0.09	0.1
17	28	476	0.1	0.1
18	18	324	0.07	0.07
19	21	399	0.08	0.08
20	14	280	0.06	0.06
21	17	357	0.07	0.08
22	17	374	0.08	0.08
23	14	322	0.07	0.07
24	16	384	0.08	0.08
25	11	275	0.06	0.06
26	16	416	0.08	0.09
27	10	270	0.06	0.06
28	10	280	0.06	0.06
29	12	348	0.07	0.07
30	14	420	0.09	0.09
31	11	341	0.07	0.07
32	5	160	0.03	0.03
33	12	396	0.08	0.08
34	4	136	0.03	0.03
35	8	280	0.06	0.06
36	11	396	0.08	0.08
37	10	370	0.08	0.08
38	7	266	0.05	0.06
39	6	234	0.05	0.05
40	7	280	0.06	0.06
41	8	328	0.07	0.07
42	4	168	0.03	0.04
43	8	344	0.07	0.07
44	6	264	0.05	0.06
45	6	270	0.06	0.06
46	9	414	0.08	0.09
47	16	752	0.15	0.16

48	6	288	0.06	0.06
49	6	294	0.06	0.06
50	6	300	0.06	0.06
51	4	204	0.04	0.04
52	6	312	0.06	0.07
53	2	106	0.02	0.02
54	5	270	0.06	0.06
55	11	605	0.12	0.13
56	3	168	0.03	0.04
57	2	114	0.02	0.02
58	5	290	0.06	0.06
59	5	295	0.06	0.06
60	2	120	0.02	0.03
61	6	366	0.07	0.08
62	2	124	0.03	0.03
63	3	189	0.04	0.04
67	3	201	0.04	0.04
68	6	408	0.08	0.09
69	2	138	0.03	0.03
70	3	210	0.04	0.04
71	3	213	0.04	0.05
72	3	216	0.04	0.05
74	1	74	0.02	0.02
75	2	150	0.03	0.03
76	6	456	0.09	0.1
77	2	154	0.03	0.03
78	3	234	0.05	0.05
79	2	158	0.03	0.03
80	4	320	0.07	0.07
81	1	81	0.02	0.02
82	3	246	0.05	0.05
84	1	84	0.02	0.02
85	3	255	0.05	0.05
86	2	172	0.04	0.04
87	1	87	0.02	0.02
88	1	88	0.02	0.02
89	4	356	0.07	0.08
91	2	182	0.04	0.04
92	1	92	0.02	0.02
93	1	93	0.02	0.02
94	3	282	0.06	0.06
95	1	95	0.02	0.02
96	1	96	0.02	0.02
97	1	97	0.02	0.02
99	1	99	0.02	0.02
100	1	100	0.02	0.02
102	1	102	0.02	0.02
103	1	103	0.02	0.02
104	4	416	0.08	0.09

105	1	105	0.02	0.02
107	1	107	0.02	0.02
109	1	109	0.02	0.02
110	2	220	0.04	0.05
113	3	339	0.07	0.07
116	2	232	0.05	0.05
117	3	351	0.07	0.07
119	5	595	0.12	0.13
120	1	120	0.02	0.03
122	2	244	0.05	0.05
124	1	124	0.03	0.03
127	1	127	0.03	0.03
128	1	128	0.03	0.03
129	2	258	0.05	0.05
130	1	130	0.03	0.03
133	1	133	0.03	0.03
134	1	134	0.03	0.03
135	1	135	0.03	0.03
136	2	272	0.06	0.06
137	3	411	0.08	0.09
139	3	417	0.09	0.09
140	1	140	0.03	0.03
141	1	141	0.03	0.03
146	1	146	0.03	0.03
148	1	148	0.03	0.03
149	1	149	0.03	0.03
152	1	152	0.03	0.03
154	1	154	0.03	0.03
159	1	159	0.03	0.03
160	1	160	0.03	0.03
165	2	330	0.07	0.07
166	1	166	0.03	0.04
177	1	177	0.04	0.04
178	1	178	0.04	0.04
179	1	179	0.04	0.04
182	1	182	0.04	0.04
186	2	372	0.08	0.08
188	1	188	0.04	0.04
190	1	190	0.04	0.04
194	1	194	0.04	0.04
195	1	195	0.04	0.04
197	1	197	0.04	0.04
200	2	400	0.08	0.08
203	1	203	0.04	0.04
204	1	204	0.04	0.04
205	1	205	0.04	0.04
210	1	210	0.04	0.04
211	1	211	0.04	0.04
226	1	226	0.05	0.05

234	1	234	0.05	0.05
237	1	237	0.05	0.05
238	2	476	0.1	0.1
239	1	239	0.05	0.05
242	1	242	0.05	0.05
245	1	245	0.05	0.05
249	1	249	0.05	0.05
256	1	256	0.05	0.05
257	1	257	0.05	0.05
260	1	260	0.05	0.05
271	1	271	0.06	0.06
280	1	280	0.06	0.06
284	1	284	0.06	0.06
298	1	298	0.06	0.06
303	2	606	0.12	0.13
308	1	308	0.06	0.07
313	1	313	0.06	0.07
316	1	316	0.06	0.07
324	1	324	0.07	0.07
325	1	325	0.07	0.07
334	1	334	0.07	0.07
336	1	336	0.07	0.07
340	1	340	0.07	0.07
344	1	344	0.07	0.07
350	1	350	0.07	0.07
357	2	714	0.15	0.15
367	1	367	0.07	0.08
369	1	369	0.08	0.08
372	1	372	0.08	0.08
388	1	388	0.08	0.08
399	1	399	0.08	0.08
400	1	400	0.08	0.08
406	1	406	0.08	0.09
423	1	423	0.09	0.09
426	1	426	0.09	0.09
427	1	427	0.09	0.09
430	1	430	0.09	0.09
460	1	460	0.09	0.1
466	1	466	0.1	0.1
511	1	511	0.1	0.11
518	1	518	0.11	0.11
536	1	536	0.11	0.11
537	1	537	0.11	0.11
549	1	549	0.11	0.12
565	1	565	0.12	0.12
575	1	575	0.12	0.12
618	1	618	0.13	0.13
648	1	648	0.13	0.14
649	1	649	0.13	0.14

713	1	713	0.15	0.15
760	1	760	0.16	0.16
798	1	798	0.16	0.17
809	1	809	0.17	0.17
811	1	811	0.17	0.17
822	1	822	0.17	0.17
839	1	839	0.17	0.18
867	1	867	0.18	0.18
923	1	923	0.19	0.2
973	2	1946	0.4	0.41
990	1	990	0.2	0.21
1046	1	1046	0.21	0.22
1058	1	1058	0.22	0.22
1095	1	1095	0.22	0.23
1249	1	1249	0.25	0.26
1322	1	1322	0.27	0.28
1465	1	1465	0.3	0.31
1703	1	1703	0.35	0.36
2206	1	2206	0.45	0.47
2226	1	2226	0.45	0.47
2984	1	2984	0.61	0.63
3028	1	3028	0.62	0.64
3416	1	3416	0.7	0.72
8944	1	8944	1.82	1.89
9305	1	9305	1.9	1.97
360863	1	360863	73.62	76.27

Table 3.2 Analysis of the occurrences of the sequences of the right terminal domain of genomic HDV RNA obtained from high-throughput sequencing (taken from Beeharry Y et al, Virology, 2014).

The first column indicates the number of occurrences of any given variant, the second column gives the number of different variants, the third column the total number of reads for these variants. The two last columns indicate the percentage that this total number of reads for these variants (third column) represents compared to the total number of sequences (fourth column) and or compared to the population of sequences present at least in 3 copies (fifth column).

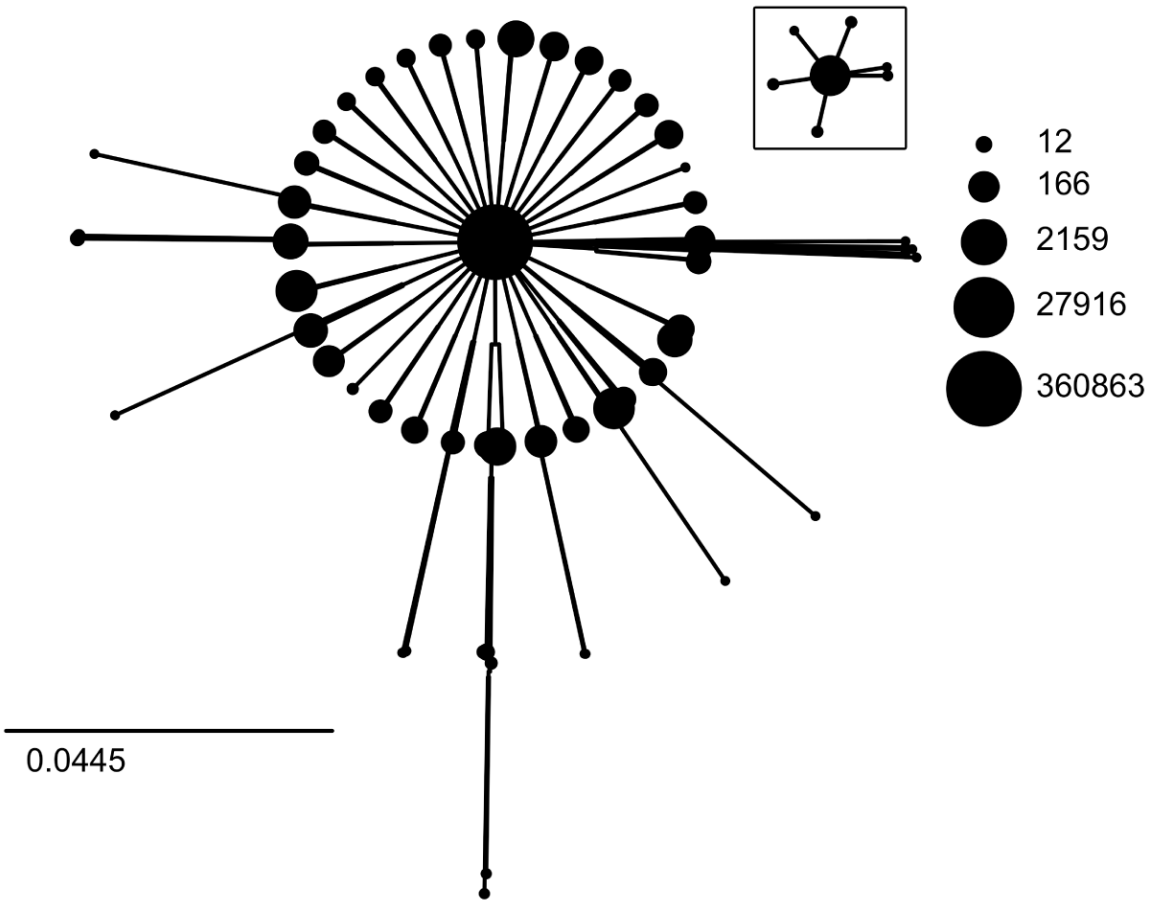

12
166
2159
27916
360863
0.0445

Figure 3.1.5 Evaluation of the sequence space occupied by the two populations.

Neighbour-joining phylogenetic trees rooted on the sequence that occurs dominantly were generated for both populations and plotted as circular dendrograms. The sizes of the clusters are proportional to the occurrence of the sequences composing this cluster (scale on right; $\log_2$ relationship). The inset represents the sequence space occupied by the control population, using the same scale as the viral population (scale on left; taken from Beeharry et al., 2014).

3.1.5 Analysis of the variability per site of HDV right terminal region

In order to calculate the variability at each position of the alignment, the number of SNPs was calculated and divided by the number of sequences. This variability rate was represented on the consensus sequence of R199G (fig. 3.1.6). A cut-off to distinguish the background variation from significant variability peaks was calculated by using four outlier tests (GESD, boxplot, medmad and shorth). A colour code allows the localization of the background variation, situated below the grey zone (fig. 3.1.6). Thirteen positions were found to have an "intermediate" variability, namely a variability above the cut-off of at least one of the statistical test (represented by the blue bar plots on fig. 3.1.6). Eleven positions were found to be hot spots of variability according to all the four statistical tests (the red bar plots on fig. 3.1.6). Notably, the highest mutation rate for this region corresponded to the second nucleotide of the first codon of HDAg-S, at position 1597: a mutation A>G for 1.2% of the viral population (4,583 reads; table 3.1.1). This UAC corresponds to the AUG initiation codon of the HDV mRNA (Gudima et al., 1999). For the nucleotides located at the tip of the stem, close to the loop, the primary sequence was more conserved (fig. 3.1.6 ; from position 1632 to 1557). For the control sequences, the variability per site of all these 24 positions remained below the background. Altogether, these results indicate that for the viral population sequences, the variability per site is not homogeneous for each position of the alignment and that the sequence at the tip of the stem is conserved. I also analyzed the nature of the mutations by looking at the composition in nucleotides at each position of the alignment. For the control, the proportion of transitions (mutation changing a purine to a purine or a pyrimidine to a pyrimidine) and transversions (mutation changing a purine to a pyrimidine or a pyrimidine to a purine) was comparable while for the viral population, the proportion of transitions was approximately 10 times higher than the transversions, namely

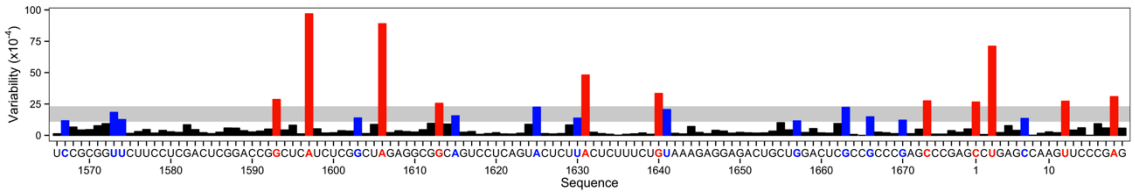

Variability (x10^{-4})
100
75
50
25
0
UCCGCGGUUCUUCCUCGACUCGGACCGGCUCAUCUCGGCUAGAGGCGGCAGUCCUCAGUACUCUUACUCUUUCUGUAAAGAGGAGACUGCUGGACUCGCCGCCCGAGCCCGAGCCUGAGCCAAGUUCCCGAG
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1
10
Sequence

Figure 3.1.6 Position-specific variability of the right terminal domain of genomic HDV RNA obtained from high-throughput sequencing.

Representation of nucleotide variability for each position obtained from the alignment. The consensus sequence displayed corresponds to a region from nucleotides 1566 to 18 of the genomic polarity of HDV RNA. Four outlier tests were used to identify positions that appear to deviate from background variations and their cut-offs was used to define the "grey zone". The calculated cut-offs were 15.58×10^{-4}, 23.33×10^{-4}, 10.79×10^{-4}, 11.19×10^{-4} mutations/site for GESD, boxplot, medmad and shorth, respectively. Blue and red indicate the position with significant variability in at least one or all four of the tests used, respectively. The U residue at location 1638 of fig. 1.3 was not used in the analysis due to high variability caused by the homopolymer effects during high-throughput 454 sequencing (Huse et al., 2007) ; taken from (Beeharry et al., 2014).

91.9% of transitions and 8.1% were transversions. Therefore, these data support the hypothesis that during HDV replication, more transitions mutations are made in R199G.

3.1.6 Covariation is observed in the highly conserved structure of the right terminal region

Since it was previously suggested that the structure of the promoter is important for the recognition of the HDV promoter by RNAP II, I investigated the covariation in the double-stranded regions (Abrahem and Pelchat, 2008; Greco-Stewart et al., 2006). There are only a few studies pertaining to the analysis of covariation for a large amount of sequences (Seemann et al., 2008; Wan et al., 2011; Westhof, 2015). Therefore, with the help of my laboratory, we developed a heuristics method to analyze covariation. We used a model of the most energetically stable secondary structure to define the base pair positions in order to calculate the covariation for each base pair position. This secondary structure model of R199G is in agreement with the one presented in the report of Beard and colleagues, which was based on *in vitro* nuclease mapping assays and natural mutants (Beard et al., 1996). The frequency and composition of nucleotides for each base pair position of the alignment was analyzed (fig. 3.1.7; table 3.1.3). In order to be able to draw a picture of the covariation per site, only variability higher than 2.1×10^{-4} mutations/base pair (equivalent to a covariation observed 100 read times) was shown on the generated consensus sequence (fig. 3.1.7.A). The majority of variation corresponded to transitions, more specifically to Wobble base pairing G-U or C-A (fig. 3.1.7.A). The C-A base pairs observed could be the results of a mutation to a G-U base pair on the antigenomic strand used as a template for the synthesis of the genomic strand. However, I cannot exclude that the C may be protonated and allowing the base pairing of a C-A Wobble base pair (Jang et al., 1998; Leontis et al., 2002). Overall, the sequence of the double-stranded promoter is highly conserved, with some

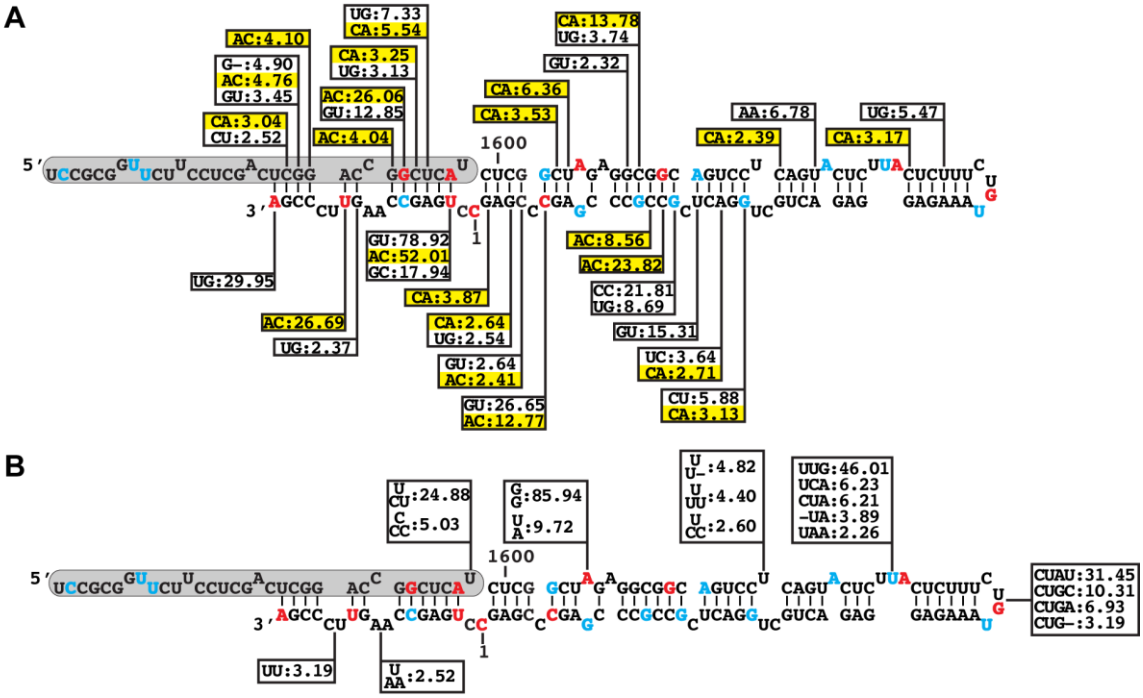

A
AC:4.10
UG:7.33
CA:5.54
CA:13.78
UG:3.74
G-:4.90
AC:4.76
GU:3.45
CA:3.25
UG:3.13
GU:2.32
CA:6.36
AC:26.06
GU:12.85
CA:3.53
AA:6.78
UG:5.47
CA:3.04
CU:2.52
AC:4.04
CA:2.39
CA:3.17
1600
5′ UCCGCG GU UCU U CCUCG A CUCGG AC C GGCUCA U CUCG GCU A G A GGCGGC AGUCC U CAGU A CUC UUA CUCUUU C U
3′ AGCC CU UG AA CCGAGU CC GAGC C CGA G C CCGCCG C UCAGG UC GUCA GAG GAGAAA U G
1
GU:78.92
AC:52.01
GC:17.94
AC:8.56
AC:23.82
UG:29.95
CA:3.87
CC:21.81
UG:8.69
AC:26.69
CA:2.64
UG:2.54
GU:15.31
UG:2.37
GU:2.64
AC:2.41
UC:3.64
CA:2.71
GU:26.65
AC:12.77
CU:5.88
CA:3.13
B
U
CU:24.88
C
CC:5.03
G
G:85.94
U
A:9.72
U
U-:4.82
U
UU:4.40
U
CC:2.60
UUG:46.01
UCA:6.23
CUA:6.21
-UA:3.89
UAA:2.26
1600
5′ UCCGCG GU UCU U CCUCG A CUCGG AC C GGCUCA U CUCG GCU A G A GGCGGC AGUCC U CAGU A CUC UUA CUCUUU C U
3′ AGCC CU UG AA CCGAGU CC GAGC C CGA G C CCGCCG C UCAGG UC GUCA GAG GAGAAA U G
1
CUAU:31.45
CUGC:10.31
CUGA:6.93
CUG-:3.19
UU:3.19
U
AA:2.52

Figure 3.1.7 Covariation analysis of the right terminal domain of genomic HDV RNA obtained from high-throughput sequencing.

The covariation variability of every base pair **(A)** and single-stranded region **(B)** was calculated and displayed on the consensus RNA secondary structure. Blue and red nucleotides indicate the position with significant variability, as determined in fig. 3.1.6. Yellow backgrounds indicates transitions generating C-A on genomic HDV RNA. The grey rectangles represent the 5'-end of HDAg ORF. All numbers correspond to $x10^{-4}$ mutations/site. The grey U residue at location 1638 on fig. 1.1.3 was not used in the analysis due to high variability caused by the homopolymer effects during high-throughput 454 sequencing (Huse S M et al, Genome Bio, 2007 ; taken from from (Beeharry et al., 2014).

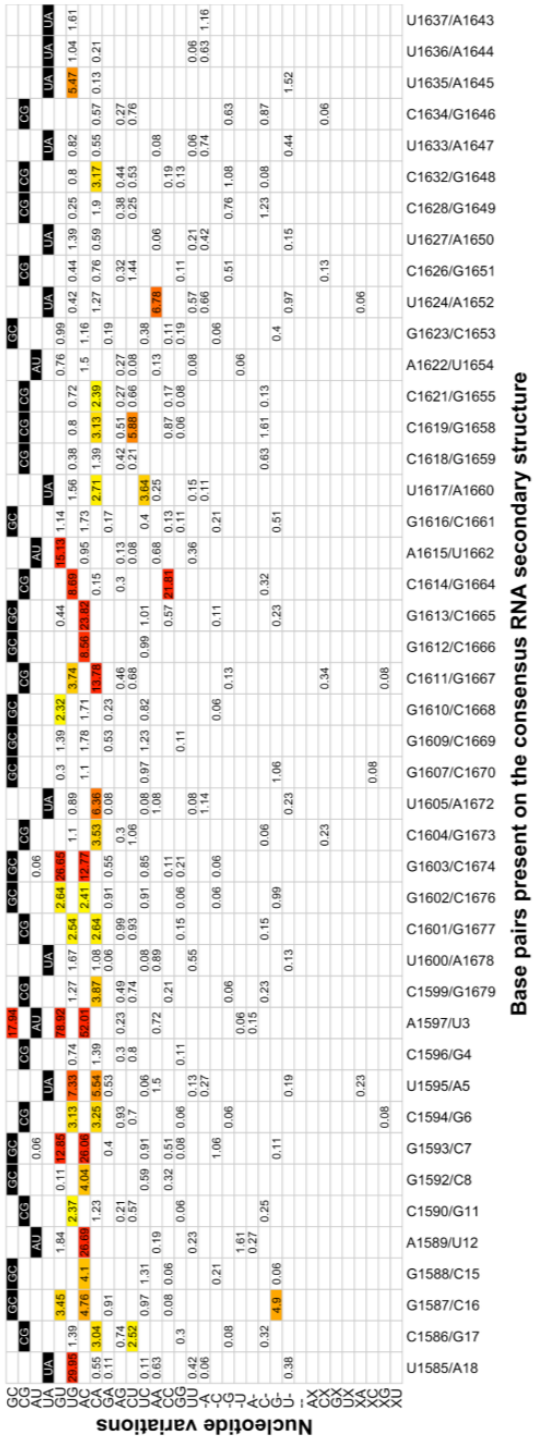


Table 3.3 Analysis of covariation in the right terminal region of an HDV population (taken from Beeharry Y et al, Virology, 2014).

The numbers indicate x10^{-4} mutation/base pair. The black represent the consensus base pairs. The higher mutation rates are highlighted by an incremental gradient of colour from yellow to red.

variability corresponding to Wobble (G-U or C-A), in favor of the conservation of the secondary structure. The mutation of a Wobble to a Watson-crick likely implies a mutation of only one of the nucleotide composing the base pair. For the second nucleotide of HDAg-S initiation codon, I observed a high level of covariation from AU to GU ($78.92X10^{-4}$), AU to AC ($52.01X10^{-4}$) and AU to GC ($17.94X10^{-4}$). This AU > GU > GC suggests a progressive transition with a GU Wobble base pairing on the other strand (C with A on the antigenomic strand). Even if the sequence is modified, the RNA secondary structure is conserved, strongly suggesting the importance of the RNA structure for transcription.

Next, I investigated the possible conservation of the tertiary structure of the RNA of R199G. Often, the tertiary structure is determined by the formation of H-bonds between nucleotides of single stranded regions (hairpins, bulges, etc) of an RNA (Mokdad and Frankel, 2008). A base has three edges: Watson-Crick, Sugar or Hoogsten. If the substitution of one of the nucleotide by another nucleotide still keeps the same distance between the two carbons 1 of the riboses of this base pair, then these two base pairs are of the same isostericity family and should adopt the same conformation. Unfortunately, Isfold could not be used for the analysis of the viral population as this software was incapable of processing such a large amount of sequences (Mokdad and Frankel, 2008). Therefore, using the same method developed to calculate the covariation of the base pairs, I calculated the mutation and composition of the bulged regions since these regions are often important for the establishment of the tertiary structure, with a cut-off of $2.1x10^{-4}$ mutations/site (fig. 3.1.7.B). For the bulge encompassing the position 1630, the proposed site of initiation of transcription, at least one uridine was conserved (fig. 3.1.7.B (Beard et al., 1996; Gudima et al., 1999). The variability rates identified were very low: U_{1599}/C_1C_2 -> U_{1599}/U_1C_2 ($24.88x10^{-4}$ mutations/site; 1,177 reads), A_{1606}/G_{1671} -> G_{1606}/G_{1671} ($85.9x10^{-4}$ mutations/site; 4,050

reads), and $U_{1629}U_{1630}A_{1631}$ -> $U_{1629}U_{1630}G_{1631}$ ($46.01x10^{-4}$ mutations/site; 2,177 reads). Based on these data, there may also be formation of a homopurine between A_{1606} and G_{1671}. However, there was not enough variation in the sequences in order to detect conserved base pairs of the same isosteric family (Leontis et al., 2002).

Chapter 3.2 Use of different HDV variants to study RNA features of the right terminal region

3.2.1 Alignment of the variants and analysis of the secondary structure

In order to characterize the RNA structure of R199G, I took advantage of 100 different sequences from HDV variants available on the Subviral database (Rocheleau, 2006). I analyzed the RNA features conserved among the different HDV isolates, which are subject to various selective pressures since they were isolated from various hosts. Based on the secondary structure previously obtained by in vitro nuclease mapping, a covariance model was established with the Infernal software, and the sequences were aligned sequentially (fig. 3.2.1; Beard et al., 1996; Nawrocki et al., 2009). The consensus composition of the base pairs is indicated in black. The base pairing conservation was quantified by using the "R package for RNA visualization and analysis", R4RNA v.0.1.3, that allows for plotting RNA secondary structures as arc diagrams with a colour code (Lai et al., 2012). The quantifications are indicated in percentage by the arcs at the bottom of the alignment (fig. 3.2.1). The quantification of the covariation is indicated by the arcs at the top. The minimum -2 indicates when there was sequence conservation but no indication on the conservation of the base pair, and +2 indicates positions where the sequence was not conserved, but the base pair was conserved. I observed that the level of covariation varies according to the base pair position for the predicted double-stranded region (fig. 3.2.1). Two patterns of covariation are observed: whether both nucleotides of the base pair were mutated and formed a Watson-Crick base pair (indicated in red on fig. 3.2.1), or whether only one nucleotide of the base pair was mutated and the base pairing (most often a Wobble base pairing) was still preserved (indicated in blue on fig. 3.2.1). The regions with a high level of covariation (blue arcs on fig. 3.2.1) represents 62% of the double-stranded region and these

are likely the regions important for the establishment of the secondary structure (yellow on fig. 3.2.2). This high level of covariation obtained from the analysis of HDV variants indicated that the host and geographic selective pressures tend to preserve the stem structure. Of interest, for the sequences of the clade 3 (bottom of the alignment) which are more divergent compared to the other sequences, there is a very high level of covariation and their structure is preserved. I found high levels of base pair covariation, which are regions where both base paired nucleotides involved in the double-stranded regions are mutated (fig. 3.2.1). Altogether, this analysis revealed which of the R199G positions have a high level of covariation. Covariation indicates evolutionary conserved base pairs that might be required. The regions of high covariation levels likely indicate conserved secondary structures important for the interaction with RNAP II.

3.2.2 Sequence features of the HDV right terminal region

In order to identify the conserved primary structure of HDV promoter for the region encompassing the initiation of transcription, from nucleotides 1592 to 8, I analyzed the composition of the nucleotides at each position of the alignment of 100 HDV variants (fig. 3.2.1). The statistical data of nucleotides composition and conservation was represented as a sliding window of four nucleotides (fig. 3.2.2., upper graph). The conservation of the primary structure was manually quantified: 33 nucleotides were at least 95% conserved, and among them 12 were 100% conserved (blue and red in fig. 3.2.2.). Overall, the primary nucleotide sequence was poorly conserved since only 13% of the nucleotides are conserved at 100% (fig. 3.2.1, fig. 3.2.2). Since I also observed high levels of covariation and that the rod like structure is preserved, this analysis of sequences subject to various selective

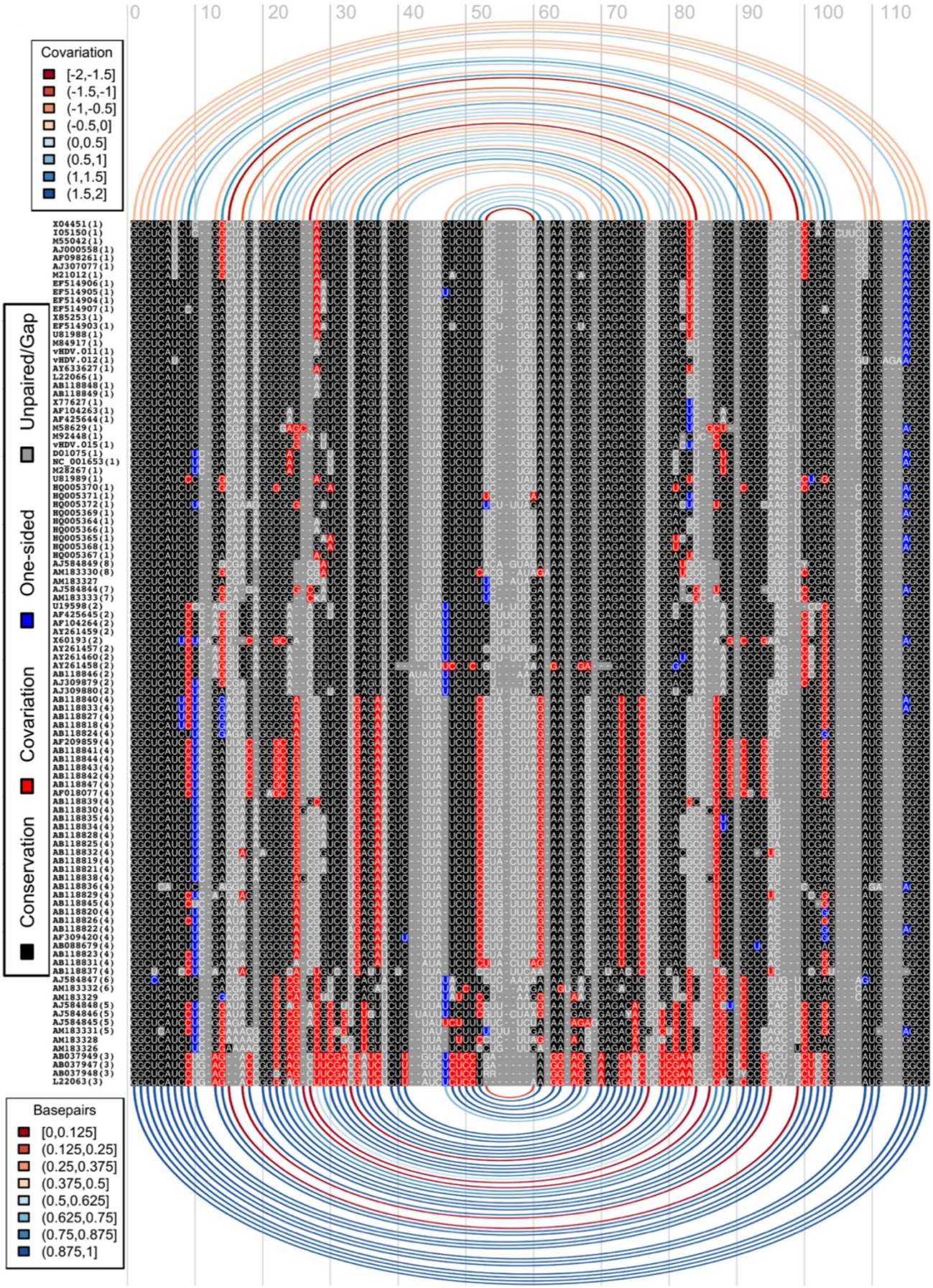
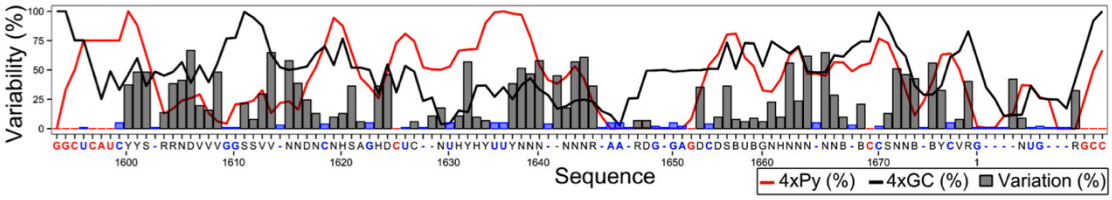

Covariation
[-2,-1.5]
(-1.5,-1]
(-1,-0.5]
(-0.5,0]
(0,0.5]
(0.5,1]
(1,1.5]
(1.5,2]
Unpaired/Gap
One-sided
Covariation
Conservation
Basepairs
[0,0.125]
(0.125,0.25]
(0.25,0.375]
(0.375,0.5]
(0.5,0.625]
(0.625,0.75]
(0.75,0.875]
(0.875,1]

Figure 3.2.1 Structural alignment of HDV variants

For the alignment, the RNA double-stranded region is represented in black, the covariation regions in red and the regions with only one nucleotide of the base pair mutated, but that still maintain the base pair, in blue. The name of each isolate is indicated on the left of the alignment, and the clade of the sequence is indicated between brackets. The base pair conservation was quantified and is indicated in percentage by the arcs at the bottom of the alignment. The covariation quantification is indicated by the arcs at the top. The minimum -2 indicates when there was sequence conservation but no indication on the conservation of the base pair, and +2 indicates positions where the sequence was not conserved, but the base pair was conserved. The covariation regions show a symmetrical pattern.



IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

Figure 3.2.2 Conserved RNA features among 100 HDV isolates.

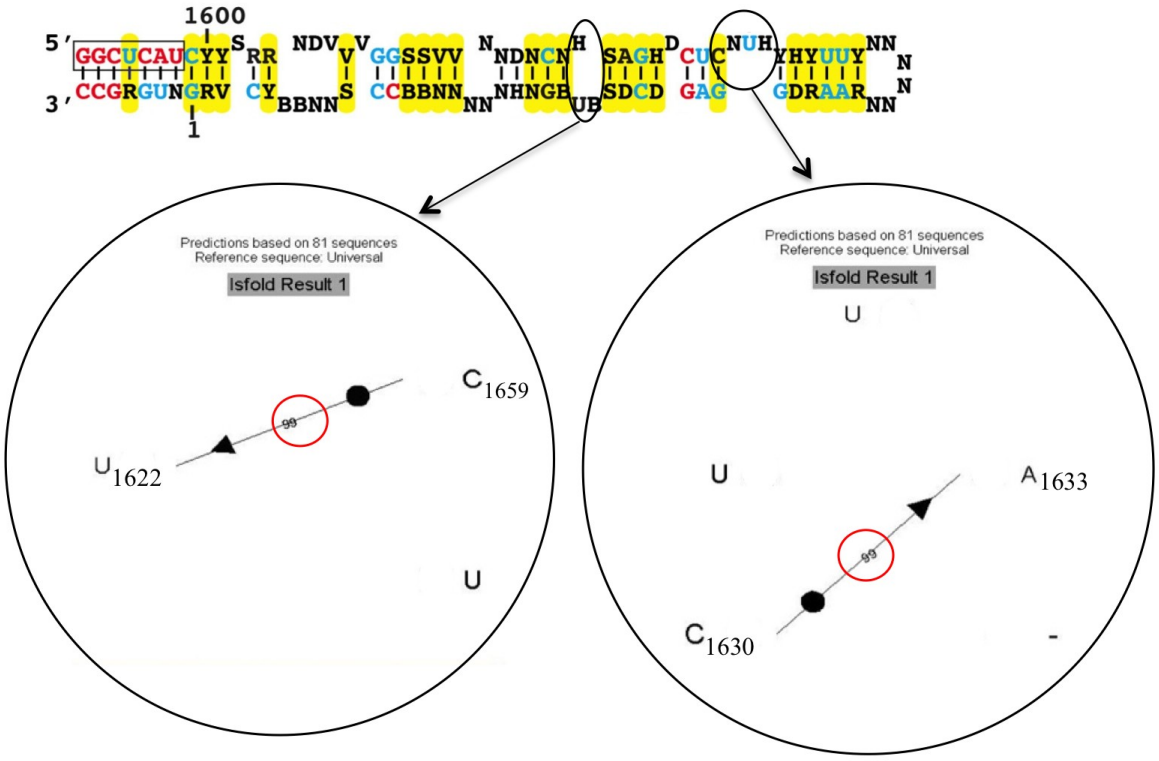
For both graphs, the consensus sequence of the alignment is represented and the nucleotides conserved at 100% and 95% are indicated in red and blue, respectively. The IUPAC abbreviation code used is detailed in the blue table. In the upper graph, the grey bars represent the variability per site and the lines represent the percentage of pyrimidines (red) and the percentage in GC (black) per nucleotide position. The region corresponding to the transcription initiation site (around the position 1630) is enriched in pyrimidines. In the lower graph, the regions with covariation are highlighted in yellow. It appears that the sequence is poorly conserved while there is a high level of covariation, suggesting that it is the secondary structure is conserved.

pressures suggests that the secondary structure is more important than the primary structure. This analysis of the nucleotide conservation also allowed for visualizing that both strands of the stem corresponding to the HDAg-S coding region are conserved (nucleotides 1592 to 1599 and 2 to 8). Therefore since this region is very conserved for both sides of the stem, we cannot deduce whether it is the primary sequence or the secondary structure that is important for the use of R199G as a promoter for transcription.

Nucleotide 1630 was shown to be the position of transcription initiation (Beard et al., 1996; Gudima et al., 1999). This site of initiation of transcription corresponds to an uridine (indicated by an arrow on fig. 3.2.2). This uridine is likely important, and moreover it has been previously reported that the transcription initiation site is often a purine (Grosveld et al., 1981; Hawley and McClure, 1983). I also observed a GC box, which is a region enriched in GC, upstream from the initiation site of transcription, at positions 1610 to 1615 (fig. 3.2.2., indicated on the lower graph and black line on the upper graph). The sequence of the loop was not conserved, since there was more than 60% of variation for this region (positions 1642 to 1646 on the upper graph of fig. 3.2.2), therefore supporting the hypothesis that the sequence conservation of the loop is not a requirement for use by RNAP II. A very conserved CUG/GAC motif was present at positions 1626-1628/1650-1652, very close to the initiation site of transcription, and this motif might be important for the recognition and interaction with RNAP II (fig. 3.2.2, lower graph). In addition, the fact that these nucleotides are conserved even among the most divergent sequences of the clade 3 and that the neighboring nucleotides are not conserved support the functional importance of these few nucleotides (fig. 3.2.1). Another feature that was conserved was a polarization in purine/pyrimidine at the tip of the stem at positions 1626-1638/1646-1652 (fig. 3.2.2). This can also be visualized on the four nucleotides sliding window, for the region of 12 nucleotides around the reported

initiation site (red line in upper graph, fig. 3.2.2.). This purine/pyrimidine polarization can be important for the recognition of RNAP II or of other cofactors involved in transcription (Peng et al., 2002). In summary, this analysis allowed for identifying conserved primary structure motifs conserved in R199G: a region enriched in purine/pyrimidine, a GC box, a conserved uridine at position 1630 and a CUG/GAC motif.

To obtain a model of the tertiary structure of the promoter R199G of HDV RNA, I used the tool Isfold (Mokdad and Frankel, 2008), which detects within an alignment of sequences the mutations that keep base pairs of the same isosterical family. I combined the results output by Isfold to the isostericity matrix and I found 2 positions where the isostericity family of the base pairs were conserved among the variants (fig. 3.2.3; i.e. positions 35-76, 43-46). Therefore, these two bulge regions are two candidates for regions that are important for the establishment of the tertiary structure of R199G.



99% of the sequences adopt a base pair type of the same isostericity family

Figure 3.2.3 Analysis of the RNA bulge regions of R199G from HDV isolates.

The Isfold software was used to analyze the base pair positions of the bulge regions: the nucleotide mutations keeping the same base pairs isostericity family across an alignment of sequences were localized. The black circles show two base pair positions (nucleotides 1622-1659 and 1630-1633) that were found to adopt the same isostericity family bound (trans Watson Crick-Hoogsteen) in 99% of the sequences. The nucleotides in red are conserved at 100% and in blue at 95%. The yellow circles show the covariation of base pairs.

3.3 Summary of findings

Firstly, I analyzed the conserved features of the R199G from sequences isolated from a 293 cellular line where an HDV population was replicating for over a year. The population was heterogeneous and hot spots of variability per nucleotide site were identified. The primary sequence was more conserved at the tip of the rod like structured genome. In addition, my results showed that the sequence of the initiation codon of HDAg-S was mutated for 1.2% of the population, but maintained the base pair. Covariation analysis indicated a strong selection for the rod-like conformation for both polarities of this region. Secondly, I analyzed the conserved features of the R199G sequences among variants isolated from various hosts. At the tip of R199G, several selected RNA motifs were identified, including a GC-rich stem (between nucleotides 1609 to 1614), a CUC/GAG motif, at least one uridine in the bulge containing the initiation site of transcription, and a polarization of purine/pyrimidine content. More importantly, base pair covariation analysis indicates a strong selection for the rod-like conformation for R199G. Altogether, these two studies support the argument that the secondary structure, rather than the primary sequence is important for the promoter region upstream from the tip of R199G.

CHAPTER 4- Link between HDV replication and the paraspeckles

Author Contributions

All experiments were performed by Yasnee Beeharry, with the exception of the following:

Ms Gabrielle Goodrum performed the immunoprecipitation step of the RIP experiments under my supervision and Mr Christian John Imperiale performed the RT-qPCR to quantify NEAT1 and IL8 mRNA under the supervision of Dr Dorota Sikora.

4.1 Involvement of PSF, p54 and PSP1 in HDV replication

4.1.1 PSP1 interacts with the HDV RNA genome

PSF interacts with the HDV RNA via its RRM (Zhang, 2013). PSF often dimerizes with its molecular partner p54. Previously in the laboratory, p54 has also been shown to interact with R199G (Sikora et al., 2009). PSF and p54 are both present in high concentrations in the nuclear paraspeckles. PSP1 is used as a paraspeckle marker (Fox et al., 2005). The proteins PSF, p54 and PSP1 are members of the DHBS family of proteins. One feature common to the three proteins is that they have an RRM. It is therefore tempting to speculate that PSP1 might also interact with HDV RNA genome, as PSF and p54.

In order to identify whether HDV RNA interacts with PSP1, a RIP was performed. For this, I used the cellular system previously developed in the laboratory of J. Taylor, where HDV replication can be induced by TET (Chang et al., 2005; fig. 2.2). As a control, I also grew 293-HDV cells where no TET was added. The cells were incubated for 48 hours after TET induction. In order to cross-link the RNA-protein interactions, all the cells were treated with formaldehyde, following the method previously developed by Niranjanakumari et al. (Niranjanakumari et al., 2002). An immunoprecipitation with the PSP1 antibody was performed. As a positive control, the same experiment for PSF was also performed since PSF binds R199G (Greco-Stewart et al., 2006). As a negative control, an unrelated antibody, β-actin, was used. The samples were heated to reverse the crosslinks and the immunoprecipitates were separated in two fractions: one fraction was used for protein-extraction and the other fraction was used for RNA extraction.

The RNA was extracted and reverse-transcribed using random primers. The cDNA was amplified by PCR and qPCR using primers targeting HDV genome in order to detect the

presence of HDV RNA. Following the RIP with the PSP1 antibody, I observed the presence of a bright band of approximately 250 bp, which is the expected length for the HDV amplicon (fig. 4.1.1). These results suggest that the HDV genome interacts with PSP1 in 293 cells replicating HDV. In cells where no HDV replication was induced, I observed a faint band of approximately 250 pb corresponding to HDV (fig. 4.1.1.A). I quantified the bands of the gel and I observed that, for both RIPs with PSF and PSP1, in 293-HDV cells treated with TET, the levels of HDV genome is 10 times higher than in untreated cells (fig. 4.1.1.B). I did not observe any band for the negative control where the immunoprecipitation was performed with an unrelated antibody (i.e. against β-actin), confirming that the band observed is not due to a non-specific binding of the RNA and the protein G agarose beads or the antibody used for the RIP. In addition, these results were confirmed by qPCR, which is a more sensitive technique: the amplification curve of the RIP with β-actin antibody is within the background signal (almost null), confirming that there is no HDV RNA for this sample (indicated by a pink arrow in fig. 4.1.1.C). The amplification curves of the RIP with PSF and PSP1 for cells treated with TET are similar (curves indicated by black arrows on fig. 4.1.1.C). The melting curve analysis showed a single peak specific of the melting temperature of the HDV amplicon only for the RIP with PSF and PSP1. Altogether, these data show that both PSF and PSP1 interact with HDV genome.

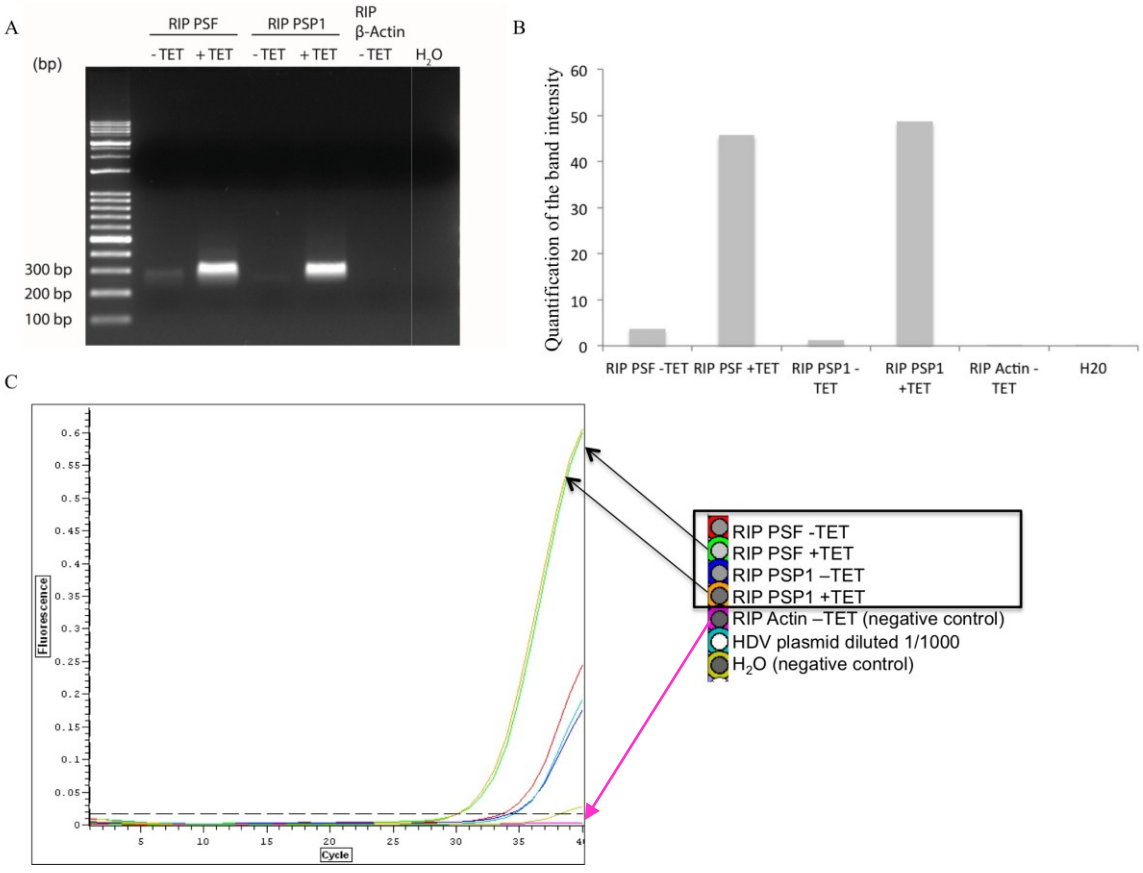

A
RIP PSF
RIP PSP1
RIP β-Actin
- TET + TET - TET + TET - TET H2O
(bp)
300 bp
200 bp
100 bp
B
Quantification of the band intensity
60
50
40
30
20
10
0
RIP PSF -TET
RIP PSF +TET
RIP PSP1 - TET
RIP PSP1 +TET
RIP Actin - TET
H20
C
Fluorescence
Cycle
RIP PSF -TET
RIP PSF +TET
RIP PSP1 –TET
RIP PSP1 +TET
RIP Actin –TET (negative control)
HDV plasmid diluted 1/1000
H2O (negative control)

Figure 4.1.1 Interaction of PSF and PSP1 with HDV RNA

293 cells replicating HDV were treated with formaldehyde. The lysates were used for RIP using PSF and PSP1 antibodies, and the β-actin antibody was used for RIP as a negative control. The isolated RNA was reverse-transcribed with random primers and amplified by PCR. For the PCR and qPCR, primers targeting the HDV ribozyme region were used. (A) The resulting PCR product was resolved on an agarose gel. Water was used as a negative control to assess that the PCR was not contaminated. (B) Quantification of the bands intensities of the agarose gel in (A). (C) Amplification curves showing the amplification of HDV RNA. The two black arrows indicate the samples of the RIP performed with PSF and PSP1 in cells treated with tetracycline. The pink arrow indicates the curve of the negative control, the RIP performed with β-actin.

4.1.2 The knockdown of the paraspeckle proteins PSF, p54 and PSP1 leads to a decrease of HDV accumulation

To clarify whether the paraspeckle markers PSF, p54 and PSP1 are involved in HDV replication, I investigated whether a knockdown of PSF, p54 or PSP1 would affect HDV accumulation. I performed this experiment in 293-HDV cells. Twenty-four hours after plating the cells, when about 60% confluence was reached, I transfected a pool of three siRNAs to knockdown the expression of respectively the proteins PSF, p54 and PSP1 for each cell line. The next day, I added TET to induce HDV replication. As a negative control, I also transfected the same siRNAs without inducing HDV replication. At 48 hours post transfection, I lysed the cells and after centrifugation, I separated the cell lysate in two samples, one for protein extraction and one for RNA extraction. The efficiency of the knock down of the targeted proteins was assessed by Western Blot. For the Western Blot, the protein levels were normalized to β-actin. PSF was knocked down by 80 %, p54 to 88 % and PSP1 by 90 %, compared to the cells transfected with a scramble RNA (fig. 4.1.2).

The total RNA was reverse transcribed using random primers. I measured by qPCR the amount of HDV genome copies following PSF, p54 and PSP1 knockdown, using the *Beta*-2 microglobulin mRNA as an endogenous control. For cells treated with TET, I observed an increase of the number of HDV genomes compared to the uninduced cells (fig. 4.1.2,D). When PSF, p54 or PSP1 were knocked down before the addition of TET, I observed a drastic decrease of the accumulation of HDV RNAs, of more than 90% compared to the cells treated with TET (fig. 4.1.2 D).

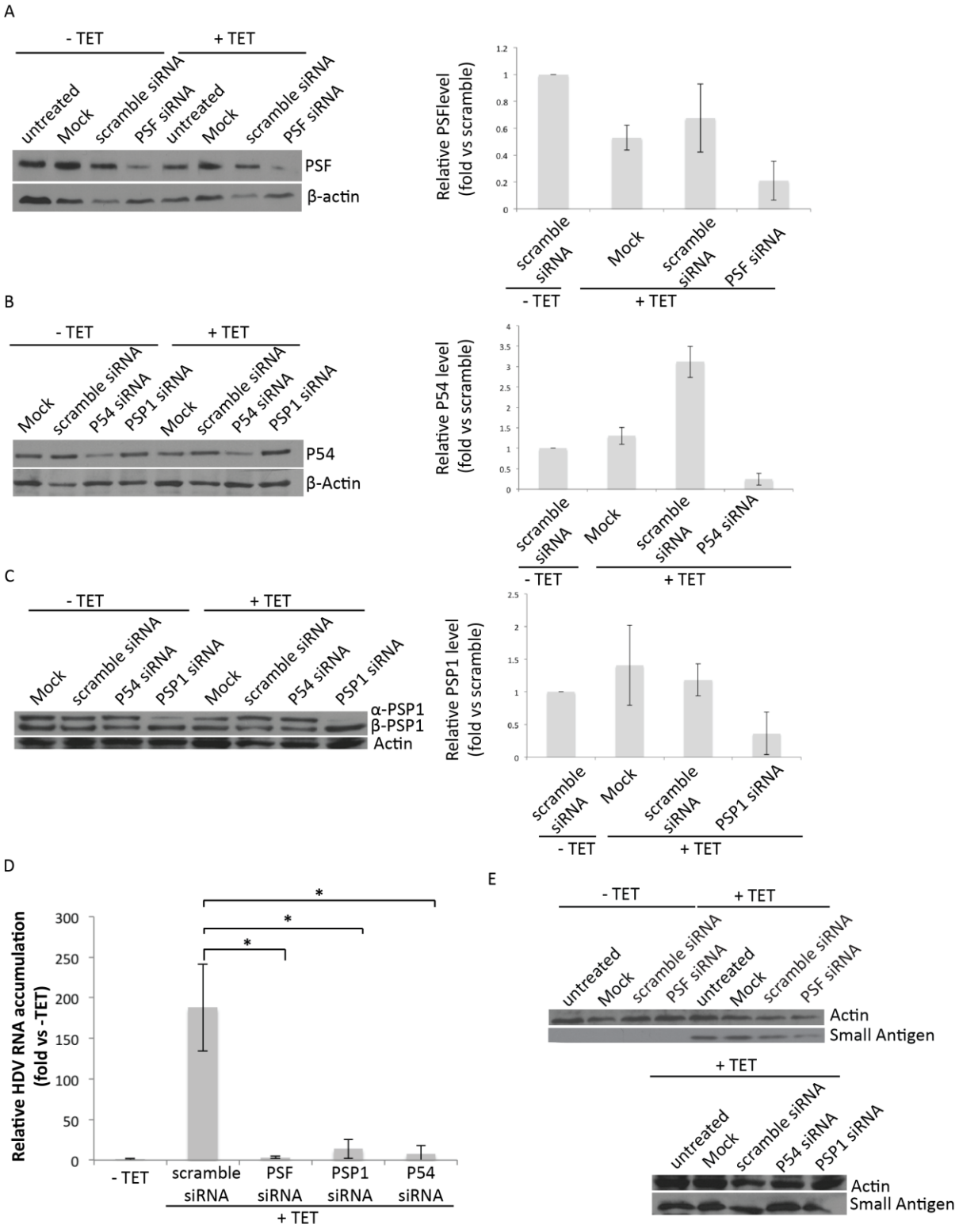

A
- TET
+ TET
untreated
Mock
scramble siRNA
PSF siRNA
untreated
Mock
scramble siRNA
PSF siRNA
PSF
β-actin
Relative PSFlevel (fold vs scramble)
scramble siRNA
Mock
scramble siRNA
PSF siRNA
- TET
+ TET
B
- TET
+ TET
Mock
scramble siRNA
P54 siRNA
PSP1 siRNA
Mock
scramble siRNA
P54 siRNA
PSP1 siRNA
P54
β-Actin
Relative P54 level (fold vs scramble)
scramble siRNA
Mock
scramble siRNA
P54 siRNA
- TET
+ TET
C
- TET
+ TET
Mock
scramble siRNA
P54 siRNA
PSP1 siRNA
Mock
scramble siRNA
P54 siRNA
PSP1 siRNA
α-PSP1
β-PSP1
Actin
Relative PSP1 level (fold vs scramble)
scramble siRNA
Mock
scramble siRNA
PSP1 siRNA
- TET
+ TET
D
Relative HDV RNA accumulation (fold vs -TET)
*
*
*
- TET
scramble siRNA
PSF siRNA
PSP1 siRNA
P54 siRNA
+ TET
E
- TET
+ TET
untreated
Mock
scramble siRNA
PSF siRNA
untreated
Mock
scramble siRNA
PSF siRNA
Actin
Small Antigen
+ TET
untreated
Mock
scramble siRNA
P54 siRNA
PSP1 siRNA
Actin
Small Antigen

Figure 4.1.2 Knockdown of PSF, p54 and PSP1 reduces HDV RNA genome accumulation.

Following the transfection with either siRNAs against PSF mRNA, siRNAs against PSP1 mRNAs and siRNAs against p54 mRNA, a pool of five scrambled siRNAs (Scramble), water (mock) or without transfection (untreated), HDV replication was induced by addition of TET in 293 cells. Western blotting (left) and its quantification (right) showing the relative amount of PSF (A), PSP1 (B) and p54 (C) in each treatment compared to β-actin. (D) Quantification by RT-qPCR of the HDV RNA genome levels normalized to the amount of the house-keeping gene *Beta*-2 microglobulin mRNA and to the untreated cells. Values represent the mean and standard deviation of three biological replicates. Unpaired two-tailed t-tests between the three treatments and the cells induced with TET (+TET) were performed: the asterisk indicates a p-value < 0.05. (E) Western Blotting showing the relative amount of HDAg-S compared to β-actin following the knockdown of PSF, PSP1 and p54.

This effect could be due to reduced levels of HDAg-S, which is required for HDV replication. In order to assess if this change in HDV levels is not caused by a change in HDAg-S levels, I measured the levels of HDAg-S proteins. I performed a Western Blot to detect the levels of HDAg-S in the proteins extract treated with scramble siRNAs, or siRNAs against PSF, PSP1 and p54. For the Western Blot, the proteins levels were normalized to β-actin and I used an antibody against HDAg-S (provided by the laboratory of John M. Taylor) for the HDAg. There was no significant difference in HDAg-S levels in cells transfected with a siRNA targeting these proteins compared to cells transfected with a pool of scramble siRNAs in cells replicating HDV (fig. 4.1.2.E).

In summary, I showed that the knockdown of PSF, p54 and PSP1 lead to a decrease of HDV genomes accumulation in the cell, but does not change the levels of HDAg-S. These results suggest that PSF, p54 and PSP1 are required directly or indirectly for the accumulation of HDV genomes in these cells.

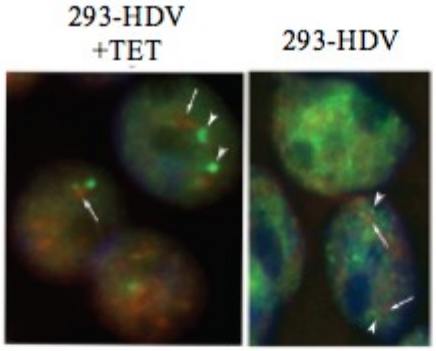
4.2 Effect of HDV replication on the paraspeckles

4.2.1 Large foci of PSP1 outside of the nucleus

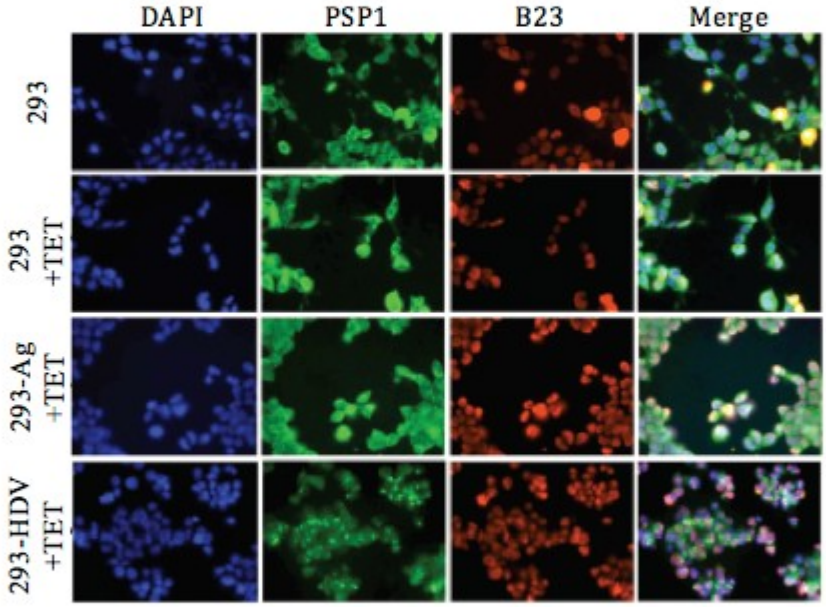
Previous studies indicated that PSP1 is present as small bright foci in the paraspeckle structures, which are located close to the speckle structures (Fox et al., 2002; Shelkovnikova et al., 2014). Since HDV genome interacts with the paraspeckle proteins PSF, p54 and PSP1, and that these proteins are required for the accumulation of HDV RNA, I hypothesized that the paraspeckles may be disrupted upon HDV replication. First, as a preliminary experiment, in order to localize the paraspeckles in the cellular system used, I carried out an immunostaining of a paraspeckle marker and of a speckle marker. I compared the cellular localization of PSP1 with Y12 in cells replicating HDV and control cells that are not

replicating HDV. A day after plating the cells, I added TET to induce HDV replication and for the negative control cells I only changed the media without adding TET. Fourty-eight hours after the induction of HDV replication, I performed an immuno-staining using a polyclonal antibody directed against the C-terminal domain of PSP1 and using an antibody against Y12 (gift of Dr Jocelyn Côté laboratory), which has been previously used as a speckle marker (Y12 Smith antigen; Lamond and Carmo-Fonseca, 1993; Spector and Lamond, 2011; fig. 4.2.1.A). Cells were mounted in Vectashield containing DAPI. In untreated cells, I observed that PSP1 is present everywhere in the cell, but with brighter foci in the nucleus close to Y12, suggesting a localization of PSP1 within paraspeckles, as previously observed (fig. 4.2.1.A; Fox et al., 2002). In cells where HDV replication was induced (293-HDV + tetracycline), PSP1 foci appeared bigger and brighter. Therefore, in cells treated with TET to induce HDV replication, I observed a different pattern of PSP1 localization when compared to 293-HDV cells not treated with TET. This observed pattern of PSP1 foci seemed similar to that previously described upon transcriptional arrest of the cell (Fox et al., 2005). In this last case, PSP1 relocalized to perinucleolar caps (Fox et al., 2005).

A


293-HDV
+TET
293-HDV

B


DAPI
PSP1
B23
Merge
293
293
+TET
293-Ag
+TET
293-HDV
+TET

C

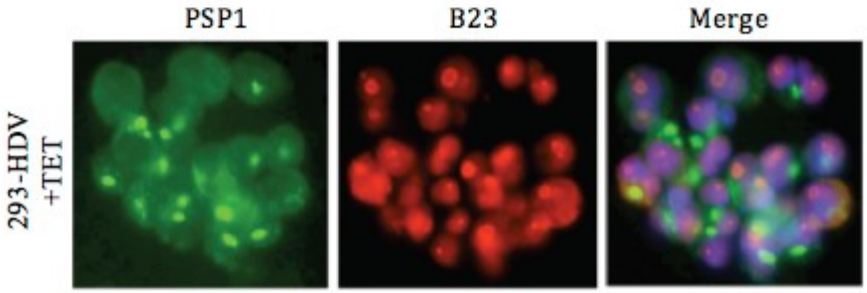

PSP1
B23
Merge
293-HDV
+TET

Figure 4.2.1 The immunostaining pattern of PSP1 is different in 293 cells replicating HDV

(A) 293-HDV cells, treated with TET and untreated, were immunostained with antibodies against PSP1 and Y12, markers of the paraspeckles and speckles respectively. The arrowheads indicate the PSP1 foci and the arrows indicate Y12 foci. In the untreated 293-HDV cells, the PSP1 foci are localized around the Y12 foci. Red: Y12. Green: PSP1. Blue: DAPI. (B) 293 cells, 293 cells treated with TET, 293-Ag cells treated with TET and 293-HDV cells treated with TET were immunostained with DAPI, PSP1 antibody and the nucleolus marker B23 antibody. For all cell lines, no colocalization of PSP1 with B23 is observed in 293 cells replicating HDV. In 293 cells, 293 cells treated with TET and 293-Ag cells treated with TET, the green immunostaining of PSP1 is diffuse. In 293-HDV cells treated with TET, bright green spots outside of the nucleus are observed. Green: PSP1. B23: Red. Blue: DAPI. Merge: Green, Red, Blue (C) Zoomed images of the 293-HDV cells treated with TET in (B).

I next investigated PSP1 relative localization to another nuclear compartment. Since it was previously shown that the HDAg-S relocalized to the nucleolus and that upon transcriptional inhibition PSP1 relocalized to the perinucleolar caps, I stained a nucleolus marker (Lee et al., 1998; Fox et al., 2005). I plated 293 cells, 293-Ag cells and 293-HDV cells. For each cell line, a day after plating the cells, I added TET to induce HDV replication. For each cell line, I also included a negative control without addition of TET. Fourty-eight hours after the induction of HDV replication, I performed an immuno-staining using two antibodies respectively directed against the C-terminal domain of the protein PSP1 and against the nucleolus marker B23 (gift of Dr Jocelyn Côté's laboratory). Cells were mounted in Vectashield containing DAPI. I did not observe a colocalization of PSP1 with the nucleolus marker B23 in cells replicating HDV (fig. 4.2.1.B). However, I observed that PSP1 was present outside of the nucleus as a large bright focus (fig. 4.2.1.C). This pattern was not observed in 293 cells treated with TET or 293-Ag cells treated with TET to induce the production of the HDAg-S. Therefore, the observed delocalization of PSP1 outside of the nucleus is solely due to the accumulation of HDV genomes.

In the cells where HDV replication was induced, I observed a different morphological phenotype: the cells were round shaped compared to the uninduced cells (fig. 4.2.2.A). In addition, the cells started detaching of the culture dish after two days of induction. The team of Dr J. M. Taylor previously reported that in this cellular system, most cells replicating HDV were arrested in G1/G0. Given the phenotype and the cellular cycle arrest previously reported, it is very likely that HDV replication might cause a replication stress in 293 cells. Therefore, I next hypothesized that such a stress might induce the relocalization of PSP1 to

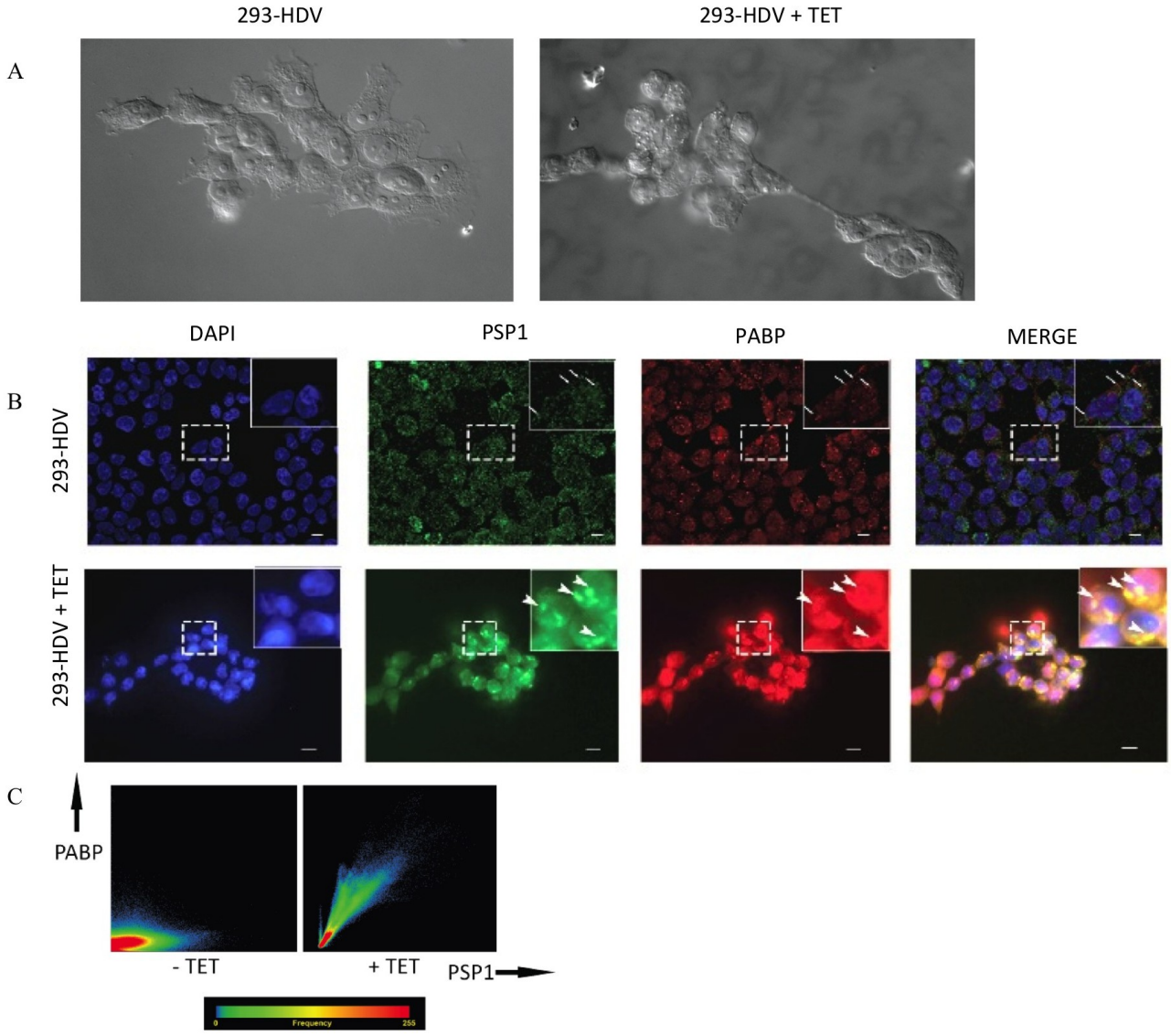

293-HDV
293-HDV + TET
A
DAPI
PSP1
PABP
MERGE
B
293-HDV
293-HDV + TET
C
PABP
- TET
+ TET
PSP1
0
Frequency
255

Figure 4.2.2 Colocalization of PSP1 with PABP

(A) Aspect of the 293 cells replicating HDV observed under a microscope bright light. The 293-HDV cells treated with TET had a more spherical cell shape than the untreated cells. (B) 293-HDV cells untreated and treated with TET were immunostained with antibodies against PSP1 and PABP. DAPI: Blue; PSP1: Green; PABP: Red; Merge: Green, Red, Blue. The squares in the top right corner shows zoomed pictures of the dotted areas. The green and red dots colocalize strongly in 293-HDV cells treated with TET. The arrows indicate the localization of the small foci of PSP1. The arrowheads show the localization of large foci of PSP1 outside of the nucleus. The scale bar on the bottom right corner corresponds to 0.01 μm. The scatter plot on the right is a quantification of the colocalization of the PABP marker (y axis) with the PSP1 marker (x axis).

the cytoplasmic stress granules. I immunostained PSP1 and the PABP, a stress granule marker (Buchan and Parker, 2009). Both PABP and PSP1 observed signals are very strong and highly overlapping in cells where HDV replication was induced. This colocalization was quantified (Pearson coefficient of 0.94 and Manders coefficient of 0.97). In contrast, I did not observe this effect in non-induced cells (fig. 4.2.2.B and C). This suggests that PSP1 is present as large foci outside of the nucleus, probably relocalized in stress granules, during HDV replication in the cellular system used.

4.2.2 PSP1 colocalization with the paraspeckle markers PSF, p54 and NEAT1 is decreased

The paraspeckle marker PSP1 is present as bright foci outside of the paraspeckles in cells replicating HDV (fig. 4.2.1, fig. 4.2.2). In order to determine whether paraspeckles are disrupted in cells replicating HDV, I immunostained PSF and p54 which are two proteins present in paraspeckles and involved in HDV replication cycle. I examined whether PSP1 still colocalized with PSF and p54. A day after plating 293 cells, 293-Ag cells and 293-HDV cells, I added TET to induce HDV replication. I also plated a negative control for each cell line, without addition of TET. Fourty-eight hours after the induction of HDV replication, I performed an immunostaining using a polyclonal antibody directed against the C-terminal domain of the protein PSP1 and using an antibody against PSF or p54. Cells were mounted in Vectashield containing DAPI. Twenty-five cells of a representative field of cells replicating HDV were analyzed, with at least 25 Z stack to capture the signal through the cellular volume. PSF, p54 and PSP1 are at different localizations in the nucleus, and therefore the fluorescent labeling allows visualizing these various locations. However, they colocalize as brighter small foci in the nucleus, which correspond to paraspeckles. In this case, automatic detection of the colocalization of PSP1 with p54 and PSF by a software was not appropriate

since these proteins are present in other nuclear locations than paraspeckles, which would result in background signal. I observed that PSF still remains in the nucleus. Since PSP1 is present outside of the nucleus, there are less foci of PSP1 that colocalize with PSF for the cells where HDV replication was induced (fig. 4.2.3.A and B). I did not observe this decrease of colocalization of PSF with PSP1 for the negative controls, 293 cells, 293 cells treated with TET or 293-Ag cells induced with TET for the production of the HDAg-S. I observed the same results for the colocalization of PSP1 with p54 (fig. 4.2.3.C). Therefore, these experiments show that small bright foci of PSF and p54 are still present in the nucleus upon HDV replication in these cells and are supportive of the hypothesis that PSF and p54 are still present in high concentrations in the paraspeckles.

Previously, the team of J. B. Lawrence showed that the long non coding RNA NEAT1 was essential for the formation of paraspeckles (Clemson et al., 2009). Since NEAT1 is required for the formation of paraspeckles, I further investigated if the paraspeckles were disrupted upon HDV replication by investigating whether NEAT1 remained in the nucleus and whether the relative localization of NEAT1 and PSP1 changes in cells replicating HDV. A day after plating 293-HDV cells, I added TET to induce HDV replication. I also plated a negative control, where no TET was added. After 48 hours, I measured PSP1 and NEAT1 localization respectively by immunostaining, as previously described, and RNA-FISH. The nucleus was stained by DAPI. The immunofluorescence pictures shown are the combination of at least 25 Z stacks, to capture signal in the whole volume of the cell. I observed that most of the foci corresponding to the staining of PSP1 are present outside of the nucleus, which is in accordance with my previous results (fig. 4.2.4.A).

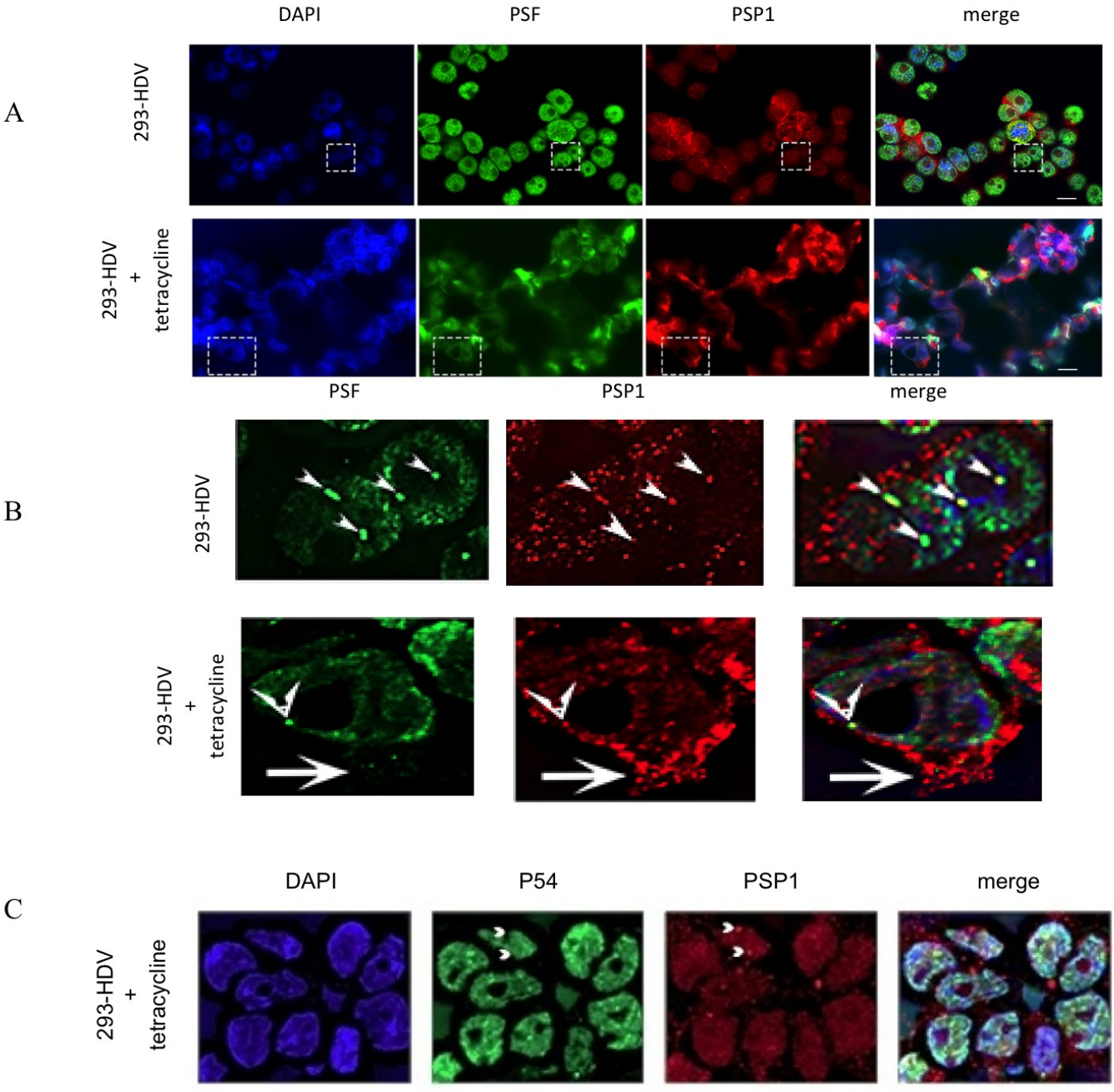

A
DAPI
PSF
PSP1
merge
293-HDV
293-HDV + tetracycline
B
PSF
PSP1
merge
293-HDV
293-HDV + tetracycline
C
DAPI
P54
PSP1
merge
293-HDV + tetracycline

Figure 4.2.3 Decrease of colocalization of PSF and p54 with PSP1 upon HDV replication in 293 cells.

(A) 293-HDV cells untreated and treated with TET were immunostained using antibodies directed against PSP1 and PSF. Blue: DAPI. Green: PSF. Red : PSP1. The scale bar on the right bottom of the merge pictures corresponds to 0.01 μm. (B) Zoomed pictures of the dotted areas in (A). In untreated cells, the green and red stainings colocalize as small bright dots (arrowheads). In cells treated with TET, the majority of the red dots does not colocalize with green (arrowheads), but the majority of the bright red dots localize outside of the nucleus (arrows). Each picture corresponds to the merge of pictures of at least 25 Z-stacks. The same pattern as 293-HDV cells without TET was observed for 293 and 293-Ag cells (data not shown). The experiments were repeated in three independent experiments for 293-HDV cells with or without TET induction. (C) 293-HDV cells treated with TET was immunostained by using antibodies directed against PSF and p54. Blue: DAPI. Green: p54. Red: PSP1. In cells treated with TET, the majority of the red dots does not colocalize with green (arrows), but the majority of the bright red dots localize outside of the nucleus. Each picture corresponds to the merge of pictures of at least 25 Z-stacks.

However, the stain corresponding to NEAT1 still appears as several bright foci in the nucleus, as previously described for the paraspeckle structure (fig. 4.2.4.A). This result confirms that PSP1 is present outside of the paraspeckle structure and moreover shows that NEAT1 speckles remain in the nucleus. These results also suggest that the paraspeckles are not disrupted upon HDV replication. Since PSP1 is present in the cytoplasm while NEAT1 remains in the nucleus, I hypothesized that I would also observe a decrease of interaction of PSP1 with NEAT1 upon HDV replication in 293 cells.

In order to further explore whether there is a decrease of NEAT1 levels physically interacting with PSP1 upon HDV replication, I performed a RIP. My previous immunofluorescence experiments indicate that PSF remains in the nucleus, therefore I expect that PSF will still interact with NEAT1, which is an essential component of the paraspeckles (Sasaki and Hirose, 2009). In order to further investigate whether PSF still interacts with NEAT1 in cells replicating HDV, I also quantified the change of NEAT1 levels physically interacting with PSF by performing this RIP.

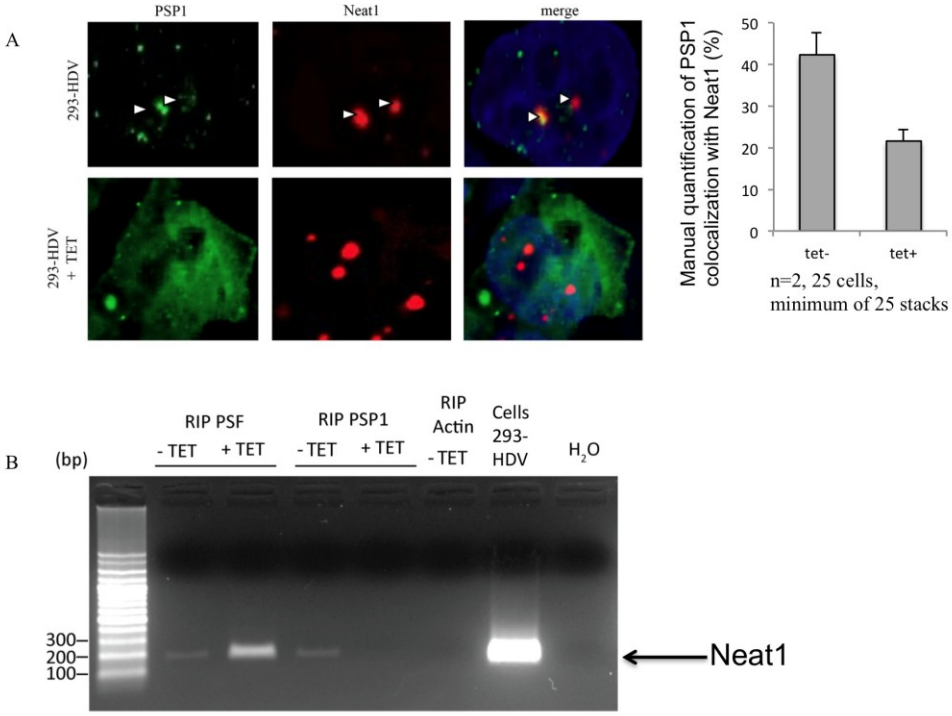

A
PSP1
Neat1
merge
293-HDV
293-HDV + TET
Manual quantification of PSP1 colocalization with Neat1 (%)
50
40
30
20
10
0
tet-
tet+
n=2, 25 cells, minimum of 25 stacks
B
(bp)
RIP PSF
- TET
+ TET
RIP PSP1
- TET
+ TET
RIP Actin
- TET
Cells 293-HDV
H2O
300
200
100
Neat1

Figure 4.2.4 Disruption of NEAT1 interaction with PSP1 and enlarged NEAT1 foci

(A) Decrease of colocalization of NEAT1 with PSP1 upon HDV replication in 293 cells. Left: The NEAT1 foci remained in the nucleus in 293 cells replicating HDV, while large foci of PSP1 are observed outside of the nucleus in cells replicating HDV. Right: The number of PSP1 foci and NEAT1 foci colocalizing was manually quantified and the results suggest a decrease of PSP1 colocalization with NEAT1 in the in 293 cells replicating HDV. Average on 25 cells, n=2 independent experiments.
(B) Decrease of interaction of NEAT1 with PSP1 upon HDV replication in 293 cells. 293 cells replicating HDV were treated with formaldehyde. The lysates were used for RIP using PSF and PSP1 antibodies, and the β actin antibody was used as a negative control. The isolated RNA was reverse-transcribed with random primers and amplified by PCR with primers targeting the NEAT1_2 gene. The resulting PCR product was resolved on an agarose gel. An aliquot of the 293-HDV cell lysates was used as a positive control, to check that the size of the NEAT1_2 amplicon was of the right size. Water was used as a negative control to assess that the PCR was not contaminated.

Twenty-four hours after growing 293 HDV cells, I added TET to induce HDV replication. I also grew control cells where no TET was added. After seventy-two hours, all the cells were treated with formaldehyde in order to cross-link the RNA-proteins conformations. I performed immunoprecipitations using antibodies directed against PSF or PSP1. I controlled for non-specific interactions with NEAT1 by using an anti β-actin antibody. The samples were heated to decrosslink RNA-protein complexes and the immunoprecipitates were separated in two fractions: one was used for protein-extraction and the other for RNA extraction. The RNA was extracted and reverse-transcribed using random primers. I semi-quantified the levels of NEAT1 by RT-PCR. In cells where no HDV replication was induced and the immunoprecipitation performed using the PSP1 or PSF antibodies, I observed the same faint band of 200 base pairs corresponding to the size of NEAT1 amplicon (fig. 4.2.4.B). This result was expected since NEAT1 interacts with PSF and PSP1 in the paraspeckles. Following the RIP with the PSP1 antibody in cells replicating HDV, I observed no band. However, upon over-exposition of the agarose gel, I observed a very faint band of 200 base pairs, which corresponds to the size of NEAT1 amplicon. For the RIP with cells where HDV replication was induced and the immunoprecipitation performed using the PSF antibody, I observed a bright band of approximately 200 base pairs, corresponding to NEAT1 amplicon (fig. 4.2.4.B). This band is brighter than the band of the lane corresponding to the immunoprecipitation with PSF or PSP1 in cells with no addition of tetracycline. I did not observe this band with the negative control antibody (i.e. against β-actin), thus confirming that the band observed is not due to a non-specific binding of NEAT1 RNA and the protein G agarose beads or the antibody used for the RIP (fig. 4.2.4.B). These results indicate that upon HDV replication, NEAT1 interaction with PSP1 is decreased, which is in accordance with the immuno-staining results showing that PSP1 is present outside of the nuclear paraspeckles

upon HDV replication. These RIP results also indicate that there is an increase of levels of NEAT1 interacting with PSF. This could be due to an increased ability of NEAT1 to interact with PSF or to higher levels of NEAT1 transcripts.

4.2.3 Change in expression levels of NEAT1 and IL8 upon HDV replication in 293 cells

It was previously reported that in cells replicating Influenza Virus A, NEAT1 expression was upregulated as a consequence of the recruitment of PSF to paraspeckles (Imamura et al., 2014). Therefore I tested whether there was an increase of NEAT1 expression in cells replicating HDV. A day after plating 293-HDV cells, I added TET to induce HDV replication. I also plated a negative control, where no TET was added. Fourty-eight hours after the induction of HDV replication, I measured NEAT1 localization by RNA FISH and DAPI staining. My analysis indicates that the number of NEAT1 foci does not vary in 293 cells transfected with HDV, with or without TET induction (fig. 4.2.5.A and B). By using FISH to detect NEAT1_2 transcripts, Imamura and colleagues showed that there was an enlargement of NEAT1 foci upon IAV replication (Imamura et al., 2014). Therefore, I investigated whether HDV replication induced an increase of the size and the intensity of NEAT1 foci in 293 cells. Quantification of the area of NEAT1 foci suggests a significant increase of size in cells replicating HDV (fig. 4.2.5.A and C). Quantification of the grey value of NEAT1 foci suggests an increase of the intensity of NEAT1 foci in cells replicating HDV (Student t-test, p-value < 0.05 ; fig. 4.2.5.A and D). In summary, while the number of NEAT1 foci per cell remained the same, the size and intensity of NEAT1 foci is increased in cells where HDV replication was induced (fig. 4.2.5). These data would support the hypothesis that the number of NEAT1 transcripts present in the paraspeckles is increased during HDV replication.

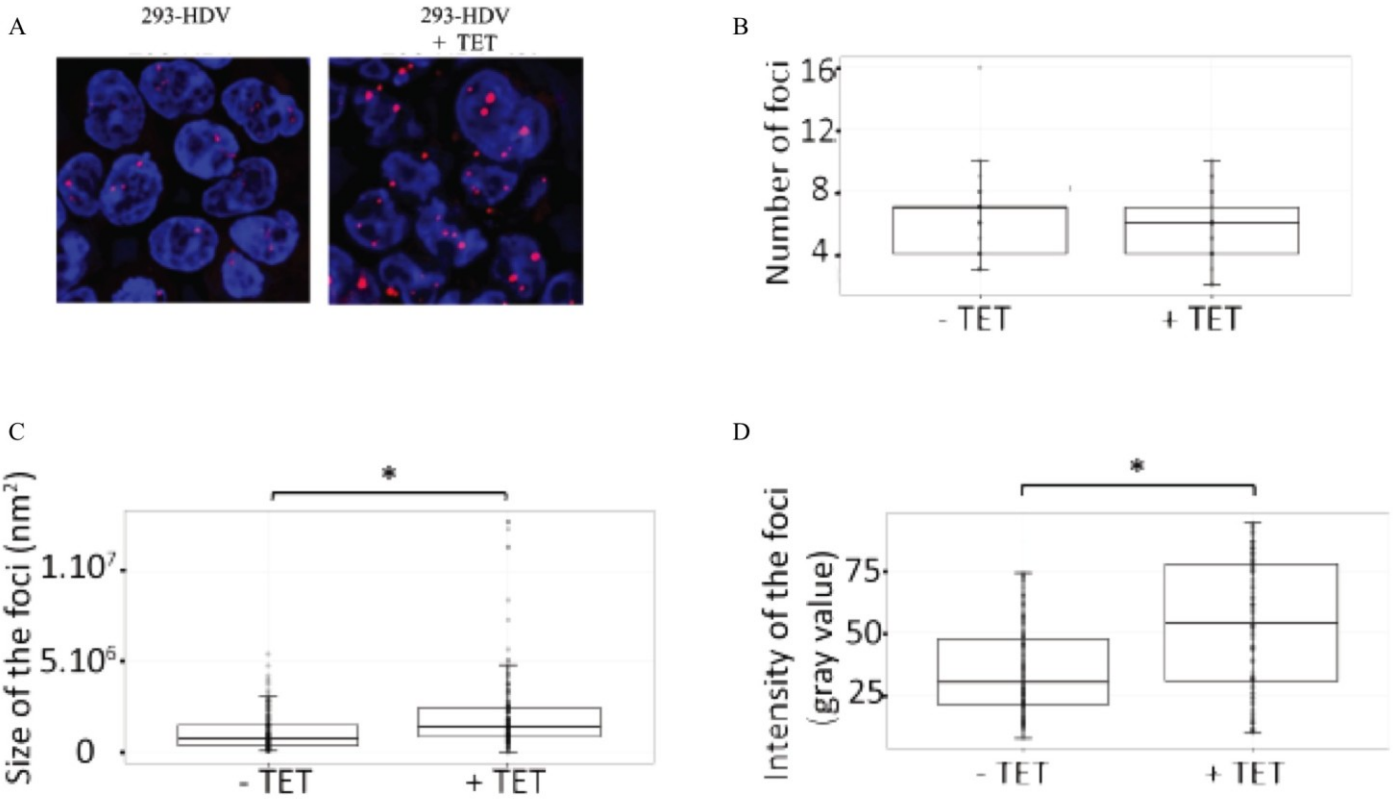

A
293-HDV
293-HDV
+ TET
B
Number of foci
16
12
8
4
- TET
+ TET
C
Size of the foci (nm²)
1.10⁷
5.10⁶
0
*
- TET
+ TET
D
Intensity of the foci
(gray value)
75
50
25
*
- TET
+ TET

Figure 4.2.5 The size and intensity of the NEAT1 foci is increased in cells replicating HDV.

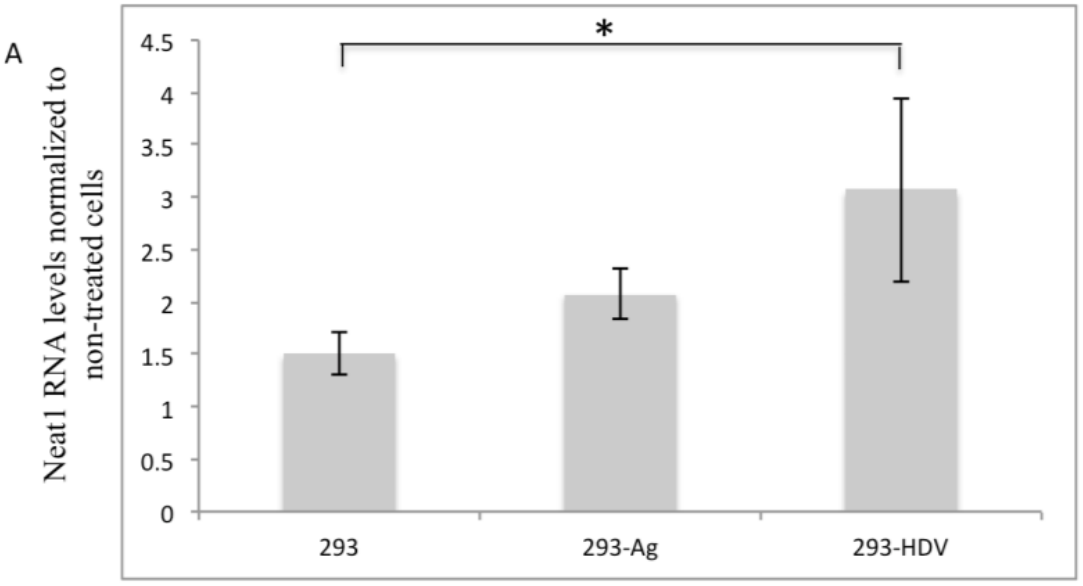
(A) NEAT1 RNA was detected by FISH in 293-HDV cells untreated and treated with TET (+ TET) and the nucleus was stained by DAPI. (B) A quantification of the immunofluorescence data shown in (A) shows that the number of foci of NEAT1 is similar in 293 cells untreated and treated with TET (C) Quantification of the area of NEAT1 foci indicates an increase of size in cells treated with TET. Average on 25 cells, n=3 independent experiments. (D) Quantification of the grey value of NEAT1 foci indicates an increase of the intensity of NEAT1 foci in cells treated with TET. Average on 25 cells, n=3 independent experiments. Unpaired two-tailed t-tests between the untreated cells and the cells induced with TET were performed: the asterisk indicates a p-value < 0.05. For (B), (C) and (D) values were obtained from three biological replicates.

To assess NEAT1 upregulation or increased accumulation, we quantified the amount of NEAT1 transcripts by RT-qPCR in cells replicating HDV compared to 293 cells and 293 cells producing only the HDAg-S. We used a set of primers targeting the human NEAT1_2 gene and a set of primers targeting the house-keeping mRNA *Beta*-2 microglobulin. For the cells treated with TET, I observed an increase of approximately 2 folds of NEAT1_2 RNA levels when compared to 293 cells (fig. 4.2.6.A). These RT-qPCR results are consistent with the previous results showing that there is an increase of approximately 2 folds of the size and intensity of NEAT1 foci (fig. 4.2.5.C and D). Since the team of K. Imamura found that the IAV replication also led to an increase of the cytokine IL8 levels, we also performed a qPCR to measure IL8 mRNA. Similarly to the results found for NEAT1_2, we found an increase of 2 fold of the IL8 transcripts (fig. 4.2.6.B).

4.3 Summary of main findings

In a cellular system allowing HDV replication, the paraspeckle proteins PSF, PSP1 and p54 are required for the accumulation of HDV RNA. Similarly to PSF and p54, the protein PSP1 associates with HDV RNA. My immunostaining and RIP experiments indicate that in 293 cells replicating HDV, the paraspeckle marker PSP1 is present outside of the nucleus and that there is less colocalization of PSP1 with PSF, p54 and NEAT1. There is also an increase of NEAT1 transcripts in the paraspeckles. Altogether, these data support the hypothesis that the host DBHS proteins are required for HDV replication and that HDV replication induces a delocalization of a fraction of the nuclear PSP1 outside of the paraspeckles to the cytoplasm.

A


*
Neat1 RNA levels normalized to non-treated cells
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
293
293-Ag
293-HDV

B

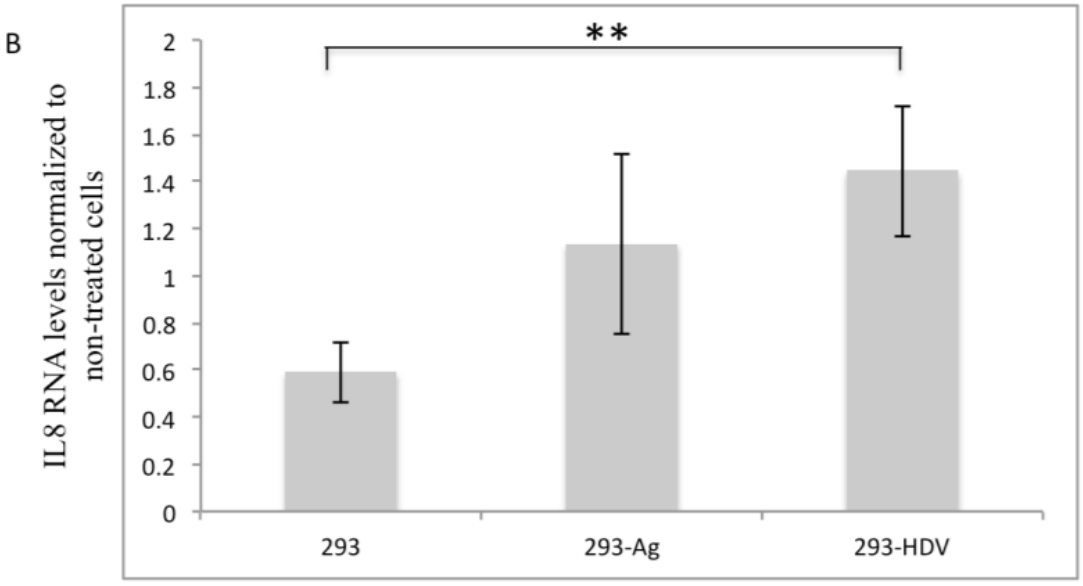

**
IL8 RNA levels normalized to non-treated cells
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
293
293-Ag
293-HDV

Figure 4.2.6 Increase of NEAT1 and IL8 levels upon HDV replication in 293 cells

293 cells, 293-Ag cells and 293-HDV cells were grown for 24 h and induced for HDV replication by addition of TET. After 48 h, the total RNA was extracted, reverse-transcribed with random primers and the cDNA was amplified by qPCR with primers directed against (A) the NEAT1-2 human gene, (B) the IL8 human gene. The *delta* Ct was calculated by normalizing the respective gene levels to *Beta*-2 microglobulin and the *delta delta* Ct was calculated by normalizing the gene expression of each cell line to the expression of non treated corresponding cell line. Values represent the mean and standard deviation of three biological replicates (*$p<0.05$, **$p<0.01$; two-tailed Student's t-test).

CHAPTER 5: DISCUSSION

5.1 Summary of findings

Previous studies have identified R199G as an RNAP II promoter able to initiate transcription, but the structural features required for HDV RNA to be used as an RNA promoter were unknown. In order to evaluate the heterogeneity of this region in a single host system, I generated 473,139 sequences representing 2,351 new R199G variants by high-throughput sequencing of a viral population replicating in 293 cells. To complement this analysis, I analyzed the conserved features of the R199G among variants isolated from various hosts. Base pair covariation analysis on these two datasets indicates a strong selection for the rod-like conformation. Altogether, these findings support the argument that the secondary structure, rather than the primary sequence is important for the region upstream from the tip of R199G. Several RNA motifs were identified, including a GC-rich stem, a CUC/GAG motif, a uridine at the initiation site of transcription, and a polarization of purine/pyrimidine content. The purine enrichment identified is reminiscent of RNA features that can be bound by proteins.

Previously it was shown that the same R199G mutants able to bind RNAP II were able to bind PSF, and that the addition of PSF resulted in an increase of the HDV RNA products of *in vitro* transcription. Previous work also identified that R199G associates with p54, a protein often forming complexes with PSF. The DBHS family of proteins is composed of three mammalian members: PSF, p54 and PSP1. I showed that PSP1 also associates with HDV RNA and that there is a drastic decrease of HDV RNA levels upon knock down of PSF, p54 and PSP1. These nuclear proteins are highly concentrated in paraspeckles, structures involved in storage of transcripts generated by RNAP II. NEAT1 is a long non-coding RNA essential for the formation of paraspeckles. I observed that upon viral replication in 293 cells, the PSP1 foci appear bigger, that the colocalization of PSP1 with

PSF and p54 decreased, that large bright focus of PSP1 are present in the cytoplasm and no longer colocalizes with NEAT1 in the nucleus. I also found that upon HDV replication, there was less NEAT1 that associates with PSP1 and that NEAT1 foci were enlarged. This finding is consistent with data that showed an increase of NEAT1 transcripts in cells replicating HDV. Altogether, these data provide evidence that the paraspeckles are altered upon HDV replication in 293 cells.

5.2 Conserved primary and secondary structural features of an HDV promoter for RNAP II

5.2.1 Model of the conserved features of HDV right terminal region

I took advantage of high-throughput sequencing technologies to extract information from thousands of sequences from a whole viral population actively replicating. We are looking at a population where most sequences are able to ensure replication. All the sequences are subject to the same selective pressure, as we also wanted to test if the same features would be conserved. In addition, I analyzed 100 different HDV variants subject to a high selective pressure. Natural selection allows for identifying essential RNA structural features. This provides *in vivo* information since we used isolates from different geographical regions and from various hosts.

One important finding is that the level of covariation depends on the nucleotide positions, suggesting some positions are more important for the establishment of the secondary structure and the recognition by RNAP II (Chapter 3, fig. 3.1.7; fig. 3.2.1; fig. 3.2.2.B). At the tip of the rod, the sequence was more conserved and upstream from the tip, the high level of base pairs covariation strongly suggests the importance of the secondary structure (fig. 5.1). My model is in accordance with the features established in previously published data from site directed mutagenesis, immunoprecipitations and *in vitro*

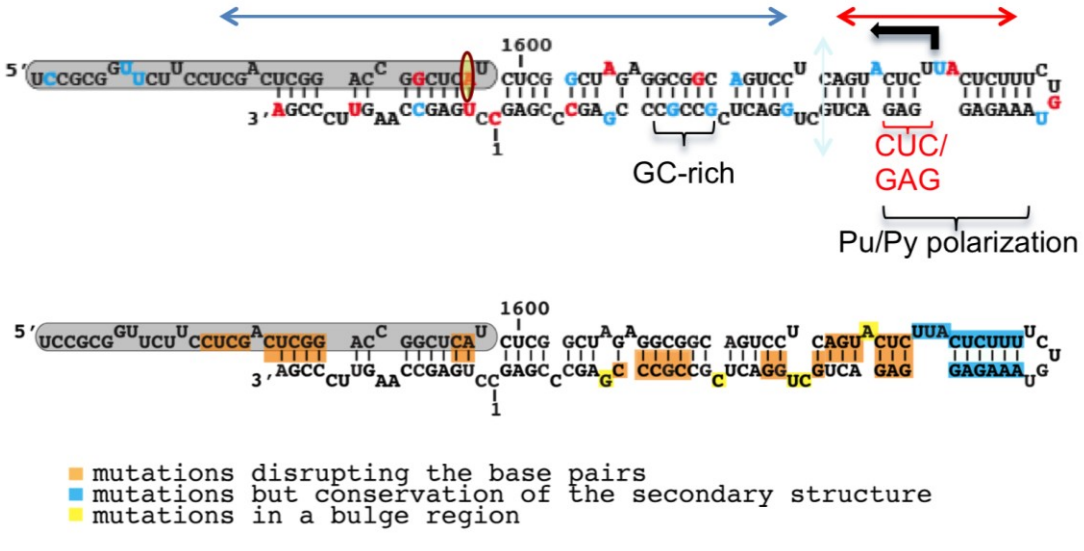

1600
GC-rich
CUC/
GAG
Pu/Py polarization
1600
mutations disrupting the base pairs
mutations but conservation of the secondary structure
mutations in a bulge region

Figure 5.1 Model of R199G conserved features derived from the analysis of the sequences of HDV isolates and of a whole viral population replicating.

(Upper figure) Model of R199G conserved features derived from the analysis of a whole HDV population replication (Chapter 3.1) and from the analysis of 100 HDV variants (Chapter 3.2). The region at the tip (region located on the right side of the pale blue vertical arrow) has a conserved sequence (indicated by a red horizontal arrow). The base-pair covariation identified for the region upstream from the tip suggests that this region has a conserved secondary structure (indicated by a blue horizontal arrow). The second nucleotide of the HDAg initiation codon is also the position with the highest variability (red circle filled in yellow). Other conserved motifs were identified: a GC-rich stem, a CUC/GAG motif, a uridine at the initiation site of transcription, and a polarization of purine/pyrimidine content.

(Lower figure) Model derived from a review of the literature of the R199G features required for the interaction with RNAP II or the accumulation of HDV RNAs in cells (presented in Chapter 1). For the region at the tip of the stem (region between positions 1632 to 1648), mutations that changes the sequence (blue), but allows for preserving the secondary structure resulted in a decrease of interaction of HDV with RNAP (Greco-Stewart, 2009) or lower levels of HDV RNA accumulated in cells (Wu et al., 1997). An alteration of the bulge "UUA" (blue; positions 1629 to 1631) located at the tip of the sequence also resulted in a decrease of HDV RNA accumulated in cells (Gudima et al., 1999). Sequence deletions or mutations of the bulges (yellow) and/or the base pairs (orange) in the regions located between positions 1585-1628 and 1649-18 results in a decrease of HDV RNA accumulated in cells (Beard et al., 1996; Liao et al., 2012; Wang et al., 1997).

transcription assays (fig. 5.1; Appendix I; Beard et al., 1996; Greco-Stewart et al., 2007; Macnaughton and Lai, 1993; Wu et al., 1997).

Importantly, at position 1597 and 3, I identified a progressive covariation of AU to GU/AC and then to GC, which still allowed a base pairing with the opposite strand in this double-stranded region. This high level of covariation in the start codon of HDAg suggests the importance of the secondary structure of the promoter rather than its primary structure. My results are in agreement with the data of two other teams (Beeharry et al., 2014; Griffin et al., 2014; Liao et al., 2012). One team studied whether the initiation codon was important for the establishment of the secondary structure and they reported that a mutation disrupting the base pairing of the nucleotides 1670 to 1674 resulted in an abrogation of the accumulation of HDV genomes (Liao et al., 2012). The other team performed a Selective 2'OH acylation analyzed by primer extension (SHAPE) and mutational assays to alter the bulges and internal loops. Their data suggested that the binding of HDAg-S, which is involved in replication, does not alter the stem-loop structure of HDV, and that the secondary structure of the RNA, rather than its primary sequence, determines HDAg binding specificity (Griffin et al., 2014; Yamaguchi et al., 2007).

5.2.2 Comparison between the features of R199G and other motifs that bind to RNAP II

5.2.2.A Initiation site of transcription

I found that at least one uridine is conserved in the bulge at position 1630, which was previously proposed to be the initiation site of transcription (fig. 5.1; Abrahem and Pelchat, 2008; Beard et al., 1996; Gudima et al., 1999). The conservation of this uridine is in accordance with *in vitro* mutations data showing that the change of this U to C and G results in a drastic decrease of the HDV RNA accumulation, of respectively 75% and 94%

(Appendix I; Gudima et al., 1999). This uridine is likely important as a transcription initiation site by RNAP II, and moreover it has been previously reported that RNAP II usually starts the transcription with a purine (Grosveld et al., 1981; Hawley and McClure, 1983). Similarly to HDV, the plant pathogen PLMVd also uses U for the initiation of transcription of its RNA promoter, however it uses a different host polymerase (Pelchat et al., 2002 ; Delgado et al., 2005; Motard et al., 2008).

5.2.2.B GC BOX

I found a conserved GC rich box on R199G, reminiscent of a feature of some DNA promoters (fig. 5.1; Novina and Roy, 1996). Indeed, CpG islands, which are enrichments in C and G, are often present in eukaryotic promoters that lack the TATA box or DPE (Antequera and Bird, 1993; Butler and Kadonaga, 2002; Carninci et al., 2006; Novina and Roy, 1996; Roeder, 1996; Smale, 1997). For DNA promoters, these CG rich regions are weak promoters (Butler and Kadonaga, 2002). Therefore, this GC motif present on HDV might play a role in the binding of the PIC or for the reinitiation of the RNAP II transcription. However, HDV uses an RNA template unlike the canonical eukaryotic transcription. Similarly to HDV, the PSTVd hijacks its plant host RNAP II to replicate its hairpin structured RNA genome, and a GC rich box has been identified around the region of initiation of transcription (Fels et al., 2001). These data support the hypothesis that the GC box might have a role in the use of RNAP II by RNA promoters such as R199G and future mutagenesis experiments upon these GC rich regions will allow for further defining their importance. In the case of the Hepatitis C virus, SELEX experiments identified that the viral RdRp had a high affinity for aptamers enriched in GC (Kanamori et al., 2009). Similar experiments with the RNAP II might help quantifying its affinity for GC rich motifs.

5.2.2.C CUC/GAG motif

Closer to the initiation site, I also identified a CUC/GAG motif that is conserved at least for 95% of the R199G variants (fig. 5.1). This high conservation supports the hypothesis that this motif is important for the recognition of the host proteins. For instance, it is known that CUG repeats is a motif specifically recognized by many host proteins, such as the CUG binding protein 1 (Teplova et al., 2010; Tsuda et al., 2009). Further experiments will allow the identification of which host proteins bind to this CUC motif during HDV replication. In addition, a GUC motif was also identified on the PLMVd RNA promoter (Delgado et al., 2005). The complementary GAG motif identified on HDV is reminiscent of the triplet present on both PLMVd RNA promoter in the CAGACCG motif and on *E. coli* DNA template for the *E. coli* RNA polymerase in the CAGACT motif (Motard et al., 2008). Therefore, this CUC/GAG motif could be important for the recruitment of the polymerase or other TFs associated with RNAP II during HDV replication.

5.2.2.D Pyrimidine/purine

Importantly, at the tip of the R199G stem, I identified a polarization of purine/pyrimidine (fig. 5.1). It is known for the mammalian DNA promoter template that the RNAP II binds with strong affinity to a pyrimidine rich region upstream of the initiation site of transcription (Szybalski et al., 1966; Sandelin et al., 2007). Therefore, the pyrimidine rich region in the proposed region of initiation of transcription of R199G could facilitate the binding of RNAP II.

In addition, previous studies showed that the conformation of the DNA is dependent of proportions whether a sequence is a mix of purines/pyrimidines or whether it is alternate regions enriched in purines or pyrimidines (Hantz et al., 2001; Huang and Lo, 2010; Konopka et al., 1985; Ng et al., 2000; Rehm et al., 2015; Ussery et al., 2002). For instance,

chemical bonds can be established between the purines and pyrimidines rich regions, therefore contributing to the structure of the nucleic acids (Buske et al., 2010). Therefore, this alternate purines/pyrimidines content identified on R199G likely influences the RNA conformation.

5.2.2.E Stem structure

I found a high level of covariation, suggesting a strong selection to maintain the secondary structure and preserve the rod-like structure (fig. 5.1; Beeharry et al., 2014). Usually RNA stem loop structures adopt the A conformation (Saenger, 1984; Weeks and Crothers, 1993). Unlike the DNA helix that can adopt A, B or Z conformations, the B form is not favourable for the RNA because of a steric clash due to the 2'OH of its ribose (Moore, 1999). In the eukaryotic cells, the B form of the DNA is most abundant and is generally the one recognized by the RNAP II. However, the A conformation of DNA can also be recognized by RNAP II (Kettenberger et al., 2006; Lehmann et al., 2007). RNAP II interacts with the major groove of the DNA of B- form, which is wider. The minor groove in the DNA in B-form is narrow (Kim et al., 1993). The binding of TBP to motifs, such as the TATA box, on the minor groove of the DNA allows for curving, partially unwinding and changing the DNA conformation (Kim et al., 1993). The binding of TAFs to DNA facilitates the interaction of the RNAP II with the nucleotide of the template (Helmann and deHaseth, 1999; Nikolov et al., 1992; Seeman et al., 1976; Verrijzer and Tjian, 1996). In the case of the RNA of A form, the minor groove is wide and shallow (Saenger, 1984). For the HDV RNA, the rod-like structure rather than the sequence might be more important for recognition of the RNAP II.

For the potato spindle tuber viroid (PSTVd), the promoter region of the genome can adopt both a Y-shaped or rod-like structure (Bojic et al., 2012). However, mutational studies

identified that the rod-like structure is the form that binds tomato RNAP II (Bojic et al., 2012). These findings further support that in case of a RNA template using RNAP II, the rod structure is important. Other studies of viroids that have a hairpin or stem-loop structure show that the promoter region is located in the terminal region (Beard et al., 1996; Filipovska and Konarska, 2000; Fels et al., 2001; Kolonko et al., 2006; Navarro and Flores, 2000; Pelchat et al., 2002; Flores et al., 2012; Motard et al., 2008; Pelchat and Perreault, 2004). Unlike the DNA template helix, for the rod-like RNA helix of these RNA pathogens, a stem with a terminal loop is recognized by the polymerase (Kolonko et al., 2006; Navarro and Flores, 2000; Pelchat and Perreault, 2004). These studies of other RNA templates for RdRps support the hypothesis that the stem-loop structure of R199G is important for HDV replication.

5.2.2.F Terminal loop

I found that the single-stranded terminal loop region is conserved among the sequences analyzed (fig. 5.1). Previous studies showed that for a DNA template that has a similar internal loop, electrostatic interactions between phosphates allow for the binding of the polymerase (Helmann and deHaseth, 1999). These interactions would be possible with the RNA stem-loop, and therefore this would support the hypothesis that the loop of the RNA is important for the interaction with RNAP II. Furthermore, in the case of the bacterial 6S RNA, it has been shown that the in is internal loop is important for its interaction with the bacterial RNAP (Gildehaus et al., 2007; Trotochaud and Wassarman, 2005; Wassarman and Saecker, 2006). It was proposed that this RNA template can act as a competitor to the host DNA of B form in order to take over the use of the host polymerase (Kettenberger et al., 2006; Schroeder and deHaseth, 2005; Thomas et al., 1997; Wassarman and Saecker, 2006). If the length of the central bubble of the bacterial RNA is altered, this impedes transcription.

However flip mutants or changes in sequence composition of the bubble only slightly interfere with the transcription (Trotochaud and Wassarman, 2005; Wassarman and Saecker, 2006). For homologues of the 6S RNA of more than 100 species of *E. coli*, the sequence is not conserved but a high level of secondary structure and a single-stranded region are conserved (Wassarman, 2007). It was suggested that internal loops are reminiscent of an open bubble of a DNA helix during transcription (Helmann and deHaseth, 1999; Steuten et al., 2014; Trotochaud and Wassarman, 2005; Wassarman, 2007).

For other RNA in eukaryotic cells, such as the structured tRNA, it is known that proteins usually bind the single stranded internal loop region (Moras and Poterszman, 1995). The presence of this bulge might contribute to the widening of the RNA A form, therefore facilitating the contact of the nucleosides with the amino acids of the polymerase (Steuten et al., 2014; Weeks and Crothers, 1993). Therefore, for the unconventional RNA promoter of RNAP II, that does not contain TATA boxes, the RNA structure and internal loop might be important (Novina and Roy, 1996). Based on these other studies of RNA promoters, it is highly plausible that the loop conserved for the R199G sequences might be important for the binding of RNAP II, by contributing to an RNA conformation favourable for the interaction with the host polymerase or by mimicking a transcription bubble.

5.2.2.G Bulges

A preliminary analysis of the conservation of the potential base pairs formed between the R199G bulges allowed for localizing two positions where for 99% of the sequences, there is a possibility that the base pairs adopt the same isosteric family (fig. 3.2.3; Mokdad and Frankel, 2008). These data support the hypothesis that the bulges might be involved in the establishment of the RNA tertiary structure. This hypothesis was initially proposed by the Leontis team that first developed these isostericity matrices (Leontis et al., 2002). *In silico*

localization of conserved isosteric base pairs between bulges of viroid sequences combined with mutational assays also allows for the identification of internal loops important for the function of PSTVd and supports the importance of the RNA structure (Takeda et al., 2011; Zhong et al., 2006; Zhong and Ding, 2008). Mutations affecting the loop or the bulge of PSTVd RNA result in a disruption of interaction with the plant RNAP II (Bojic et al., 2012). For other RNA viruses such a the Brome Mosaic Virus, the Dengue Virus and the Norovirus, the importance of the loop and the bulge of the promoter regions for the interaction with the viral RdRp has also been demonstrated (Kim et al., 1993; Lin et al., 2015; Lodeiro et al., 2009; Ranjith-Kumar et al., 2003). Based on the literature of the different RNA templates for cellular or viral RdRps and based on the conservation of the bulges regions identified in the present study, further mutational assays will expose the importance of the bulges of R199G for the structure of the RNA template for RNAP II.

5.2.3 Differences between an HDV template for RNAP II and other RNA promoters

The conserved R199G features could provide the foundation for a model of an RNA template for RNAP II. In addition, the conserved stem-loop structure and the bulges identified in the present study are in accordance with studies of bacterial, viroids or viral RNA templates for the bacteria, T7 or plant RNAPs (Carpenter and Simon, 1998; Fels et al., 2001; Kolonko et al., 2006; Lauber et al., 1997; Navarro and Flores, 2000). Structural analysis of RNAP II and these RNAPs suggested homologies between the core domains and the same requirement for two metal ions in order to ensure their functions (Poole and Logan, 2005). One study showed that the *E. coli* RNAP can transcribe the PLMVd promoters (Pelchat and Perreault, 2004). However, this use of an RNA template by the polymerase of a different host is not always possible. For instance, the X and Y are RNA templates for T7

polymerase, but they cannot be synthesized by the SP6 polymerase nor the *E. coli* RNAP (Konarska and Sharp, 1989, 1990).

Even if the different RdRps have highly similar cores, the RNA templates of these different RdRps also present differences in their features. For instance, there are A/U rich sequences for the RNA promoter of the bacteriophage T7 (Konarska and Sharp, 1989; Konarska and Sharp, 1990), for the *E. coli* 6S RNA (Biebricher and Orgel, 1973; Gildehaus et al., 2007) for PLMVd and ASBVd (Navarro and Flores, 2000; Pallas et al., 1988; Rakowski and Symons, 1989) and I did not identify any AU-rich motif in the analysis of the R199G sequences. Another difference is the nucleotide used as a start site: while for the R199G we identified a U as the start site, for the RNA promoter of PSTVd an A was identified as the start site (Fels et al., 2001). One explanation for the differences in features of the various RNA promoters might rely on the distinctions between the bacteriophage, *E. coli*, plant and chloroplastic polymerases, and the human RNAP II. For instance, the *E. coli* and T7 RNA polymerases are more divergent (Cermakian et al., 1997; Iyer et al., 2003; Velazquez et al., 2012). RNAP II is a multi-subunit polymerase unlike the bacteriophage, mitochondrial and chloroplastic polymerases. This difference between the polymerases might be at the origin of the differences in their function during the initial recognition and interaction with their promoter (Velazquez et al., 2012). Therefore, this might explain some disparities in RNA features identified on R199G and other bacterial RNA promoters.

Another explanation for these various RNA promoter features might be the cellular components and environments of the different hosts in which the polymerases evolved. For instance, since the concentrations in NTPs of the environment influences the error rate and the composition of nucleotides incorporated, the RNA promoter features may vary according to the cellular NTP concentrations environments (Anglana et al., 2003; Barker and Gourse,

2001; Gaal et al., 1997; Guajardo et al., 1998; Mathews, 2006). Based on this, the function of the RNA promoters might not be necessarily conserved among the different RNA templates-RNAP couples. For instance, even if both the bacterial 6S RNA and the murine B2 RNA have similar structural RNA features and interfere with the host DNA transcription by binding the RNAP, the mechanism of function of the RNA template is not necessarily conserved. The 6S RNA competes with the host DNA for binding of the RNAP while the B2 RNA interferes with the host DNA-RNAP II complex initially formed (Espinoza et al., 2004; Goodrich and Kugel, 2006; Kettenberger et al., 2006; Wagner et al., 2013; Wassarman and Saecker, 2006). Therefore, this evolution of the various RNAPs in their respective host cells could explain that some features identified on the HDV R199G region are different from features of other RNA templates. Moreover, unlike the viroids that do not have any additional proteins, HDV uses HDAg-S for its replication (Lee and Sheu, 2008; Lin et al., 1990; Yamaguchi et al., 2001). Therefore, the HDAg-S might be another factor explaining why the HDV promoter features required for use by the human RNAP II present some disparities with the viroid promoters using the plant polymerases.

Nonetheless, the study of RNA virus RdRps, could still provide insight on general/ fundamental characteristics of cellular RNAPs (Ng et al., 2008). This characterization of the mechanism of recognition of RNAP II by the RNA might help better understand the degree of evolution and genetic distances separating the different RNAPs. It has been proposed that all RNAPs might have evolved from a common ancestor, and moreover phylogenetic studies have shown that RNAPs share many homologies with RdRPs (Gilbert, 1986; Iyer et al., 2003). Therefore, it is not surprising that the features I have identified for HDV RNA template for RNAP II include some that have been previously encountered for the DNA template of RNAP II, and for the RNA templates of other viral or host RNAPs.

5.2.4 Mutation rate of HDV right terminal region

I analyzed the sequence space of the HDV promoter region, which has been previously reported to be the most conserved region of HDV genome (Chang et al., 2005). I found that one of the sequence variant represents 76.3 % of the population, which confirms that the R199G region is well conserved (table 3.2, Appendix IV). However, even in a cellular system where the HDAg-S is given in trans, therefore reducing the selective pressure, HDV right terminal region exists as a heterogeneous sequence population. This is in accordance with the characteristic of RNA viruses to have a high polymerase error-rate, giving rise to a quasi-species population (Domingo et al., 2012). To deduce the mutation rate due to the experimental procedures, I used a control that has the same sequence as the sequence initially used to transfect the 293-HDV cells. The global mutation rate calculated for HDV in the present study is 8.1×10^{-4} mutation/site which is within the mutation rate for RNA viruses of 10^{-5} to 10^{-4} mutation per site, but higher than the mutation rate of the human RNAP II when transcribing DNA templates, which is approximately 10^{-5} mutation/site (Cramer, 2004; Domingo et al., 2012). However, it is possible that the mutation rate of the HDV genome is also influenced by other human transcriptional factors such as NELF and DSIF, which are involved in the RNAP II pausing during cellular transcription (Yamaguchi et al., 2002).

I calculated the mutation rate (mutation/nucleotide), but not the mutation frequency (mutation/nucleotide/replication cycle), since the exact number of viral cycle was not known (the exact number of passages and also the exact number of viral cycles per passage). I can only estimate that approximately one year separates the time of the initial transfection of the clonal sequence and the time of extraction of the viral RNA for deep-sequencing. Moreover, this mutation rate is to be taken with caution since it only reflects the error rate during HDV

replication in a particular cellular system, where the HDAg-S is provided in trans. I am not testing the ability to make HDAg mRNA, therefore it is close to a viroid replication system. The viroids, other small RNA pathogens, rely exclusively on their genome and their host proteins for their viral cycle since they do not encode any protein. Viroids have been shown to have the highest error-rate in nature, with 10^{-3} mutations/sequence/cycle (Gago et al., 2009; Glouzon et al., 2014). However, the results of S Gago and JP Glouzon studies and my present study are not directly comparable as different approaches were used. For instance, one difference is the sample size of the population analyzed. S Gago and colleagues amplified and cloned the isolated viroids RNA, sequenced the insert of each clone by Sanger sequencing and obtained approximately 300 sequences, repeated in three independent experiments (Gago et al., 2009). J.P. Glouzon and colleagues performed a deep-sequencing with two sets of primers directed against different regions of the genome, and analyzed 259, 486 sequences representing 2186 variants (Gago et al., 2009; Glouzon et al., 2014). Additionally, I cannot directly compare the error-rate of these two studies with my present study as the viroid studies have been performed in the context of an *in vivo* natural infection, and have more selective pressure from the viroid plant host, such as the plant immunity (Goodman et al., 1984). The mutation rate of RNAP II may vary due to the interaction with other cellular components present in the host cell, such as transcriptional factors (Yamaguchi et al., 2002). A combination of studies of HDV populations from different systems and *in vivo* isolates will provide insights on the variability of the HDV genome in the presence of the transcriptional factors in these various environments.

Another major difference being that viroids are only composed of their RNA genome while HDV also encodes a protein. The HDAg-S has been reported to be involved in the replication step and might help regulate the progression of RNAP II (Yamaguchi et al., 2001;

Yamaguchi et al., 2002; Yamaguchi et al., 2007). Using a DNA template, Y Yamaguchi and colleagues showed that the mutation rates of RNAP II is increased by the action of the HDAg-S. Based on this, it is possible that HDAg-S contributes to the progression of RNAP II on the HDV RNA and affects the nature of the nucleosides incorporated. If it is the case, the presence of high levels of HDAg-S would contribute to an increase in the diversity of the quasispecies populations. Future experiments, such as *in vitro* transcription on HDV RNA in the presence of different concentrations of HDAg-S, will shed light on the ability of the HDAg-S to modulate HDV mutation rate. Understanding how HDAg-S modulates the quasispecies populations could provide a foundation to design therapeutics targeting HDV replication, for instance by driving the equilibrium of the HDV quasispecies population to a lethal sequence space.

For the promoter region, I report that there is 91.9% of transitions. To date, this is the only deep-sequencing study of a HDV population and provides around 500 000 sequences of the promoter region, composing a same viral population. The majority of the studies analyzed the antigen ORF region (Lee et al., 1992; Netter et al., 1995). Previous similar studies based on Sanger sequencing provided dozens, to at best a few dozens sequences, which does not provide a high statistical power of the mutation rate calculated as the totality of the population is not represented (Casey et al., 2006; Deny et al., 1991; Krushkal and Li, 1995; Netter et al., 1995; Polson et al., 1998). Based on the analysis of the entire genome of three HDV sequences isolated from a patient chronically infected, Lee CM et al estimated that the HDV mutation rate was around $3x10^{-2}$ to $3x10^{-3}$ substitution/nucleotide/year, which is higher than the mutation rate that I calculated for the promoter region (Lee et al., 1992). This is expected since the promoter region is more conserved than the rest of HDV genome (Chang et al., 2005). Chang et al used the same cellular system as in my study; after performing

cloning and Sanger sequencing of nine sequences, they found that there was 2.1 changes/nucleotide/year. They found that the rod-like structure was conserved at 100%. Since the antigen is provided exogenously by the plasmid transfected in the cells, the selective pressure on the sequence of the ORF is lowered, but not the selective pressure on the features required for HDV replication. In their previous studies, it was found that 90% of the mutations were transition U>C and A>G, and they suggested that 70% of the mutations are likely caused by ADAR editing (Chang et al., 2005).

Previous studies have suggested the presence of viruses defective for the synthesis of HDAg in the HDV population (Chang and Taylor, 2002). Defective viruses have a role in the quasi-species viral population, such as contributing to the diversity of the viral population and thereby participating the immune escape of the virus (Chang and Taylor, 2002; Wu et al., 2005). Therefore, I cannot exclude that the mutation rate calculated in my report may be changed due to the presence of defective viruses or due to recombination events (Table 3.1.1, Appendix III). Also, I cannot exclude the possibility that there is genome recombination in the HDV population, as previous studies have suggested that recombination occurs in the region coding for HDAg (Anisimova and Yang, 2004; Chao et al., 2006; Gudima et al., 2005). In addition, it was previously proposed that recombination is more likely to occur in regions with high levels of secondary structure (Lai, 1992; Nagy and Simon, 1997; Wu et al., 1999).

Importantly, for 1.2% of the population, there is a mutation in the second nucleotide of the first codon UAC of the HDAg-S, which corresponds to the AUG initiation codon of the HDV mRNA. This finding is in accordance with the hypothesis that the selective pressure would favour a lost of the ability of HDV RNA to code for the HDAg-S: since in this cellular

system the antigen is already provided, therefore the synthesis of HDAg-S would represent a cost to recruit the translation machinery. My study also suggests that even without the HDAg-S initiation codon, HDV is still able to ensure its replication, which is in contrast with a recent study suggesting that the HDAg initiation codon was required for HDV replication in a similar cellular system where the HDAg-S was provided (Liao et al., 2012). The team of Liao F-T showed that the mutation of the HDAg initiation codon to a stop codon resulted in a decrease of accumulation of HDV RNA dimers. One could argue that a limitation of this study is the system used: they transfected a DNA plasmid, while in the normal HDV cycle there are only RNA intermediates. Therefore, the replication and transcription process might not be representative of the promoter required in the normal HDV cycle.

I looked at the global number of mutations per sequence, but I also looked at the number of mutations per position, which allowed for the identification of hot spots of mutations present in the viral population, but not for the control. I cannot affirm whether these mutations are due to modifying enzymes, to the behavior of RNAP II in this nucleotidic and/or structural context or if it is due to selective pressures on HDV genome (Blake et al., 1992; Domingo, 2015; Escarmis et al., 1996; Lee et al., 2012; Long et al., 2015; Maris et al., 2005; Medvedeva et al., 2013). Based on the few previous analyses, the mutation rate is higher in other regions of the HDV genome, such as the coding region of HDAg that may allow Wobble codons (Huang and Lo, 2010). Future research using high-throughput sequencing technologies will be necessary in order to obtain a statistical confirmation of the variability of other regions of the HDV genome (Lee et al., 1992; Netter et al., 1995; Polson et al., 1998).

5.3 Involvement of the host paraspeckle components in HDV replication

5.3.1 Comparison of R199G features to other RNA motifs that bind to PSF, p54 and PSP1

5.3.1.A Purine and pyrimidine motifs that bind PSF, p54 and PSP1

Previously, the laboratory showed that R199G binds PSF (Greco-Stewart et al., 2006). *In vitro* immunoprecipitation assays indicated that both the change in sequence and the disruption of the secondary structure of R199G resulted in an abrogation of the interaction with PSF (Greco-Stewart et al., 2006). The change in primary sequence might disrupt the RNA structure, and it was hypothesized that the R199G stem loop structure rather than its primary sequence might be important for the interaction with PSF (Greco-Stewart, 2009; Greco-Stewart et al., 2006). Furthermore, it was proposed that the side of the R199G stem encompassing the trinucleotide bulge region (positions 1629 to 1631) or the tip containing the purine/pyrimidine polarization might be important for the binding of PSF (Greco-Stewart, 2009; Greco-Stewart et al., 2006).

My model obtained from the analysis of HDV sequences is in accordance with these data: for the region bound by PSF, I found a purine/pyrimidine polarization at the tip of HDV (fig. 5.1). The enrichment in purines is a characteristic of the RRM bound by PSF (Abrahem and Pelchat, 2008; Peng et al., 2002). p54, PSF and PSP1 have been shown to interact with both cellular single and double stranded DNA and RNA and I have summarized the structural features of their RRMs in Appendix V. p54 and PSF are associated with DNA folded around histone octamers (Basu et al., 1997). A previous study showed that p54 binds with high affinity to the long terminal repeats region of the enhancer element of intracisternal A particles, a cellular DNA enriched in purines (Basu et al., 1997). In addition, the purine rich motif "AGGGA" was found in a majority of sequences that preferentially bind to the DNA-binding domain of p54, situated in the N-terminal protein region (Basu et al., 1997).

Therefore, the binding of PSF and p54 to cellular DNA enriched in purines is in accordance with the hypothesis that the purine rich motif of R199G is a site preferential for binding to PSF, p54 and PSP1.

Using a SELEX approach, the Patton team investigated the nucleotide composition of synthetic sequences that preferentially bind to PSF and p54; they found that there was an enrichment in purines (50 to 67%; Peng et al., 2002). Furthermore, they discovered that this consensus sequence 5'-UGGAGAGAGGAAC-3' was present on the purine rich sequence close to the 3' terminal end of the single stranded U5 snRNA bound by either PSF, p54 or by both (Peng et al., 2002). The GAG motif present in this consensus motif is also present in the DNA of the Insulin-like Growth Response Element (IGFRE) box bound by PSF (Peng et al., 2002; Urban et al., 2000). Moreover, in the R199G sequences analyzed, I identified the presence of a highly conserved GAG motif (fig. 5.1). This supports the hypothesis that the GAG motif could also be important for the binding of PSF and p54 to R199G.

Several studies reported that both p54 and PSF have a high affinity for RNA rich in G and inosines (DeCerbo and Carmichael, 2005; Prasanth et al., 2005; Zolotukhin et al., 2003). Under the form of p54-PSF heterodimer, p54 has a high affinity for A-to-I edited RNAs, therefore enabling the sequestration of double-stranded RNAs in the nucleus (Zhang and Carmichael, 2001). The formation of paraspeckles depends on to the ability of PSF, p54 and PSP1 to bind the NEAT1 RNA, which is enriched in the purine G (Bruelle et al., 2011). In addition to the cellular DNA and RNA, several teams have reported the ability of PSF and p54 to bind RNA pathogens (Schwartz et al., 1992; Zolotukhin et al., 2003). HIV also hijacks the cellular heterodimers of PSF/p54 that bind to the Gag mRNA cis-acting instability elements (INS) sequence enriched in purines (Schwartz et al., 1992). In summary, these data showing that both cellular and viral RNAs enriched in purines bind PSF, p54 and PSP1 with

high affinity are concordant with the model I presented in fig. 5.1, showing an enrichment in G and A at one side of the tip of R199G.

However, PSF was initially identified as the polypyrimidine-tract splicing factor (Patton et al., 1993). PSF binds polypyrimidine tracts of both cellular DNA and viral RNAs (Heise et al., 2006; Kidd-Ljunggren et al., 2000; Song et al., 2004). For instance, PSF binds the polyprimidine tract of the 3' splice site of HBV mRNA (Heise et al., 2006; Kidd-Ljunggren et al., 2000). Moreover, for the mouse retro-element Virus-Like 30S (VL30), PSF RRM binds to two polyprimidine tracts named ptb-A and ptb-B (Song et al., 2004; Song et al., 2005). It was suggested that this binding releases PSF from the DNA region enriched in pyrimidine in the GAGE 6 gene (Song et al., 2005). Therefore, it is possible that PSF, p54 and PSP1 also bind the pyrimidine motif of R199G. However, the affinity of the binding of PSF, p54 and PSP1 might be different for a purine rich or pyrimidine rich region.

Yet another study suggests that PSF would bind RNA enriched in purines with higher affinity than DNA (Peng et al., 2002). It seems that PSF, p54 and PSP1 can bind different motifs, both purine and poly-pyrimidine tracts of RNA or DNA, with different affinity and/ or avidity (Song et al., 2004; Song et al., 2005). In addition, different post-translational modifications of PSF/p54/PSP1 such as phosphorylation might also dictate their affinity for nucleic acid, as well as their association as homo or hetero dimer (Bruelle et al., 2011; Zhang and Carmichael, 2001; Zolotukhin et al., 2003).

The effect of the phosphorylation state of p54 has been linked to its ability to bind an RNA *in vitro*: the phosphorylated form of p54 abrogates its ability to bind poly(A), poly(C) and poly(U) synthetic RNAs, but does not change its ability to interact with poly(G) (Bruelle et al., 2011). Accordingly, *in vivo* the NEAT1 RNA which is G-rich still interacts with p54 in interphase or with its mitotic phosphorylated form (Bruelle et al., 2011). Similarly, PSF and

PSP1 recognize similar RRM since they have an affinity for poly(A), poly(U) and poly(G) RNAs and for NEAT1 RNA (Baltz et al., 2012; Clemson et al., 2009). Therefore, these findings support the hypothesis that PSF, p54 and PSP1 have a higher affinity for purines than polypyrimidines sequences. Altogether, these previous studies support the proposition that the purine rich sequence of R199G is the region that binds with stronger affinity to the human proteins involved in HDV replication. Nonetheless, I cannot completely exclude the possibility that the complementary pyrimidine region on the other side could also be bound by the host factors. However, the RNAP II of high molecular weight occupies this site, so it is unlikely that other cellular proteins are also able to bind this site concomitantly. As previously done for PSF, future mutagenesis experiments will clarify the importance of the purine/pyrimidine polarization for binding p54 and PSP1.

I also found a GC box upstream from the region enriched in pyrimidines. Similarly, a DNAse I footprint has previously identified a GC box close to the palindrome motif bound by PSF on the IGFRE motif of the porcine P450scc gene (Urban et al., 2000). Therefore, I cannot exclude that the GC box of R199G situated upstream from the region enriched in pyrimidines also favours the binding of PSF, p54 and PSP1.

HDV relies heavily on its cellular host to ensure its replication and hijacks many host proteins (Beeharry and Pelchat, 2011). It is also possible that other cellular proteins influence the binding of PSF, p54 and PSP1 to R199G. For instance, it was previously shown that the PSF RRM binding affinity can be modulated through association with the host protein Hakai (Figueroa et al., 2009b). When the tumour promoting factor Hakai binds PSF, it triggers the dissociation of PSF from DNA and increases the association of PSF with other host mRNA

linked to oncogenes (Figueroa et al., 2009b). Based on this study, I cannot exclude that the recruitment of PSF, p54 and PSP1 to R199G is indirect and relies on the hijacking of another cellular component by HDV.

5.3.1.B Stem loop structure of the RNA ligands for PSF, p54 and PSP1

For the region upstream from the tip of R199G, a high level of covariation indicated that the stem-loop secondary structure seemed to be important. In my thesis, I used isostericity matrices to analyze the importance of the bulge of R199G stem and found that two positions where mutations present in the sequences studied seem to preserve the same isosteric family. The structural context on the RNA influences the ability of the sequence to be recognized by a protein (Maris et al., 2005). In some cases, the binding of a protein to the RNA can remodel its structure, which makes a proximal sequence permissive to the interaction with the protein (Allain et al., 1996; Showalter and Hall, 2004). Based on this, I hypothesized that the stem-loop structure of HDV might be important to bind PSF, p54 and PSP1. This is in accordance with another study indicating that the heterodimer PSF-p54, or PSF and p54 separately have a high affinity for a stem loop structured region of the host U5 snRNA, the stem 1b (Peng et al., 2002). Interestingly, the authors showed that this hairpin-loop structure of snRNAs which binds PSF is quite conserved among eukaryotes, which suggests the importance of the RNA stem-loop structure for the interaction with PSF and p54 (Peng et al., 2002). *In vitro* mutagenesis experiments showed that the RNA structure of the stem 1b was required for the interaction with PSF and p54 but the change of sequence of the 3' terminal end of stem 1b did not abrogate the PSF-p54-RNA interaction (Peng et al., 2002). Based on *in vitro* immunoprecipitation experiments, the authors suggest that both the sequence and the structure of the stem 1b RNA contributes to the interaction with PSF and p54 (Peng et al., 2002). Furthermore, PSF, p54 and PSP1 bind NEAT1, which is a highly

structured RNA (Appendix V; Bruelle et al., 2011; Clemson et al., 2009; Murthy and Rangarajan, 2010; Sasaki and Hirose, 2009). In summary, these studies suggest the importance of the stem loop for the interaction of R199G with PSF, p54 and PSP1.

PSF binds the HBV mRNA in a hairpin and stem loop region (Heise et al., 2006; Kidd-Ljunggren et al., 2000). Different teams have characterized the binding of PSF, p54 and PSP1 to the stem loop structures of the mRNA containing the Rev-responsive Element of HIV (Mann et al., 1994; Naji et al., 2012; Nathans et al., 2009; Zolotukhin et al., 2003). These studies of other RNA pathogens further confirm that the stem-loop of R199G might be important to interact with PSF, p54 and PSP1. For EBV, p54 has a role in the functional regulation of the viral cycle: the hyper-edited origin of replication forms a hairpin RNA structure which interacts with p54 and therefore promotes lytic transcription (Cao et al., 2015). For HIV RNA, a GO analysis of the interactants of the Rev Recognition Element (RRE), an RNA with an important secondary structure and stem loop regions, identified PSP1 as a binding protein (Mann et al., 1994; Naji et al., 2012). PSF and p54 also binds with high fidelity to the INS sequence of HIV RNA transcripts in a structure dependent manner and recruit the viral RNA in paraspeckles, therefore restricting their expression (Houzet and Jeang, 2011; Huang et al., 2012; Liu et al., 2011; Nathans et al., 2009; Zhang et al., 2013a; Zolotukhin et al., 2003; Kula et al., 2013). For instance, the PSF/p54 heterodimer strongly binds the INS sequence of p37 gag mRNA (Zolotukhin et al., 2003). p54 is subsequently transported to the cytoplasm where it interacts with an INS-enriched HIV mRNA (Zolotukhin et al., 2003). PSF represses the expression of HIV-1 Rev-responsive element-containing mRNAs by interacting with cis-acting regulatory elements of HIV mRNAs (Zolotukhin et al., 2003). The Rev-responsive element-containing mRNAs are highly structured and form hairpin and stem loop domains reminiscent of the R199G structure

(Beard et al., 1996; Jain and Belasco, 2001; Wolff et al., 2003). These studies of the interaction of EBV and HIV RNA to PSF and p54 lend support to the hypothesis that the stem-loop structure of R199G might be important for the interaction with PSF, p54 and PSP1.

Since HDV RNA competes with cellular nucleic acids for PSF interaction, HDV RNA might use "molecular mimicry" to hijack its host protein PSF (Colussi et al., 2014; Greco-Stewart, 2009). Another example of how an RNA pathogen can use molecular mimicry to hijack PSF is the study of VL30 (Song et al., 2004; Song et al., 2002). VL30 is a pathogen RNA similar to proviruses (Itin and Keshet, 1983; Song et al., 2005). It was suggested that when the RRM of PSF interacts with the mouse VL30, this may release the DBD of PSF interaction with the cellular DNA (Song et al., 2005; Song et al., 2002). Mutations changing the sequence but not the composition of the two polypyrimidine tracts, ptb-A and ptb-B, resulted in an abrogation of the interaction with PSF (Song et al., 2004). Ptb-A and ptb-B of VL30 compete with the IGFRE motif of the porcine gene P450scc (Song et al., 2004; Urban et al., 2000). The authors suggest that the occupancy of the RRM by VL30 RNA might create allosteric change and thereby reduce the bonding of the DBD for its DNA ligand (Song et al., 2004). The quantity of VL30 is one of the factors that determine this competitiveness of the RNA and DNA for PSF binding (Song et al., 2004). Future investigations will determine whether the interaction of PSF, p54 and PSP1 RRMs to HDV competes with the interaction of these proteins to host nucleic acids. One approach would be to screen for cellular gene deregulation that could be triggered by the titration of these proteins by HDV RNA.

5.3.2 Role of PSF, p54 and PSP1 in HDV viral cycle

I observed that upon knockdown of PSF, p54 and PSP1, the HDV levels are decreased by more than 90%, which support the hypothesis that PSF, p54 and PSP1 are important for HDV accumulation and/or replication in cells. An interaction of HDV with these proteins may result in their sequestration, and I would expect that HDV replication produces a similar effect than the knockdown of the DBHS proteins in the cells. PSF and p54 are at the crossroads of multiple RNA pathways, such as splicing, polyadenylation, transcriptional regulation and nucleic acid unwinding (Shav-Tal and Zipori, 2002), while PSP1 role in the cell remains largely unknown and its involvement in the RNA pathways remains to be characterized (Fox et al., 2002; Gao et al., 2016; Myojin et al., 2004). The knockdown of these proteins affects other cellular proteins involved in the RNA pathway metabolism and can cause effects such as cell cycle arrest, deregulation of the transport and stabilization of RNAs, changes in transcripts and protein expression and deregulation of the cellular immunity cytokines (Yarosh, 2015).

5.3.2.A Cell Cycle Arrest

In my study, PSF was knocked down to approximately 90%, which is comparable to previous knockdown studies with a decrease between 70% to almost 100% (Kula et al., 2013; Rajesh et al., 2010; Sasaki and Hirose, 2009). I was expecting that it would not be a complete knockdown of PSF since previously, many studies have described PSF as a protein vital for the cell, and suggest that a complete knockout of PSF would be non-viable, but a partial knockdown of PSF would be possible (Gozani et al., 1994; Patton et al., 1993; Shav-Tal and Zipori, 2002; Yarosh et al., 2015; Lowery L A, 2007; Ha et al., 2011; Kaneko et al., 2007; Mahaney et al., 2009). One explanation to the discrepancy of the different research teams concerning the proteins levels observed upon a PSF knockdown might rely in the different

cell lines used, the region targeted by the siRNA used, the number of siRNA used and also the number of days between the transfection of the siRNA and the measurement of the protein levels. Upon p54 knockdown, I found a decrease of 88%, which is higher than the 60% decrease of a previous study (Sasaki and Hirose, 2009). For the knockdown of PSP1, the protein levels decreased by 90 to 95%, which is comparable to the decrease of 85% reported in the literature (Gao et al., 2016; Li et al., 2014; Sasaki and Hirose, 2009). PSF, p54 and PSP1 have a role in the regulation of the cellular DNA replication and in the repair of DNA damage response, which is in accordance with data showing that the silencing of PSF, p54 and PSP1 leads to a cellular cycle arrest (Akhmedov et al., 1995; Akhmedov and Lopez, 2000; Bladen et al., 2005; Gao et al., 2014; Gao et al., 2016; Ha et al., 2011; Lin et al., 2000; Mahaney et al., 2009; Matsuoka et al., 2007; Rajesh et al., 2011; Rajesh et al., 2010; Udayakumar et al., 2003; Wu et al., 2010).

Similarly to cells knocked down for PSF, p54 and PSP1, the cell cycle of cells replicating HDV seems to be perturbed. The induction of HDV replication caused 293-HDV cells to stop in the G1/G0 phase and to detach from the culture dishes after 72 hours (Chang et al., 2005). One hypothesis to explain this would be that the recruitment of PSF, p54 and PSP1 by HDV result in an impairment in the DNA damage repair pathway, preventing the cell to pass the checkpoint from G1 to S. It is known that the Herpes Viruses, Myxomavirus T-5, HTLV-1, SARS, EBV, Cytomegalovirus and Murine Coronavirus mouse hepatitis virus induce their host cell arrest in the G1/G0 state in order to pursue their viral cycle (Cayrol and Flemington, 1996; Flemington, 2001; Hobbs and DeLuca, 1999; Johnston et al., 2005; Kannian and Green, 2010; Lomonte and Everett, 1999; Lu and Shenk, 1999; Song et al., 2001; Wiebusch and Hagemeier, 1999; Yuan et al., 2006). It has been proposed that by forcing cellular arrest in G1/G0 and inhibiting cellular replication, this creates a cellular

environment enriched in components usually involved in replication, benefiting the replication of the viruses (Flemington, 2001). In order to understand what causes this G1/G0 arrest in HDV viral cycle, it will be interesting to compare among the direct or indirect interactants of PSF, p54 and PSP1 whether there are players, such as kinases, that are important for the cell cycle transition to G1/G0 state. It will also be interesting to quantify the amount of the different post-translational forms of proteins important for the G1/G0 arrest, such as the cyclin E-Cdk2 or the retinoblastoma protein in cells during HDV replication. Further investigation will be required to understand whether this G1/G0 resting cell state in the 293 cells benefits HDV replication.

5.3.2.B Transport and stabilization of cellular and viral RNAs

PSF, p54 and PSP1 are involved in the dynamic regulation of the cellular RNA localization, transport and its stabilization (Chen and Carmichael, 2009; Hundley and Bass, 2010; Lu and Sewer, 2015; Major et al., 2015; Mao et al., 2011a; Nakagawa and Hirose, 2012; Sasaki and Hirose, 2009; Yarosh et al., 2015; Zhang and Carmichael, 2001). Since PSF, p54 and PSP1 participate in the stabilization of cellular RNA, one possible consequence of the interaction of HDV with these DBHS proteins is the deregulation of the transport and stability of RNA transcripts and an alteration of the cellular RNA metabolism (Lu and Sewer, 2015). It is also known that PSF is involved in the stabilization of the viral RNA and maintains a nuclear pool of unspliced HIV-1 RNA, making it readily available for the proteins Rev and MATR3 (Kula et al., 2013). A. Kula and colleagues reported that the knockdown of PSF interfered with the nuclear export of the HIV-1 RNA via Rev, but did not affect the viral transcription levels (Kula et al., 2013). Upon binding of Rev to the Matrin3-PSF-HIV1 RNA complex, it allowed nuclear export of Rev and the HIV-1 RNA (Kula et al., 2013). This illustrates how PSF can be involved in the viral cycle by playing an intermediate

role between the viral RNA and other cellular proteins. This hypothesis is also very likely in the case of the HDV: PSF, by binding the viral HDV promoter, could allow the recruitment of other host proteins required for transcription and replication of the HDV genome. The report of A. Kula and colleagues shows that the hijacked PSF function is not necessarily only dependent of RNAP II activity. Based on this example, we therefore cannot assert that the decrease of HDV upon PSF, p54 and PSP1 knockdown is solely due to the disruption of these proteins role in HDV transcription via RNAP II. This effect could also be caused by other roles of PSF, p54 and PSP1 in HDV cycle, such as stabilizing the HDV RNA.

5.3.2.C Deregulation of splicing and changes in protein expression

PSF and p54 are both heavily involved in splicing and their depletion leads to a deregulation of the transcriptional pathways (Emili et al., 2002; Gozani et al., 1994; Kameoka et al., 2004; Kaneko et al., 2007; Patton et al., 1993; Peng et al., 2002; Dong et al., 2005; Dong et al., 2007; Ishitani et al., 2003; Mathur et al., 2001; Song et al., 2005; Urban et al., 2000; Urban et al., 2002; Yarosh et al., 2015). HDV hijacks its host proteins greatly to ensure its replication, and could therefore interfere with a great number of cellular pathways (Beeharry and Pelchat, 2011). Concordantly, upon HDV replication, the cell transcriptional expression is largely deregulated (Mota et al., 2009; Mota et al., 2008; Cao et al., 2009; Monga and Cagle, 2010). Another study by A Garren and colleagues shows that the binding of PSF to the P450ssc gene results in the repression of steroid synthesis (Song et al., 2004). The knockdown of PSF resulted in an effect similar to the titration of PSF by VL30, i.e. an overexpression of the steroids (Song et al., 2004). Similarly, the cellular consequences of the knockdown of PSF, p54 and PSP1 could be indicative of the consequence of PSF, p54 and PSP1 titration by HDV. Therefore, we cannot exclude the possibility that the effect of the reported decrease of HDV levels upon knock down of PSF, p54 and PSP1 could be indirect

and caused by the deregulation of other genes, RNA and proteins downstream of PSF, p54 and PSP1.

Analogously to our findings, the knockdown of PSF in A549 cells led to a decrease of the IAV gene expression by a 2-5 log fold change (Landeras-Bueno et al., 2011). However, the knockdown of p54 had no effect on the IAV replication (Landeras-Bueno et al., 2011). The use of a virus replicon system allowed for identifying that the knockdown of PSF affected IAV transcription but not IAV replication: the mRNA levels were decreased by 5 fold but not the vRNA levels. More specifically, the knockout of PSF led to a 5 fold decrease in polyadenylation, but did not affect cap-snatching (Landeras-Bueno et al., 2011). One similar feature between IAV transcription and HDV replication is that they are both linked to the cellular RNAP II transcription activity (Beeharry and Pelchat, 2011; Engelhardt et al., 2005). Moreover, PSF is intrinsically linked to RNAP II activities (Emili et al., 2002). For HDV it is harder to differenciate transcription from replication, since both mechanisms use RNAP II. Further work will be required to determine whether it is HDV transcription and/or replication that is affected by the knockdown of the DBHS proteins. By identifying whether HDV mRNA polyadenylation and capping are affected at the same levels upon partial PSF knockdown, it may be possible to gain insights in the involvement of PSF in HDV transcription. The better understanding of how HDV and IAV uses RNAP II for their viral cycle might also help further understand the involvement of PSF in the cellular transcriptional pathways.

5.3.3 PSF is involved in RNAP II directed replication of HDV RNA

PSF and p54 bind the CTD of RNAP II and regulate elongation of transcription (Cao et al., 2015; Emili et al., 2002). Given that PSF, p54 and PSP1 have different roles in many

cellular pathways, I cannot exclude that the silencing of PSF, p54 and PSP1 also leads to down or up-regulation of other proteins heavily involved in several RNA pathways or in antiviral immunity. For instance, the knockdown of these proteins could change the expression of other transcripts and inhibit the accumulation of HDV genomes in the nucleus. Therefore, only based on these experiments I cannot affirm that PSF, p54 and PSP1 are solely involved in the viral cycle part of the hijacking of RNAP II by HDV RNA. However, in addition to the present experiments, previous immunoprecipitation experiments carried out in the laboratory show that the same HDV genome mutants that are unable to bind RNAP II are also unable to bind PSF (Greco-Stewart, 2009). Moreover, the transcription assay previously developed shows that HDV is able to transcribe its genome using the nuclear RNAP II, but when PSF is immunodepleted, this leads to a loss of the transcriptional signal (Sikora, 2012). Accordingly, the addition of PSF to the immunodepleted extract allows the recovery of the HDV transcriptional signal (Sikora, 2012). Altogether, this strongly supports the hypothesis that PSF is required for HDV to recognize and use RNAP II to synthesize its genome.

5.4 Link between HDV replication and paraspeckles

5.4.1 The effect of HDV replication on the host paraspeckle components PSF, p54, PSP1 and NEAT1

I showed that in this cellular system allowing the replication of HDV, PSP1 is present as large bright foci outside of the nucleus (fig. 4.2.1 ; fig. 5.2). To date, this is the first report of a concentration of PSP1 as a large bright foci in the cytoplasm. Other studies have shown that under transcriptional arrest and under stress conditions, PSP1 relocalized to the perinucleolar caps (Andersen et al., 2002; Fox et al., 2005; Shelkovnikova et al., 2014).

In the cellular system used, it was previously reported that the majority of cells remain blocked in the G1/G0 phase. In addition, a day after addition of TET, I observed that there was a decrease of cell adherence and by day four almost all the cells were detached from the plate. This detachment of the cell is not observed if I add TET to 293 cells or 293-Ag cells. Therefore, this effect is likely caused by HDV replication. Altogether, these observations suggest that the cells are undergoing a stress caused by the viral replication, perhaps a cellular replication stress since they are arrested in G1/G0 (Zeman and Cimprich, 2014). Often, in cells undergoing a stress stimulus, there is formation of cytoplasmic stress granules. Since the aspect of stress granules are very similar to the aspect of PSP1 delocalizing from the nucleus upon HDV replication, and because these cells may be under a stress response, I investigated whether PSP1 was delocalized to stress granules. I found that PSP1 greatly colocalize with the PABP, a stress granule marker (Buchan and Parker, 2009; Kedersha et al., 1999). PABP is an RNA binding protein which binds the polyA tail of the cellular mRNA, is involved in the polyadenylation, in the protection of the RNA from degradation, in the shuttling of mRNA between the nucleus and the cytoplasm and in non-sense mediated decay as well as in the initiation of translation (Adam et al., 1986; Burd et al., 1991; Eliseeva et al., 2013; Matunis et al., 1993; Query et al., 1989). PABP is usually located in association with the polyA tail of mRNA in the cytoplasm and a subset of PABP cytoplasmic pool is recruited to the stress granules in stressed cells (Kedersha et al., 1999; Goss and Kleiman, 2013; Grey et al., 2015). The cytoplasmic foci of PABP that I visualized following its immunostaining is similar to what is described for stress granules, a cytoplasmic structure composed of RNA and protein, forming upon a stress stimulus (Buchan and Parker, 2009). Stress granules contain non-translating mRNAs and factors associated with the stability of mRNA (Buchan and Parker, 2009).

The paraspeckle marker PSP1 is not essential for the formation of paraspeckles since the knock-out of PSP1 protein does not lead to their disassembly (Sasaki and Hirose, 2009). NEAT1 is a lnc RNA that was shown to be essential for the formation of paraspeckles (Clemson et al., 2009; Hirose et al., 2014). I found that nuclear NEAT1 foci were still present upon HDV replication, suggesting that paraspeckles are still present. However, since PSP1 is delocalized from the paraspeckles, this suggests that they are altered upon HDV replication. It was recently reported that the IAV also induces an increase in size of the NEAT1 foci and an increase of IL8 levels caused by a release of the delocalization of the transcriptional repressor PSF from the host gene to the paraspeckles (Imamura et al., 2014). Similarly, we showed that HDV replication induces an increase in the size of NEAT1 foci and in levels of NEAT1 and IL8 mRNA.

My results are also consistent with previous reports suggesting that paraspeckles might be part of a cellular response to stress events such as viral infections (Prasanth et al., 2005; Sunwoo, 2009 ; Nakagawa and Hirose, 2012; Zhang et al., 2013a; Hirose et al., 2014). Furthermore, the study by T. Hirose and colleagues shows that the proteasome inhibition, which is a mechanism involved in the stress response of the cell, leads to the expression of a higher level of NEAT1 transcripts. This observed increase in NEAT1 results in a bulge of the paraspeckles foci, which allowed for recruitment of paraspeckle proteins (Clemson et al., 2009; Hirose et al., 2014; Sasaki and Hirose, 2009). In our present study, upon HDV replication, we observed both an enlargement of Neat foci and an up-regulation of NEAT1 transcripts. In addition, this bulge of the Neat foci previously described in stressed cells further support the hypothesis mentioned earlier (in 4.3.1), that PSP1 might colocalize with PABP in stress granules.

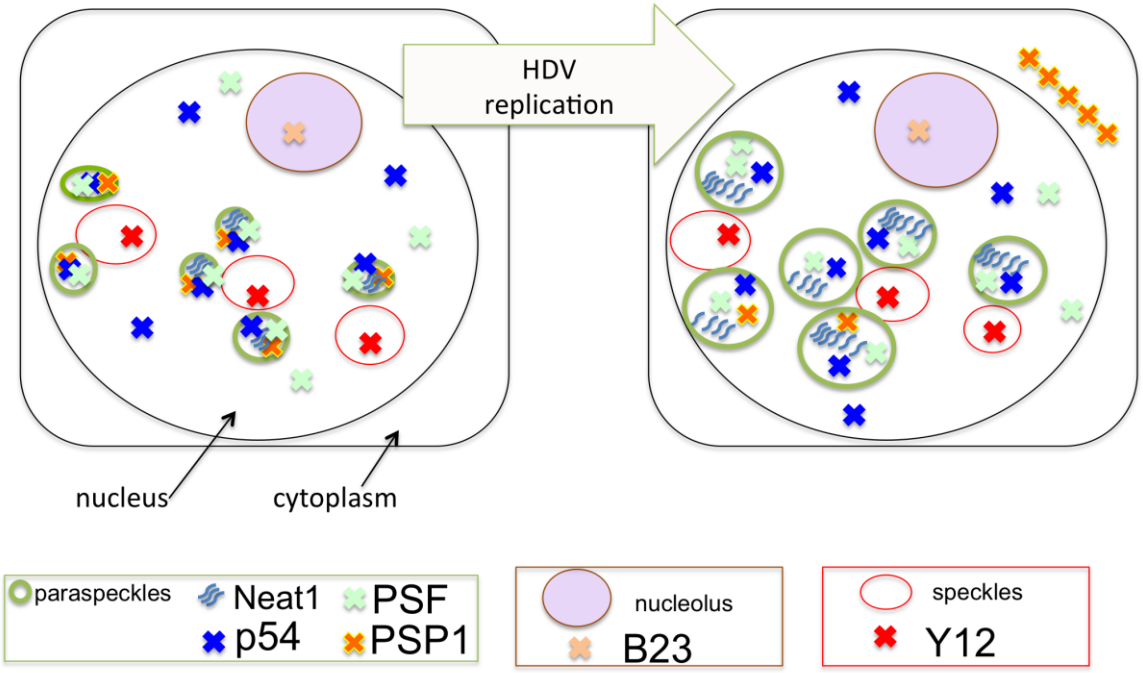

HDV
replication
nucleus
cytoplasm
paraspeckles
Neat1
PSF
p54
PSP1
nucleolus
B23
speckles
Y12

Figure 5.2 Model of the paraspeckles alteration in 293 cells replicating HDV.

In cells replicating HDV, PSP1 is delocalized outside of the nucleus. The other paraspeckle markers PSF, p54 and NEAT1 remain in the nuclear paraspeckle. The NEAT1 RNA transcription levels are increased and as a result, the NEAT1 foci appear brighter and bigger.

5.4.2 Possible link between PSF, NEAT1 and the development of cancer

PSF has previously been reported to regulate NEAT1 transcription (Imamura et al., 2014). My study suggests that upon HDV replication there is an upregulation of NEAT1. In a recent report, NEAT1 regulation has been associated with the regulation of oncogenes for the lung, liver, esophageal and retinal tissues (Choudhry et al., 2015; Gernapudi et al., 2016; Gibb et al., 2011; Li et al., 2015). Interestingly, it has also been shown that high levels of NEAT1 correlated with hepatocellular carcinoma tissues of patients (Guo et al., 2015). Therefore, in the future it will be worth investigating whether our results showing that NEAT1 is upregulated upon HDV replication may partly explain the development of hepatocellular carcinoma in HDV infected patients. In addition to the regulation of NEAT1, PSF and p54 have been linked to cancer in many reports (Benn et al., 2000; Clark et al., 1997; Mathur et al., 2003; Buxade et al., 2008; Figueroa et al., 2009a; Figueroa et al., 2009b; Galietta et al., 2007; Lukong et al., 2009; Miyamoto et al., 2008; Song et al., 2004; Song et al., 2005). The mutation of PSF has been linked to cervical cancer and papillary renal cancer tissues (Benn et al., 2000; Clark et al., 1997; Mathur et al., 2003). Further examples include a study showing a connexion between PSF and Hakai, a ubiquitin ligase involved in the mechanism of down-regulation of the cell to cell contact, which promotes the progression to cancerous cells (Figueroa et al., 2009b). The knockdown of PSF prevented the cellular over-growth mediated by Hakai. Upon overexpression of Hakai, PSF interacts with transcripts encoding tumour inducing proteins (Figueroa et al., 2009b). The authors proposed that Hakai probably indirectly promotes the phosphorylation of PSF, which upsets PSF ability to bind DNA and to repress oncogenes, and enhances PSF ability to bind and stabilize cellular

oncogenes mRNA through its RRM domain (Figueroa et al., 2009b). In addition, PSF has been shown to regulate other oncogenes upon viral or RNA pathogens replication. Interestingly, an up-regulation of several cellular genes and oncogenes is observed during the interaction of PSF with a small mouse RNA transposon, VL30. Song and al have shown that the DBD of PSF was binding to the oncogene GAGE 6 promoter while RRM of PSF was binding to VL30 RNA (Song et al., 2004; Song et al., 2005). Since PSF is a transcriptional repressor acting as a tumour suppressor, it is tempting to speculate that the HDV perturbs the interaction of PSF with cellular promoters and leads to an up regulation of oncogenes. Further experiments will allow for shedding light on which oncogenes may be deregulated upon HDV replication. An upregulation of oncogenes in HDV infected hepatocytes may explain the carcinogenous effects leading to fulminous hepatocellular carcinomas (Romeo et al., 2009). However, since PSF has many other cellular roles such as the retention of hyperedited RNA or the localization and shuttling of RNA, HDV titration of PSF proteins may have other consequences on the cell fate, such as a cytopathic effect. This would also be in accordance with the pathogenicity previously described as the cytoxicity of HDV leading to fulminant cirrhosis (Cole et al., 1991; Fattovich et al., 1987). Future experiments will be important to determine what genes are up-regulated following the interaction of HDV with PSF.

5.4.3 Cellular localization of HDV RNA and interaction of HDV with the host cellular components

Previous studies identified that the HDV foci composed of HDAg and HDV genomes are localized close to speckles (Bichko and Taylor, 1996; Cunha et al., 1998). At the time of this study, the paraspeckles were not characterized yet. The two studies using FISH against

HDV RNA suggest that the HDV foci do not seem to be the sites of replication or transcription (Bichko and Taylor, 1996; Cunha et al., 1998).

We report that upon HDV replication in the 293 cells, the paraspeckle structures seem disrupted and that a paraspeckle component, PSP1 is delocalized outside of the nucleus. One possible explanation would be that the HDV foci containing HDAg and HDV genomes disrupt the paraspeckles by recruiting its components. Another explanation is that the HDV foci colocalize with the paraspeckles. PSP1 has a NLS and is usually present in the nucleoplasm and concentrated in paraspeckles (Fox et al., 2005). A more recent study showed that importin *alpha* 2 allows for PSP1 transport from the cytoplasm to the nucleus (Major et al., 2015). I showed that upon HDV replication, PSP1 is exported to the cytoplasm, but that PSF remains in the nucleus. These results are similar to the observations of A. T. Major which show that changes in the PSP1 traffic did not influence the localization of the PSF foci and suggested that PSP1 is not required as a scaffold for the paraspeckles (Major et al., 2015). Previous studies reported that HDAg NLS allows for targeting of HDRNP to the nucleus via importin *alpha* 2 (Alves et al., 2008; Chou et al., 1998). One possibility would be that the HDV RNA that enters the cell with the HDAg-S hijacks the importin *alpha* 2, thereby inhibiting PSP1 to use importin *alpha* 2. Another possibility would be that HDV RNA forces PSP1 transport to the cytoplasm.

The M. Carmo-Fonseca team suggested that the RNA genome might be responsible for the export of HDRNP to the cytoplasm (Tavanez et al., 2002). Since the HDRNP do not seem to accumulate in the cytoplasm following passive diffusion, the authors speculate that one possibility might be that the binding of the HDV RNA to the HDAg triggers the exposure of a NES of the HDAg or that the HDV RNA itself has a cytoplasmic signalization signal. It is known for other viruses that the viral RNA may interact with the cellular host

machinery to promote shuttling between the cytoplasm and the nucleus. For instance, for the retrovirus type 1, the folding of the RNA in a hairpin loop constitutes the CTE signal, which allows the interaction with host protein TAP (Tavanez et al., 2002). Since the HDV RNA genome is highly structured, it will be worth investigating whether the RNA structure itself constitutes a signal for nuclear export.

5.4.4 Future directions

5.4.4.1 Interaction of PSP1 and HDV RNA

In my present study, I found that upon HDV replication, PSP1 associates with HDV genome and that PSP1 is delocalized to the cytoplasm. Future work will allow for identifying the exact binding sites of PSP1 on HDV RNA. It also remains to be clarified whether PSP1 is co-transported with HDV RNA and whether the viral RNA is responsible of this export of the cellular PSP1. For instance, the use of FISH in a time point assay to precisely locate HDV genome in its cellular host will allow localizing HDV at different steps of the replication, and will help identifying at which step of HDV replication PSP1 is exported. This could be combined with the immunostaining of markers of the nuclear membrane to clarify whether PSP1 induces the export of HDV genome or if PSP1 is indirectly exported out of the nucleus. It will be worth elucidating the link between the importin *alpha* 2, HDV replication and PSP1.

5.4.4.2.Involvement of host proteins in HDV replication

PSF and p54 facilitate transcriptional regulation and help build the paraspeckle structures in the host cell (Cao et al., 2015). My present work supports the hypothesis that the three paraspeckle proteins PSF, p54 and PSP1 are linked to HDV replication. Future *in vitro* transcription assays will provide further clues on the requirements of p54 and PSP1 for HDV genome replication, as previously established for PSF (Sikora, 2012 ; Zhang, 2013). Further

experiments such as immunoprecipitations of HDV mutants with p54 and PSP1 or *in vitro* transcription assays with nuclear extracts immunodepleted of p54 and PSP1 will shed light on whether p54 and PSP1 have a direct role in the use of RNAP II by HDV for its replication. It will be worth investigating what effect the knocking down of NEAT1 has on HDV levels. In the case of HIV, the silencing of NEAT1 results in higher levels of HIV-1 containing INS elements RNAs (Barichievy et al., 2015; Zhang et al., 2013a). Future endeavors will allow placing the current results in the mechanistic pathways including paraspeckles and the regulation of the host transcriptional pathway during HDV replication. It will worth further testing whether PSF is recruited to the paraspeckles and induces a derepression of the host cellular genes for HDV, as it is the case for Influenza A Virus (Imamura et al., 2014).

In order to further explore the link between the paraspeckles and HDV replication, it will be of interest to study the interplay of the 33 recently identified paraspeckles components in HDV replication (Fong et al., 2013; Nakagawa and Hirose, 2012; Quiskamp et al., 2014; Yamazaki and Hirose, 2015). The paraspeckles also contain human lnc RNAs edited by the host enzyme ADAR (Chen and Carmichael, 2009; Prasanth et al., 2005). A recent study showed that PSF promotes ADARB2 transcription and that the recruitment of PSF to paraspeckles resulted in a decrease of ADARB2 expression (Hirose et al., 2014). These studies suggest that paraspeckles have a role in regulating the cellular RNA editing processes. ADAR is responsible for HDV mRNA editing (Hartwig et al., 2006; Wong and Lazinski, 2002). As paraspeckles are localized close to the speckles and HDV genomic RNA has also been localized close to the speckles (Cunha et al., 1998; Fox et al., 2002), it would be worth investigating the implication of paraspeckles ADAR in HDV RNA editing.

5.4.4.3. Effect of HDV replication on the paraspeckle proteins mediated immunity

For the EBV, it has been proposed that the interaction of p54 with a highly structured EBV RNA is an innate antiviral immune pathway mediated by the paraspeckles (Cao et al., 2015). I would expect that the titration of PSF by HDV could trigger the innate response; further work will be necessary to test this hypothesis. This would be consistent with the fact that HDV replication caused a cytopathic effect in cells (Wang et al., 2001). We show an increase in IL8 expression upon HDV replication. Future work will be required to test whether mRNA of other cytokines reported to be increased following HDV replication, such as interleukin 2, interleukine 10, interferon gamma and Tumour Necrosis factor *alpha*, are bound by PSF (Ataseven et al., 2006; Grabowski et al., 2011; Nisini et al., 1997; Park et al., 2009; Pugnale et al., 2009). Previous studies showed that PSF and p54 are involved in cellular immunity (Buxade et al., 2008). For instance, PSF and p54 had a high affinity for RNAs with AU-rich elements, including pro-inflammatory cytokine mRNAs (Buxade et al., 2008). Mitogen Activated Protein-kinases phosphorylate PSF, which increases PSF affinity to Tumor Necrosis factor *alpha* mRNA (Buxade et al., 2008). It will be useful to test whether HDV RNA binds with stronger affinity to the paraspeckle proteins, thereby releasing these proteins from the mRNA encoding proinflammatory cytokines.

5.5 Global significance of the research

5.5.1 Role of RNA genome structure

Since R199G is an RNA transcribed by the human RNAP II, it is likely to use both the same polymerase and cellular environment of potential human RNA templates for RNAP II. Therefore, one of the most exciting prospects provided by the HDV promoter model presented in our study is to find reminiscent features in other cellular RNAs in order to

identify cellular RNA templates for RNAP II. In the past decade, Molecular Biology have been marked by the important discoveries of human lnc RNAs, human endo retroviruses and circular RNAs (Chen and Carmichael, 2009; Lasda and Parker, 2014; Mercer et al., 2009; Morris and Mattick, 2014; Quinn and Chang, 2016). It will be of interest to screen among these RNAs for candidates for human RNA that could use the human RNAP II. One first piece of evidence towards the hypothesis that human RNAs can be used by RNAP II was provided by the team of J A Goodrich: they showed that the B2 mammalian non-coding RNA, which is transcribed from Short Interspersed Elements, could be used as a template by RNAP II *in vitro* and the RNA was extended at its 3' end (Wagner et al., 2013). In addition, based on the similarity with HDV, which also has a circular RNA genome and replicates via RNAP II, it is tempting to hypothesize that the circular RNAs are potential human RNA that are able to use the RNAP II (Taylor, 2014). The work of Zhang and colleagues on the ci-ankrd52 RNA showed that circular RNAs can associate with RNAP II complexes to act as transcriptional regulators (Zhang et al., 2013b). However, much work remains to be done to elucidate the function of theses circular RNAs, as well as the importance of the RNA structure in their function (Hansen T B et al., 2013; Lasda and Parker, 2014; Memczak et al., 2013; Salzman et al., 2012). In order to scan the human chromosomes and detect putative RNAs that could be templates for RNAP II, the model of HDV promoter presented in this thesis could be used to search for similar RNA structural motifs in RNA databases (Burge et al., 2013; Griffiths-Jones et al., 2003).

In this study, I identified the features conserved among HDV sequences subject to the replicative selective pressure. I localized regions with higher levels of covariation and regions where the primary structure was conserved, and overall these data show a pressure to preserve the stem-loop structure of R199G. This finding is a contribution to the field of HDV

replication research since it provides a model of an HDV RNA promoter and therefore brings insights about requirements that allow for HDV to replicate its genome.

Since the number of methods available to study the secondary structure of an RNA from high-throughput data were very limited, we have developed a heuristic method to use deep sequencing data for the analysis of the RNA secondary structure (Beeharry et al., 2014; Westhof, 2015). This approach can also be used to analyze RNA structural features from SELEX and SHAPE data (Lucks et al., 2011; Spitale et al., 2014). I have herein described the variability rate of HDV R199G region and found a high mutation rate of $8x10^{-4}$ mutation/site, reminiscent of the high mutation rate of RNA viruses (Domingo, 2015).

This bio-informatics approach will also be beneficial for the study of the secondary structure of other RNA sequence populations, such as human RNA pathogenic viruses, especially since there will be much work to be done, as it is estimated that there are at least 320 000 mammalian viruses that have yet to be sequenced (Anthony et al., 2013; Marz et al., 2014). Since viroids are highly similar and belong to the same circular RNA group as HDV, our bio-informatics pipeline will also be of significant importance to further characterize the RNA structure of viroids, plant pathogens which cause severe agricultural damage (Flores et al., 2011). Moreover, the better characterization of the RNA structure is important to better understand its functional activity such as RNA mimicry to hijack the host machinery (Bojic et al., 2012; Colussi et al., 2014; Flores et al., 2011; Lasda and Parker, 2014, 2016). Better overall understanding of the RNA structure is crucial to better counteract RNA pathogens.

5.5.2 Role of paraspeckle components in HDV replication

Previously, PSF potential role in HDV replication was identified and it was shown that p54 binds R199G. But this was only the tip of the iceberg for the involvement of the DBHS

proteins involved in the HDV viral cycle. Indeed, my present work shows that PSP1 also interacts with R199G, and furthermore that the knockdown of PSF, p54 and PSP1 leads to a dramatic decrease of HDV RNA levels in 239 cells. This study provides a foundation for further research in order to test whether the DBHS proteins are specifically involved in the recognition of R199G by RNAP II. Better characterizing the function of PSF, p54 and PSP1 in the viral replication cycle and understanding what RNA structural properties allows HDV to hijack these host proteins will help better targeting and impeding HDV take-over of the cellular machinery. HDV is a virus responsible of high morbidity and still no specific treatment is available. Currently, for patients severely infected, the only therapeutics to constrain HDV infection remains the liver transplantation (Wedemeyer and Manns, 2010). It is therefore urgent to better characterize the host proteins involved in HDV replication cycle in order to design specific anti-viral therapeutic and predict their consequence on the host proteins.

The better characterization of the paraspeckles disruption and of the delocalization of PSP1 will help better understanding the consequence on HDV replication. In the coming years, the exploration of paraspeckle proteins functions will allow better characterizing what are the cytokines secreted upon HDV replication and provide a foundation to further study their prevalence as part of an innate immunity pathway against HDV and to which extent this cellular pathway is hijacked by HDV for its replication. While identifying the host components that interact with the virus, a better understanding of the hijacked host cellular factors and pathways is granted (Gaglia and Glaunsinger, 2010; Lai, 1998). For example, the interactions between transcriptions factors and nuclear lnc RNAs composing paraspeckles remain to be characterized thoroughly.

5.6 General conclusion

HDV is a satellite RNA virus that is able to heavily hijack its host cell machinery and to take advantage of HBV envelope proteins in order to ensure its viral cycle. Surprisingly, HDV is even able to usurp the human DNA-dependent RNA polymerase II. However, the RNA features that allow for this unconventional use of RNAP II with an RNA template instead of a DNA template remained to be elucidated. A previous study in the laboratory identified that R199G, the right terminal region of HDV of 199 nucleotides, was a promoter for HDV and suggested the importance of its RNA structure. I investigated the selected promoter features that are likely important for HDV replication. My results suggest importance of the secondary structure of HDV in the region upstream from the tip of R199G and of the primary sequence at the tip of R199G. I also brought insights into the variability rates of an HDV promoter region. Based on the selected features identified, I proposed a model of an RNA promoter, which might be useful for identifying other RNA promoters for RNAP II. My analysis is also supported by previous *in vitro* assays performed in the laboratory. These features are likely important for the binding of R199G to host proteins.

Previously, the laboratory identified that PSF and p54 bind R199G and that PSF is required for RNAP II acting on R199G in *in vitro* transcription assays. One of the novel findings of this work was that R199G associates to the paraspeckle protein PSP1, another member of the DBHS family of proteins. I also demonstrated the requirement of PSF, p54 and PSP1 proteins for HDV RNA accumulation. Furthermore, I showed that there is an alteration of the paraspeckles during HDV replication in 293 cells.

Overall, I characterized the RNA structure and investigated the requirement of several host proteins for viral replication, and provided evidence of an effect of HDV replication on

the host paraspeckles. The PhD work presented in my thesis contributes to knowledge in the field of the Hepatitis *Delta* Virus life cycle, and also to basic knowledge on cellular processes such as the regulation of RNA metabolism.

References

Abbas, Z., Abbas, M., Abbas, S., and Shazi, L. (2015). Hepatitis D and hepatocellular carcinoma. World journal of hepatology *7*, 777-786.

Abrahem, A., and 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 Res *36*, 5201-5211.

Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K., and Dreyfuss, G. (1986). mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Molecular and cellular biology *6*, 2932-2943.

Akhmedov, A.T., Bertrand, P., Corteggiani, E., and Lopez, B.S. (1995). Characterization of two nuclear mammalian homologous DNA-pairing activities that do not require associated exonuclease activity. Proc Natl Acad Sci U S A *92*, 1729-1733.

Akhmedov, A.T., and Lopez, B.S. (2000). Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. Nucleic Acids Res *28*, 3022-3030.

Allain, F.H., Gubser, C.C., Howe, P.W., Nagai, K., Neuhaus, D., and Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. Nature *380*, 646-650.

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

Amelio, A.L., Miraglia, L.J., Conkright, J.J., Mercer, B.A., Batalov, S., Cavett, V., Orth, A.P., Busby, J., Hogenesch, J.B., and Conkright, M.D. (2007). A coactivator trap identifies NONO (p54nrb) as a component of the cAMP-signaling pathway. Proc Natl Acad Sci U S A *104*, 20314-20319.

Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. Current biology : CB *12*, 1-11.

Anglana, M., Apiou, F., Bensimon, A., and Debatisse, M. (2003). Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. Cell *114*, 385-394.

Anisimova, M., and Yang, Z. (2004). Molecular evolution of the hepatitis delta virus antigen gene: recombination or positive selection? Journal of molecular evolution *59*, 815-826.

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A *90*, 11995-11999.

Anthony, S.J., Epstein, J.H., Murray, K.A., Navarrete-Macias, I., Zambrana-Torrelio, C.M., Solovyov, A., Ojeda-Flores, R., Arrigo, N.C., Islam, A., Ali Khan, S.*, et al.* (2013). A strategy to estimate unknown viral diversity in mammals. MBio *4*, e00598-00513.

Ataseven, H., Bahcecioglu, I.H., Kuzu, N., Yalniz, M., Celebi, S., Erensoy, A., and Ustundag, B. (2006). The levels of ghrelin, leptin, TNF-alpha, and IL-6 in liver cirrhosis and hepatocellular carcinoma due to HBV and HDV infection. Mediators of inflammation *2006*, 78380.

Auboeuf, D., Dowhan, D.H., Li, X., Larkin, K., Ko, L., Berget, S.M., and O'Malley, B.W. (2004). CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. Molecular and cellular biology *24*, 442-453.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M.*, et al.* (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell *46*, 674-690.

Barichievy, S., Naidoo, J., and Mhlanga, M.M. (2015). Non-coding RNAs and HIV: viral manipulation of host dark matter to shape the cellular environment. Frontiers in genetics *6*, 108.

Barker, M.M., and Gourse, R.L. (2001). Regulation of rRNA transcription correlates with nucleoside triphosphate sensing. Journal of bacteriology *183*, 6315-6323.

Basu, A., Dong, B., Krainer, A.R., and Howe, C.C. (1997). The intracisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54nrb/NonO. Molecular and cellular biology *17*, 677-686.

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

Beeharry, Y., and Pelchat, M. (2011). Subversion of RNA Processing Pathways by the Hepatitis delta Virus (INTECH Open Access Publisher).

Beeharry, Y., Rocheleau, L., and Pelchat, M. (2014). Conserved features of an RNA promoter for RNA polymerase II determined from sequence heterogeneity of a hepatitis delta virus population. Virology *450-451*, 165-173.

Beerenwinkel, N., Gunthard, H.F., Roth, V., and Metzner, K.J. (2012). Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Frontiers in microbiology *3*, 329.

Bell, P., Brazas, R., Ganem, D., and Maul, G.G. (2000). Hepatitis delta virus replication generates complexes of large hepatitis delta antigen and antigenomic RNA that affiliate with and alter nuclear domain 10. J Virol *74*, 5329-5336.

Benn, D.E., Dwight, T., Richardson, A.L., Delbridge, L., Bambach, C.P., Stowasser, M., Gordon, R.D., Marsh, D.J., and Robinson, B.G. (2000). Sporadic and familial pheochromocytomas are associated with loss of at least two discrete intervals on chromosome 1p. Cancer research 60, 7048-7051.

Bezy O, Elabd C, Cochet O, Petersen RK, Kristiansen K, Dani C, Ailhaud G, Amri EZ. (2005). Delta-interacting protein A, a new inhibitory partner of CCAAT/enhancer-binding protein beta, implicated in adipocyte differentiation. J Biol Chem 25;280(12):11432-8.

Bhogal, R.H., Hodson, J., Bartlett, D.C., Weston, C.J., Curbishley, S.M., Haughton, E., Williams, K.T., Reynolds, G.M., Newsome, P.N., Adams, D.H.*, et al.* (2011). Isolation of primary human hepatocytes from normal and diseased liver tissue: a one hundred liver experience. PLoS One *6*, e18222.

Bichko, V.V., and Taylor, J.M. (1996). Redistribution of the delta antigens in cells replicating the genome of hepatitis delta virus. J Virol *70*, 8064-8070.

Biebricher, C.K., and Orgel, L.E. (1973). An RNA that multiplies indefinitely with DNA-dependent RNA polymerase: selection from a random copolymer. Proc Natl Acad Sci U S A *70*, 934-938.

Bjorklund, S., and Gustafsson, C.M. (2005). The yeast Mediator complex and its regulation. Trends in biochemical sciences *30*, 240-244.

Bladen, C.L., Udayakumar, D., Takeda, Y., and Dynan, W.S. (2005). Identification of the polypyrimidine tract binding protein-associated splicing factor.p54(nrb) complex as a candidate DNA double-strand break rejoining factor. The Journal of biological chemistry *280*, 5205-5210.

Blake, R.D., Hess, S.T., and Nicholson-Tuell, J. (1992). The influence of nearest neighbors on the rate and pattern of spontaneous point mutations. Journal of molecular evolution *34*, 189-200.

Bojic, T., Beeharry, Y., Zhang da, J., and Pelchat, M. (2012). Tomato RNA polymerase II interacts with the rod-like conformation of the left terminal domain of the potato spindle tuber viroid positive RNA genome. The Journal of general virology *93*, 1591-1600.

Bond, C.S., and Fox, A.H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. The Journal of cell biology *186*, 637-644.

Bonino, F., Hoyer, B., Shih, J.W., Rizzetto, M., Purcell, R.H., and Gerin, J.L. (1984). Delta hepatitis agent: structural and antigenic properties of the delta-associated particle. Infection and immunity *43*, 1000-1005.

Brazas, R., and Ganem, D. (1996). A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. Science *274*, 90-94.

Brown, S.A., Ripperger, J., Kadener, S., Fleury-Olela, F., Vilbois, F., Rosbash, M., and Schibler, U. (2005). PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. Science *308*, 693-696.

Bruelle, C., Bedard, M., Blier, S., Gauthier, M., Traish, A.M., and Vincent, M. (2011). The mitotic phosphorylation of p54(nrb) modulates its RNA binding activity. Biochemistry and cell biology = Biochimie et biologie cellulaire *89*, 423-433.

Buchan, J.R., and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. Mol Cell *36*, 932-941.

Burd, C.G., Matunis, E.L., and Dreyfuss, G. (1991). The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. Molecular and cellular biology *11*, 3419-3424.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., and Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. Nucleic Acids Res *41*, D226-232.

Bushnell, D.A., Cramer, P., and Kornberg, R.D. (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. Proc Natl Acad Sci U S A *99*, 1218-1222.

Buske, F.A., Boden, M., Bauer, D.C., and Bailey, T.L. (2010). Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics *26*, 860-866.

Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev *16*, 2583-2592.

Buxade, M., Morrice, N., Krebs, D.L., and Proud, C.G. (2008). The PSF.p54nrb complex is a novel Mnk substrate that binds the mRNA for tumor necrosis factor alpha. The Journal of biological chemistry *283*, 57-65.

Cadena, D.L., and Dahmus, M.E. (1987). Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. The Journal of biological chemistry *262*, 12468-12474.

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

Cao, S., Moss, W., O'Grady, T., Concha, M., Strong, M.J., Wang, X., Yu, Y., Baddoo, M., Zhang, K., Fewell, C.*, et al.* (2015). New Noncoding Lytic Transcripts Derived from the Epstein-Barr Virus Latency Origin of Replication, oriP, Are Hyperedited, Bind the Paraspeckle Protein, NONO/p54nrb, and Support Viral Lytic Transcription. J Virol 89, 7120-7132.

Cardinale, S., Cisterna, B., Bonetti, P., Aringhieri, C., Biggiogera, M., and Barabino, S.M. (2007). Subnuclear localization and dynamics of the Pre-mRNA 3' end processing factor mammalian cleavage factor I 68-kDa subunit. Mol Biol Cell 18, 1282-1292.

Carey V. *parody: Parametric And Resistant Outlier DYtection*. R package version 1.24.0.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C.*, et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nature genetics *38*, 626-635.

Carpenter, C.D., and Simon, A.E. (1998). Analysis of sequences and predicted structures required for viral satellite RNA accumulation by in vivo genetic selection. Nucleic Acids Res *26*, 2426-2432.

Casey, J.L., Brown, T.L., Colan, E.J., Wignall, F.S., and Gerin, J.L. (1993). A genotype of hepatitis D virus that occurs in northern South America. Proc Natl Acad Sci U S A *90*, 9016-9020.

Casey, J.L., Tennant, B.C., and Gerin, J.L. (2006). Genetic changes in hepatitis delta virus from acutely and chronically infected woodchucks. J Virol *80*, 6469-6477.

Cayrol, C., and Flemington, E.K. (1996). The Epstein-Barr virus bZIP transcription factor Zta causes G0/G1 cell cycle arrest through induction of cyclin-dependent kinase inhibitors. The EMBO journal *15*, 2748-2759.

Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Grey, M.W., and Cedergren, R. (1997). On the evolution of the single-subunit RNA polymerases. Journal of molecular evolution *45*, 671-681.

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

Chang, J., Moraleda, G., and Taylor, J. (2000). Limitations to replication of hepatitis delta virus in avian cells. J Virol *74*, 8861-8866.

Chang, J., Nie, X., Chang, H.E., Han, Z., and Taylor, J. (2008). Transcription of hepatitis delta virus RNA by RNA polymerase II. J Virol *82*, 1118-1127.

Chang, J., and Taylor, J. (2002). In vivo RNA-directed transcription, with template switching, by a mammalian RNA polymerase. The EMBO journal *21*, 157-164.

Chang, M.F., Baker, S.C., Soe, L.H., Kamahora, T., Keck, J.G., Makino, S., Govindarajan, S., and Lai, M.M. (1988). Human hepatitis delta antigen is a nuclear phosphoprotein with RNA-binding activity. J Virol *62*, 2403-2410.

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

Chao, M., Hsieh, S.Y., and Taylor, J. (1990). Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. J Virol *64*, 5066-5069.

Chao, M., Wang, T.C., and Lee, S.E. (2006). Detection of hepatitis delta virus recombinants in cultured cells co-transfected with cloned genotypes I and IIb DNA sequences. Journal of virological methods *137*, 252-258.

Chen, C.W., Tsay, Y.G., Wu, H.L., Lee, C.H., Chen, D.S., and 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*, 33058-33067.

Chen, L.L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell *35*, 467-478.

Chen, P.J., Kalpana, G., Goldberg, J., Mason, W., Werner, B., Gerin, J., and Taylor, J. (1986). Structure and replication of the genome of the hepatitis delta virus. Proc Natl Acad Sci U S A *83*, 8774-8778.

Choi, S.H., Park, K.J., and Hwang, S.B. (2002). Large hepatitis delta antigen is phosphorylated at multiple sites and phosphorylation is associated with protein conformational change. Intervirology *45*, 142-149.

Chou, H.C., Hsieh, T.Y., Sheu, G.T., and Lai, M.M. (1998). Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. J Virol *72*, 3684-3690.

Choudhry, H., Albukhari, A., Morotti, M., Haider, S., Moralli, D., Smythies, J., Schodel, J., Green, C.M., Camps, C., Buffa, F.*, et al.* (2015). Tumor hypoxia induces nuclear paraspeckle formation through HIF-2alpha dependent transcriptional activation of NEAT1 leading to cancer cell survival. Oncogene *34*, 4482-4490.

Clark, J., Lu, Y.J., Sidhar, S.K., Parker, C., Gill, S., Smedley, D., Hamoudi, R., Linehan, W.M., Shipley, J., and Cooper, C.S. (1997). Fusion of splicing factor genes PSF and NonO (p54nrb) to the TFE3 gene in papillary renal cell carcinoma. Oncogene *15*, 2233-2239.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell *33*, 717-726.

Clemson, C.M., Tzekov, R., Krebs, M., Checchi, J.M., Bigelow, C., and Kaushal, S. (2011). Therapeutic potential of valproic acid for retinitis pigmentosa. The British journal of ophthalmology *95*, 89-93.

Cole, S.M., Gowans, E.J., Macnaughton, T.B., Hall, P.D., and Burrell, C.J. (1991). Direct evidence for cytotoxicity associated with expression of hepatitis delta virus antigen. Hepatology *13*, 845-851.

Colussi, T.M., Costantino, D.A., Hammond, J.A., Ruehle, G.M., Nix, J.C., and Kieft, J.S. (2014). The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA. Nature *511*, 366-369.

Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A *82*, 7934-7938.

Cramer, P. (2004). Structure and function of RNA polymerase II. Advances in protein chemistry *67*, 1-42.

Cullen, J.M., David, C., Wang, J.G., Becherer, P., and Lemon, S.M. (1995). Subcellular distribution of large and small hepatitis delta antigen in hepatocytes of hepatitis delta virus superinfected woodchucks. Hepatology *22*, 1090-1100.

Cunha, C., Monjardino, J., Cheng, D., Krause, S., and Carmo-Fonseca, M. (1998). Localization of hepatitis delta virus RNA in the nucleus of human cells. Rna *4*, 680-693.

Dandri, M., Burda, M.R., Gocht, A., Torok, E., Pollok, J.M., Rogler, C.E., Will, H., and Petersen, J. (2001). Woodchuck hepatocytes remain permissive for hepadnavirus infection and mouse liver repopulation after cryopreservation. Hepatology *34*, 824-833.

Davies, J.W., Kaesberg, P., and Diener, T.O. (1974). Potato spindle tuber viroid. XII. An investigation of viroid RNA as a messenger for protein synthesis. Virology *61*, 281-286.

DeCerbo, J., and Carmichael, G.G. (2005). Retention and repression: fates of hyperedited RNAs in the nucleus. Current opinion in cell biology *17*, 302-308.

Delgado, S., Martinez de Alba, A.E., Hernandez, C., and Flores, R. (2005). A short double-stranded RNA motif of Peach latent mosaic viroid contains the initiation and the self-cleavage sites of both polarity strands. J Virol *79*, 12934-12943.

Deny, P. (2006). Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? Current topics in microbiology and immunology *307*, 151-171.

Deny, P., Zignego, A.L., Rascalou, N., Ponzetto, A., Tiollais, P., and Brechot, C. (1991). Nucleotide sequence analysis of three different hepatitis delta viruses isolated from a woodchuck and humans. The Journal of general virology 72 (Pt 3), 735-739.

Du X, Wang Q, Hirohashi Y, Greene MI. DIPA, which can localize to the centrosome, associates with p78/MCRS1/MSP58 and acts as a repressor of gene transcription. (2006). Exp Mol Pathol. 81(3):184-90.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., and Barabino, S.M. (2004). Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. The Journal of biological chemistry *279*, 35788-35797.

Diener, T.O. (1974). Viroids: the smallest known agents of infectious disease. Annual review of microbiology *28*, 23-39.

Dingle, K., Moraleda, G., Bichko, V., and Taylor, J. (1998). Electrophoretic analysis of the ribonucleoproteins of hepatitis delta virus. Journal of virological methods *75*, 199-204.

Domingo, E. (2015). Virus as Populations: Composition, Complexity, Dynamics, and Biological Implications (Elsevier Science).

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. Microbiology and molecular biology reviews : MMBR *76*, 159-216.

Dong, B., Horowitz, D.S., Kobayashi, R., and 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 Res *21*, 4085-4092.

Dong, X., Shylnova, O., Challis, J.R., and Lye, S.J. (2005). Identification and characterization of the protein-associated splicing factor as a negative co-regulator of the progesterone receptor. The Journal of biological chemistry *280*, 13329-13340.

Dong, X., Sweet, J., Challis, J.R., Brown, T., and Lye, S.J. (2007). Transcriptional activity of androgen receptor is modulated by two RNA splicing factors, PSF and p54nrb. Molecular and cellular biology *27*, 4863-4875.

Dourakis, S., Karayiannis, P., Goldin, R., Taylor, M., Monjardino, J., and Thomas, H.C. (1991). An in situ hybridization, molecular biological and immunohistochemical study of hepatitis delta virus in woodchucks. Hepatology *14*, 534-539.

Duong, H.A., Robles, M.S., Knutti, D., and Weitz, C.J. (2011). A molecular mechanism for circadian clock negative feedback. Science *332*, 1436-1439.

Eliseeva, I.A., Lyabin, D.N., and Ovchinnikov, L.P. (2013). Poly(A)-binding proteins: structure, domain organization, and activity regulation. Biochemistry Biokhimiia *78*, 1377-1391.

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

Engelhardt, O.G., Smith, M., and Fodor, E. (2005). Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. J Virol *79*, 5812-5818.

Escarmis, C., Davila, M., Charpentier, N., Bracho, A., Moya, A., and Domingo, E. (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. Journal of molecular biology *264*, 255-267.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nature structural & molecular biology *11*, 822-829.

Fattovich, G., Boscaro, S., Noventa, F., Pornaro, E., Stenico, D., Alberti, A., Ruol, A., and Realdi, G. (1987). Influence of hepatitis delta virus infection on progression to cirrhosis in chronic hepatitis type B. The Journal of infectious diseases *155*, 931-935.

Fattovich, G., Giustina, G., Christensen, E., Pantalena, M., Zagni, I., Realdi, G., and Schalm, S.W. (2000). Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type B. The European Concerted Action on Viral Hepatitis (Eurohep). Gut *46*, 420-426.

Fauquet, C.M., and Fargette, D. (2005). International Committee on Taxonomy of Viruses and the 3,142 unassigned species. Virology journal *2*, 64.

Fels, A., Hu, K., and Riesner, D. (2001). Transcription of potato spindle tuber viroid by RNA polymerase II starts predominantly at two specific sites. Nucleic Acids Res *29*, 4589-4597.

Figueroa, A., Fujita, Y., and Gorospe, M. (2009a). Hacking RNA: Hakai promotes tumorigenesis by enhancing the RNA-binding function of PSF. Cell cycle *8*, 3648-3651.

Figueroa, A., Kotani, H., Toda, Y., Mazan-Mamczarz, K., Mueller, E.C., Otto, A., Disch, L., Norman, M., Ramdasi, R.M., Keshtgar, M.*, et al.* (2009b). Novel roles of hakai in cell proliferation and oncogenesis. Mol Biol Cell *20*, 3533-3542.

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

Flemington, E.K. (2001). Herpesvirus lytic replication and the cell cycle: arresting new developments. J Virol *75*, 4475-4481.

Flores, R., Grubb, D., Elleuch, A., Nohales, M.A., Delgado, S., and Gago, S. (2011). Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. RNA Biol *8*, 200-206.

Flores, R., Serra, P., Minoia, S., Di Serio, F., and Navarro, B. (2012). Viroids: from genotype to phenotype just relying on RNA sequence and structural motifs. Frontiers in microbiology *3*, 217.

Fong, K.W., Li, Y., Wang, W., Ma, W., Li, K., Qi, R.Z., Liu, D., Songyang, Z., and Chen, J. (2013). Whole-genome screening identifies proteins localized to distinct nuclear bodies. The Journal of cell biology *203*, 149-164.

Fox, A.H., Bond, C.S., and Lamond, A.I. (2005). p54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. Mol Biol Cell *16*, 5304-5315.

Fox, A.H., Lam, Y.W., Leung, A.K., Lyon, C.E., Andersen, J., Mann, M., and Lamond, A.I. (2002). Paraspeckles: a novel nuclear domain. Current biology : CB *12*, 13-25.

Fox, A.H., and Lamond, A.I. (2010). Paraspeckles. Cold Spring Harbor perspectives in biology *2*, a000687.

Gaal, T., Bartlett, M.S., Ross, W., Turnbough, C.L., Jr., and Gourse, R.L. (1997). Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. Science *278*, 2092-2097.

Gaeta, G.B., Stroffolini, T., Chiaramonte, M., Ascione, T., Stornaiuolo, G., Lobello, S., Sagnelli, E., Brunetto, M.R., and Rizzetto, M. (2000). Chronic hepatitis D: a vanishing Disease? An Italian multicenter study. Hepatology *32*, 824-827.

Gaglia, M.M., and Glaunsinger, B.A. (2010). Viruses and the cellular RNA decay machinery. Wiley interdisciplinary reviews RNA *1*, 47-59.

Gago, S., Elena, S.F., Flores, R., and Sanjuan, R. (2009). Extremely high mutation rate of a hammerhead viroid. Science *323*, 1308.

Galietta, A., Gunby, R.H., Redaelli, S., Stano, P., Carniti, C., Bachi, A., Tucker, P.W., Tartari, C.J., Huang, C.J., Colombo, E.*, et al.* (2007). NPM/ALK binds and phosphorylates the RNA/DNA-binding protein PSF in anaplastic large-cell lymphoma. Blood *110*, 2600-2609.

Gao, X., Kong, L., Lu, X., Zhang, G., Chi, L., Jiang, Y., Wu, Y., Yan, C., Duerksen-Hughes, P., Zhu, X.*, et al.* (2014). Paraspeckle protein 1 (PSPC1) is involved in the cisplatin induced DNA damage response--role in G1/S checkpoint. PLoS One *9*, e97174.

Gao, X., Zhang, G., Shan, S., Shang, Y., Chi, L., Li, H., Cao, Y., Zhu, X., Zhang, M., and Yang, J. (2016). Depletion of Paraspeckle Protein 1 Enhances Methyl Methanesulfonate-Induced Apoptosis through Mitotic Catastrophe. PLoS One *11*, e0146952.

Gernapudi, R., Wolfson, B., Zhang, Y., Yao, Y., Yang, P., Asahara, H., and Zhou, Q. (2016). MicroRNA 140 Promotes Expression of Long Noncoding RNA NEAT1 in Adipogenesis. Molecular and cellular biology *36*, 30-38.

Gibb, E.A., Vucic, E.A., Enfield, K.S., Stewart, G.L., Lonergan, K.M., Kennett, J.Y., Becker-Santos, D.D., MacAulay, C.E., Lam, S., Brown, C.J.*, et al.* (2011). Human cancer long non-coding RNA transcriptomes. PLoS One *6*, e25915.

Gilbert, W. (1986). Origin of life: The RNA world. Nature *319*, 618-618.

Gildehaus, N., Neusser, T., Wurm, R., and Wagner, R. (2007). Studies on the function of the riboregulator 6S RNA from E. coli: RNA polymerase binding, inhibition of in vitro transcription and synthesis of RNA-directed de novo transcripts. Nucleic Acids Res *35*, 1885-1896.

Glenn, J.S., Watson, J.A., Havel, C.M., and White, J.M. (1992). Identification of a prenylation site in delta virus large antigen. Science *256*, 1331-1333.

Glouzon, J.P., Bolduc, F., Wang, S., Najmanovich, R.J., and Perreault, J.P. (2014). Deep-sequencing of the peach latent mosaic viroid reveals new aspects of population heterogeneity. PLoS One *9*, e87297.

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

Goodrich, J.A., and Kugel, J.F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. Nature reviews Molecular cell biology *7*, 612-616.

Gorzer, I., Guelly, C., Trajanoski, S., and Puchhammer-Stockl, E. (2010). The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. Journal of virological methods *169*, 248-252.

Goss, D.J., and Kleiman, F.E. (2013). Poly(A) binding proteins: are they all created equal? Wiley interdisciplinary reviews RNA *4*, 167-179.

Gowans, E.J., Baroudy, B.M., Negro, F., Ponzetto, A., Purcell, R.H., and Gerin, J.L. (1988). Evidence for replication of hepatitis delta virus RNA in hepatocyte nuclei after in vivo infection. Virology *167*, 274-278.

Gozani, O., Patton, J.G., and Reed, R. (1994). A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. The EMBO journal *13*, 3356-3367.

Grabowski, J., Yurdaydin, C., Zachou, K., Buggisch, P., Hofmann, W.P., Jaroszewicz, J., Schlaphoff, V., Manns, M.P., Cornberg, M., Wedemeyer, H.*, et al.* (2011). Hepatitis D virus-specific cytokine responses in patients with chronic hepatitis delta before and during interferon alfa-treatment. Liver international : official journal of the International Association for the Study of the Liver *31*, 1395-1405.

Grey, N.K., Hrabalkova, L., Scanlon, J.P., and Smith, R.W. (2015). Poly(A)-binding proteins and mRNA localization: who rules the roost? Biochemical Society transactions *43*, 1277-1284.

Greco-Stewart, V.S. (2009). Characterization of the interaction of human DNA-dependent RNA polymerases and transcription factors with RNA from the hepatitis delta virus novel perspectives in HDV biology (University of Ottawa), pp. xvi, 294 leaves.

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

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

Griffin, B.L., Chasovskikh, S., Dritschilo, A., and Casey, J.L. (2014). Hepatitis delta antigen requires a flexible quasi-double-stranded RNA structure to bind and condense hepatitis delta virus RNA in a ribonucleoprotein complex. J Virol *88*, 7402-7411.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res *31*, 439-441.

Grosveld, G.C., Shewmaker, C.K., Jat, P., and Flavell, R.A. (1981). Localization of DNA sequences necessary for transcription of the rabbit beta-globin gene in vitro. Cell *25*, 215-226.

Guajardo, R., Lopez, P., Dreyfus, M., and Sousa, R. (1998). NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation. Journal of molecular biology *281*, 777-792.

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

Gudima, S., Dingle, K., Wu, T.T., Moraleda, G., and Taylor, J. (1999). Characterization of the 5' ends for polyadenylated RNAs synthesized during the replication of hepatitis delta virus. J Virol *73*, 6533-6539.

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

Gudima, S.O., Chang, J., and Taylor, J.M. (2005). Reconstitution in cultured cells of replicating HDV RNA from pairs of less than full-length RNAs. RNA *11*, 90-98.

Guo, S., Chen, W., Luo, Y., Ren, F., Zhong, T., Rong, M., Dang, Y., Feng, Z., and Chen, G. (2015). Clinical implication of long non-coding RNA NEAT1 expression in hepatocellular carcinoma patients. International journal of clinical and experimental pathology 8, 5395-5402.

Ha, K., Takeda, Y., and Dynan, W.S. (2011). Sequences in PSF/SFPQ mediate radioresistance and recruitment of PSF/SFPQ-containing complexes to DNA damage sites in human cells. DNA repair 10, 252-259.

Hansen, Thomas B; Jensen, Trine I; Clausen, Bettina H; Bramsen, Jesper B; Finsen, Bente; Damgaard, Christian K; Kjems, Jorgen. (2013). Natural RNA circles function as efficient microRNA sponges. Nature 495,7441,384-388.

Han, Z., Alves, C., Gudima, S., and Taylor, J. (2009). Intracellular localization of hepatitis delta virus proteins in the presence and absence of viral RNA accumulation. J Virol 83, 6457-6463.

Hantz, E., Larue, V., Ladam, P., Le Moyec, L., Gouyette, C., and Huynh Dinh, T. (2001). Solution conformation of an RNA--DNA hybrid duplex containing a pyrimidine RNA strand and a purine DNA strand. International journal of biological macromolecules *28*, 273-284.

Hartwig, D., Schutte, C., Warnecke, J., Dorn, I., Hennig, H., Kirchner, H., and Schlenke, P. (2006). The large form of ADAR 1 is responsible for enhanced hepatitis delta virus RNA editing in interferon-alpha-stimulated host cells. Journal of viral hepatitis *13*, 150-157.

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

Hawley, D.K., and McClure, W.R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res *11*, 2237-2255.

Heinicke, L.A., and Bevilacqua, P.C. (2012). Activation of PKR by RNA misfolding: HDV ribozyme dimers activate PKR. RNA *18*, 2157-2165.

Heise, T., Sommer, G., Reumann, K., Meyer, I., Will, H., and Schaal, H. (2006). The hepatitis B virus PRE contains a splicing regulatory element. Nucleic Acids Res *34*, 353-363.

Helmann, J.D., and deHaseth, P.L. (1999). Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. Biochemistry *38*, 5959-5967.

Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Benard, M., Fox, A.H.*, et al.* (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. Mol Biol Cell *25*, 169-183.

Hobbs, W.E., 2nd, and DeLuca, N.A. (1999). Perturbation of cell cycle progression and cellular gene expression as a function of herpes simplex virus ICP0. J Virol *73*, 8245-8255.

Houzet, L., and Jeang, K.T. (2011). Genome-wide screening using RNA interference to study host factors in viral replication and pathogenesis. Experimental biology and medicine *236*, 962-967.

Huang, C.R., and Lo, S.J. (2010). Evolution and diversity of the human hepatitis d virus genome. Advances in bioinformatics, 323654.

Huang, W.H., Mai, R.T., and Lee, Y.H. (2008). Transcription factor YY1 and its associated acetyltransferases CBP and p300 interact with hepatitis delta antigens and modulate hepatitis delta virus RNA replication. J Virol *82*, 7313-7324.

Huang, W.H., Yung, B.Y., Syu, W.J., and Lee, Y.H. (2001). The nucleolar phosphoprotein B23 interacts with hepatitis delta antigens and modulates the hepatitis delta virus RNA replication. The Journal of biological chemistry *276*, 25166-25175.

Huang, Y., Liu, N., Wang, J.P., Wang, Y.Q., Yu, X.L., Wang, Z.B., Cheng, X.C., and Zou, Q. (2012). Regulatory long non-coding RNA and its functions. Journal of physiology and biochemistry *68*, 611-618.

Hundley, H.A., and Bass, B.L. (2010). ADAR editing in double-stranded UTRs and other noncoding RNA sequences. Trends in biochemical sciences *35*, 377-383.

Huo, T.I., Wu, J.C., Wu, S.I., Chang, A.L., Lin, S.K., Pan, C.H., Huang, Y.H., Chang, F.Y., and Lee, S.D. (2004). Changing seroepidemiology of hepatitis B, C, and D virus infections in high-risk populations. Journal of medical virology *72*, 41-45.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol *8*, R143.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC genomics *8*, 39.

Hwang, S.B., and Park, K.J. (1999). Cell cycle arrest mediated by hepatitis delta antigen. FEBS letters *449*, 41-44.

Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., Kato, A., Kawaguchi, Y., Sato, H., Yoneda, M.*, et al.* (2014). Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. Mol Cell *53*, 393-406.

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

Inouye, C.J., and Seto, E. (1994). Relief of YY1-induced transcriptional repression by protein-protein interaction with the nucleolar phosphoprotein B23. The Journal of biological chemistry *269*, 6506-6510.

Ishitani, K., Yoshida, T., Kitagawa, H., Ohta, H., Nozawa, S., and Kato, S. (2003). p54nrb acts as a transcriptional coactivator for activation function 1 of the human androgen receptor. Biochemical and biophysical research communications *306*, 660-665.

Itin, A., and Keshet, E. (1983). Nucleotide sequence analysis of the long terminal repeat of murine virus-like DNA (VL30) and its adjacent sequences: resemblance to retrovirus proviruses. J Virol *47*, 656-659.

Iwasaki, T., Chin, W.W., and Ko, L. (2001). Identification and characterization of RRM-containing coactivator activator (CoAA) as TRBP-interacting protein, and its splice variant as a coactivator modulator (CoAM). The Journal of biological chemistry *276*, 33375-33383.

Iyer, L.M., Koonin, E.V., and Aravind, L. (2003). Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. BMC structural biology *3*, 1.

Jain, C., and Belasco, J.G. (2001). Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. Mol Cell *7*, 603-614.

Jang, S.B., Hung, L.W., Chi, Y.I., Holbrook, E.L., Carter, R.J., and Holbrook, S.R. (1998). Structure of an RNA internal loop consisting of tandem C-A+ base pairs. Biochemistry *37*, 11726-11731.

Jayan, G.C., and Casey, J.L. (2005). Effects of conserved RNA secondary structures on hepatitis delta virus genotype I RNA editing, replication, and virus production. J Virol *79*, 11187-11193.

Johnston, J.B., Wang, G., Barrett, J.W., Nazarian, S.H., Colwill, K., Moran, M., and McFadden, G. (2005). Myxoma virus M-T5 protects infected cells from the stress of cell cycle arrest through its interaction with host cell cullin-1. J Virol *79*, 10750-10763.

Kameoka, S., Duque, P., and Konarska, M.M. (2004). p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. The EMBO journal *23*, 1782-1791.

Kanamori, H., Yuhashi, K., Uchiyama, Y., Kodama, T., and Ohnishi, S. (2009). In vitro selection of RNA aptamers that bind the RNA-dependent RNA polymerase of hepatitis C virus: a possible role of GC-rich RNA motifs in NS5B binding. Virology *388*, 91-102.

Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2007). The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. Genes Dev *21*, 1779-1789.

Kannian, P., and Green, P.L. (2010). Human T Lymphotropic Virus Type 1 (HTLV-1): Molecular Biology and Oncogenesis. Viruses *2*, 2037-2077.

Kedersha, N.L., Gupta, M., Li, W., Miller, I., and Anderson, P. (1999). RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. The Journal of cell biology *147*, 1431-1442.

Kettenberger, H., Eisenfuhr, A., Brueckner, F., Theis, M., Famulok, M., and Cramer, P. (2006). Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. Nature structural & molecular biology *13*, 44-48.

Kidd-Ljunggren, K., Zuker, M., Hofacker, I.L., and Kidd, A.H. (2000). The hepatitis B virus pregenome: prediction of RNA structure and implications for the emergence of deletions. Intervirology *43*, 154-164.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520-527.

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

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev *14*, 2452-2460.

Konarska, M.M., and Sharp, P.A. (1989). Replication of RNA by the DNA-dependent RNA polymerase of phage T7. Cell *57*, 423-431.

Konarska, M.M., and Sharp, P.A. (1990). Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase. Cell *63*, 609-618.

Konopka, A.K., Reiter, J., Jung, M., Zarling, D.A., and Jovin, T.M. (1985). Concordance of experimentally mapped or predicted Z-DNA sites with positions of selected alternating purine-pyrimidine tracts. Nucleic Acids Res *13*, 1683-1701.

Krushkal, J., and Li, W.H. (1995). Substitution rates in hepatitis delta virus. Journal of molecular evolution *41*, 721-726.

Kula, A., Gharu, L., and Marcello, A. (2013). HIV-1 pre-mRNA commitment to Rev mediated export through PSF and Matrin 3. Virology *435*, 329-340.

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

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

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

Kuwahara, S., Ikei, A., Taguchi, Y., Tabuchi, Y., Fujimoto, N., Obinata, M., Uesugi, S., and Kurihara, Y. (2006). PSPC1, NONO, and SFPQ are expressed in mouse Sertoli cells and may function as coregulators of androgen receptor-mediated transcription. Biol Reprod *75*, 352-359.

Lai, D., Proctor, J.R., Zhu, J.Y., and Meyer, I.M. (2012). R-CHIE: a web server and R package for visualizing RNA secondary structures. Nucleic Acids Res *40*, e95.

Lai, M.M. (1992). Genetic recombination in RNA viruses. Current topics in microbiology and immunology *176*, 21-32.

Lai, M.M. (1995). The molecular biology of hepatitis delta virus. Annual review of biochemistry *64*, 259-286.

Lai, M.M. (1998). Cellular factors in the transcription and replication of viral RNA genomes: a parallel to DNA-dependent RNA transcription. Virology *244*, 1-12.

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

Lai, M.M., Chao, Y.C., Chang, M.F., Lin, J.H., and Gust, I. (1991). Functional studies of hepatitis delta antigen and delta virus RNA. Progress in clinical and biological research *364*, 283-292.

Lamond, A.I., and Carmo-Fonseca, M. (1993). Localisation of splicing snRNPs in mammalian cells. Molecular biology reports *18*, 127-133.

Landeras-Bueno, S., Jorba, N., Perez-Cidoncha, M., and Ortin, J. (2011). The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription. PLoS pathogens *7*, e1002397.

Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. RNA *20*, 1829-1842.

Lasda, E., and Parker, R. (2016). Circular RNAs Co-Precipitate with Extracellular Vesicles: A Possible Mechanism for circRNA Clearance. PLoS One *11*, e0148407.

Lauber, E., Guilley, H., Richards, K., Jonard, G., and Gilmer, D. (1997). Conformation of the 3'-end of beet necrotic yellow vein benevirus RNA 3 analysed by chemical and enzymatic probing and mutagenesis. Nucleic Acids Res *25*, 4723-4729.

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

Lazinski, D.W., and Taylor, J.M. (1994). Expression of hepatitis delta virus RNA deletions: cis and trans requirements for self-cleavage, ligation, and RNA packaging. J Virol *68*, 2879-2888.

Lazinski, D.W., and Taylor, J.M. (1995). Regulation of the hepatitis delta virus ribozymes: to cleave or not to cleave? RNA *1*, 225-233.

Le Gal, F., Gault, E., Ripault, M.P., Serpaggi, J., Trinchet, J.C., Gordien, E., and Deny, P. (2006). Eighth major clade for hepatitis delta virus. Emerging infectious diseases *12*, 1447-1450.

Lee, C.H., Chang, S.C., Chen, C.J., and Chang, M.F. (1998). The nucleolin binding activity of hepatitis delta antigen is associated with nucleolus targeting. The Journal of biological chemistry *273*, 7650-7656.

Lee, C.H., Chang, S.C., Wu, C.H., and 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*, 8142-8148.

Lee, C.M., Bih, F.Y., Chao, Y.C., Govindarajan, S., and Lai, M.M. (1992). Evolution of hepatitis delta virus RNA during chronic infection. Virology *188*, 265-273.

Lee, C.Z., Chen, P.J., and Chen, D.S. (1995). Large hepatitis delta antigen in packaging and replication inhibition: role of the carboxyl-terminal 19 amino acids and amino-terminal sequences. J Virol *69*, 5332-5336.

Lee, C.Z., and Sheu, J.C. (2008). Histone H1e interacts with small hepatitis delta antigen and affects hepatitis delta virus replication. Virology *375*, 197-204.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad Sci U S A *109*, E2774-2783.

Lee, W.P., Stromberg, M.P., Ward, A., Stewart, C., Garrison, E.P., and Marth, G.T. (2014). MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One *9*, e90581.

Lee, Y.M., and Lee, S.C. (1994). Transcriptional activation of the alpha-1 acid glycoprotein gene by YY1 is mediated by its functional interaction with a negative transcription factor. DNA and cell biology *13*, 1029-1036.

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

Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res *30*, 3497-3531.

Li, S., Li, Z., Shu, F.J., Xiong, H., Phillips, A.C., and Dynan, W.S. (2014). Double-strand break repair deficiency in NONO knockout murine embryonic fibroblasts and compensation by spontaneous upregulation of the PSPC1 paralog. Nucleic Acids Res *42*, 9771-9780.

Li, Y., Li, Y., Chen, W., He, F., Tan, Z., Zheng, J., Wang, W., Zhao, Q., and Li, J. (2015). NEAT expression is associated with tumor recurrence and unfavorable prognosis in colorectal cancer. Oncotarget *6*, 27641-27650.

Li, Y.J., Stallcup, M.R., and Lai, M.M. (2004). Hepatitis delta virus antigen is methylated at arginine residues, and methylation regulates subcellular localization and RNA replication. J Virol *78*, 13325-13334.

Liao, F.T., Hsu, L.S., Ko, J.L., Lin, C.C., and Sheu, G.T. (2012). Multiple genomic sequences of hepatitis delta virus are associated with cDNA promoter activity and RNA double rolling-circle replication. The Journal of general virology *93*, 577-587.

Lin, J.H., Chang, M.F., Baker, S.C., Govindarajan, S., and Lai, M.M. (1990). Characterization of hepatitis delta antigen: specific binding to hepatitis delta virus RNA. J Virol *64*, 4051-4058.

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

Lin, X., Thorne, L., Jin, Z., Hammad, L.A., Li, S., Deval, J., Goodfellow, I.G., and Kao, C.C. (2015). Subgenomic promoter recognition by the norovirus RNA-dependent RNA polymerases. Nucleic Acids Res *43*, 446-460.

Littlejohn, M., Locarnini, S., and Yuen, L. (2016). Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. Cold Spring Harbor perspectives in medicine *6*, a021360.

Liu, L., Oliveira, N.M., Cheney, K.M., Pade, C., Dreja, H., Bergin, A.M., Borgdorff, V., Beach, D.H., Bishop, C.L., Dittmar, M.T., et al. (2011). A whole genome screen for HIV restriction factors. Retrovirology 8, 94.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method.Methods, 25(4), 402-408.

Lodeiro, M.F., Filomatori, C.V., and Gamarnik, A.V. (2009). Structural and functional studies of the promoter element for dengue virus RNA replication. J Virol *83*, 993-1008.

Lomonte, P., and Everett, R.D. (1999). Herpes simplex virus type 1 immediate-early protein Vmw110 inhibits progression of cells through mitosis and from G(1) into S phase of the cell cycle. J Virol *73*, 9456-9467.

Long, H., Sung, W., Miller, S.F., Ackerman, M.S., Doak, T.G., and Lynch, M. (2015). Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient Pseudomonas fluorescens ATCC948. Genome biology and evolution *7*, 262-271.

Lowery, Laura Anne; Rubin, Jamie; Sive, Hazel. (2008). Whitesnake/sfpq is required for cell survival and neuronal development in the zebrafish. Developmental Dynamics 237,3,859-859.

Lu, J.Y., and Sewer, M.B. (2015). p54nrb/NONO regulates cyclic AMP-dependent glucocorticoid production by modulating phosphodiesterase mRNA splicing and degradation. Molecular and cellular biology *35*, 1223-1237.

Lu, M., and Shenk, T. (1999). Human cytomegalovirus UL69 protein induces cells to accumulate in G1 phase of the cell cycle. J Virol *73*, 676-683.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A *108*, 11063-11068.

Lukong, K.E., Huot, M.E., and Richard, S. (2009). BRK phosphorylates PSF promoting its cytoplasmic localization and cell cycle arrest. Cellular signalling *21*, 1415-1422.

Macnaughton, T.B., Gowans, E.J., Jilbert, A.R., and Burrell, C.J. (1990). Hepatitis delta virus RNA, protein synthesis and associated cytotoxicity in a stably transfected cell line. Virology *177*, 692-698.

Macnaughton, T.B., and Lai, M.M. (1993). Identification of promoters of hepatitis delta virus RNA transcription on hepatitis delta virus cDNA. Progress in clinical and biological research *382*, 13-20.

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

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. The Biochemical journal *417*, 639-650.

Major, A.T., Hogarth, C.A., Miyamoto, Y., Sarraj, M.A., Smith, C.L., Koopman, P., Kurihara, Y., Jans, D.A., and Loveland, K.L. (2015). Specific interaction with the nuclear transporter importin alpha2 can modulate paraspeckle protein 1 delivery to nuclear paraspeckles. Mol Biol Cell *26*, 1543-1558.

Manche, L., Green, S.R., Schmedt, C., and Mathews, M.B. (1992). Interactions between double-stranded RNA regulators and the protein kinase DAI. Molecular and cellular biology *12*, 5238-5248.

Mann, D.A., Mikaelian, I., Zemmel, R.W., Green, S.M., Lowe, A.D., Kimura, T., Singh, M., Butler, P.J., Gait, M.J., and Karn, J. (1994). A molecular rheostat. Co-operative

rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. Journal of molecular biology *241*, 193-207.

Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011a). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nature cell biology *13*, 95-101.

Mao, Y.S., Zhang, B., and Spector, D.L. (2011b). Biogenesis and function of nuclear bodies. Trends in genetics : TIG *27*, 295-306.

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

Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I.L., Hoffmann, D., Middendorf, M., Rattei, T., Stadler, P.F.*, et al.* (2014). Challenges in RNA virus bioinformatics. Bioinformatics *30*, 1793-1799.

Mathews, C.K. (2006). DNA precursor metabolism and genomic stability. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *20*, 1300-1314.

Mathur, M., Das, S., and Samuels, H.H. (2003). PSF-TFE3 oncoprotein in papillary renal cell carcinoma inactivates TFE3 and p53 through cytoplasmic sequestration. Oncogene *22*, 5031-5044.

Mathur, M., Tucker, P.W., and Samuels, H.H. (2001). PSF is a novel corepressor that mediates its effect through Sin3A and the DNA binding domain of nuclear hormone receptors. Molecular and cellular biology *21*, 2298-2311.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y.*, et al.* (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160-1166.

Matunis, M.J., Matunis, E.L., and Dreyfuss, G. (1993). PUB1: a major yeast poly(A)+ RNA-binding protein. Molecular and cellular biology *13*, 6114-6123.

Medvedeva, S.A., Panchin, A.Y., Alexeevski, A.V., Spirin, S.A., and Panchin, Y.V. (2013). Comparative analysis of context-dependent mutagenesis using human and mouse models. BioMed research international *2013*, 989410.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. Genes Dev *19*, 1401-1415.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M.*, et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333-338.

Mendes, M., Perez-Hernandez, D., Vazquez, J., Coelho, A.V., and Cunha, C. (2013). Proteomic changes in HEK-293 cells induced by hepatitis delta virus replication. Journal of proteomics *89*, 24-38.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nature reviews Genetics *10*, 155-159.

Miyamoto, K., Sakurai, H., and Sugiura, T. (2008). Proteomic identification of a PSF/p54nrb heterodimer as RNF43 oncoprotein-interacting proteins. Proteomics *8*, 2907-2910.

Modahl, L.E., Macnaughton, T.B., Zhu, N., Johnson, D.L., and Lai, M.M. (2000). RNA-Dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. Molecular and cellular biology *20*, 6030-6039.

Mokdad, A., and Frankel, A.D. (2008). ISFOLD: structure prediction of base pairs in non-helical RNA motifs from isostericity signatures in their sequence alignments. Journal of biomolecular structure & dynamics *25*, 467-472.

Monga, S.P.S., and Cagle, P.T. (2010). Molecular Pathology of Liver Diseases (Springer US).

Moore, P.B. (1999). In The RNA world: the nature of modern RNA suggests a prebiotic RNA, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor Laboratory Press).

Moras, D., and Poterszman, A. (1995). RNA-protein interactions. Diverse modes of recognition. Current biology : CB *5*, 249-251.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nature reviews Genetics *15*, 423-437.

Mota, S., Mendes, M., Freitas, N., Penque, D., Coelho, A.V., and Cunha, C. (2009). Proteome analysis of a human liver carcinoma cell line stably expressing hepatitis delta virus ribonucleoproteins. Journal of proteomics *72*, 616-627.

Mota, S., Mendes, M., Penque, D., Coelho, A.V., and Cunha, C. (2008). Changes in the proteome of Huh7 cells induced by transient expression of hepatitis D virus RNA and antigens. Journal of proteomics *71*, 71-79.

Motard, J., Bolduc, F., Thompson, D., and Perreault, J.P. (2008). The peach latent mosaic viroid replication initiation site is located at a universal position that appears to be defined by a conserved sequence. Virology *373*, 362-375.

Mu, J.J., Chen, D.S., and Chen, P.J. (2001). The conserved serine 177 in the delta antigen of hepatitis delta virus is one putative phosphorylation site and is required for efficient viral RNA replication. J Virol *75*, 9087-9095.

Mu, J.J., Tsay, Y.G., Juan, L.J., Fu, T.F., Huang, W.H., Chen, D.S., and Chen, P.J. (2004). The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine 72 may influence its cellular localization and viral RNA synthesis. Virology *319*, 60-70.

Mu, J.J., Wu, H.L., Chiang, B.L., Chang, R.P., Chen, D.S., and Chen, P.J. (1999). Characterization of the phosphorylated forms and the phosphorylated residues of hepatitis delta virus delta antigens. J Virol *73*, 10540-10545.

Muhlbach, H.P., and Sanger, H.L. (1979). Viroid replication is inhibited by alpha-amanitin. Nature *278*, 185-188.

Murthy, U.M., and Rangarajan, P.N. (2010). Identification of protein interaction regions of VINC/NEAT1/Men epsilon RNA. FEBS letters *584*, 1531-1535.

Myojin, R., Kuwahara, S., Yasaki, T., Matsunaga, T., Sakurai, T., Kimura, M., Uesugi, S., and Kurihara, Y. (2004). Expression and functional significance of mouse paraspeckle protein 1 on spermatogenesis. Biol Reprod *71*, 926-932.

Naganuma, T., and Hirose, T. (2013). Paraspeckle formation during the biogenesis of long non-coding RNAs. RNA Biol *10*, 456-461.

Nagy, P.D., and Simon, A.E. (1997). New insights into the mechanisms of RNA recombination. Virology *235*, 1-9.

Naji, S., Ambrus, G., Cimermancic, P., Reyes, J.R., Johnson, J.R., Filbrandt, R., Huber, M.D., Vesely, P., Krogan, N.J., Yates, J.R., 3rd*, et al.* (2012). Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. Molecular & cellular proteomics : MCP *11*, M111 015313.

Nakagawa, S., and Hirose, T. (2012). Paraspeckle nuclear bodies--useful uselessness? Cell Mol Life Sci *69*, 3027-3036.

Nathans, R., Chu, C.Y., Serquina, A.K., Lu, C.C., Cao, H., and Rana, T.M. (2009). Cellular microRNA and P bodies modulate host-HIV-1 interactions. Mol Cell *34*, 696-709.

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

Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics *25*, 1335-1337.

Negro, F., Pacchioni, D., Bussolati, G., and Bonino, F. (1991). Hepatitis delta virus heterogeneity: a study by immunofluorescence. Journal of hepatology *13 Suppl 4*, S125-129.

Netter, H.J., Kajino, K., and Taylor, J.M. (1993). Experimental transmission of human hepatitis delta virus to the laboratory mouse. J Virol *67*, 3357-3362.

Netter, H.J., Wu, T.T., Bockol, M., Cywinski, A., Ryu, W.S., Tennant, B.C., and Taylor, J.M. (1995). Nucleotide sequence stability of the genome of hepatitis delta virus. J Virol *69*, 1687-1692.

Ng, H.L., Kopka, M.L., and Dickerson, R.E. (2000). The structure of a stable intermediate in the A <--> B DNA helix transition. Proc Natl Acad Sci U S A *97*, 2035-2039.

Ng, K.K., Arnold, J.J., and Cameron, C.E. (2008). Structure-function relationships among RNA-dependent RNA polymerases. Current topics in microbiology and immunology *320*, 137-156.

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

Nikolov, D.B., Hu, S.H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.H., Roeder, R.G., and Burley, S.K. (1992). Crystal structure of TFIID TATA-box binding protein. Nature *360*, 40-46.

Ninio, J. (1991). Connections between translation, transcription and replication error-rates. Biochimie *73*, 1517-1523.

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

Nisini, R., Paroli, M., Accapezzato, D., Bonino, F., Rosina, F., Santantonio, T., Sallusto, F., Amoroso, A., Houghton, M., and Barnaba, V. (1997). Human CD4+ T-cell response to hepatitis delta virus: identification of multiple epitopes and characterization of T-helper cytokine profiles. J Virol *71*, 2241-2251.

Novina, C.D., and Roy, A.L. (1996). Core promoters and transcriptional control. Trends in genetics : TIG *12*, 351-355.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

Ossipow, V., Fonjallaz, P., and Schibler, U. (1999). An RNA polymerase II complex containing all essential initiation factors binds to the activation domain of PAR leucine zipper transcription factor thyroid embryonic factor. Molecular and cellular biology *19*, 1242-1250.

Otto, J.C., and 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*, 4569-4572.

Pallas, V., Garcia-Luque, I., Domingo, E., and Flores, R. (1988). Sequence variability in avocado sunblotch viroid (ASBV). Nucleic Acids Res *16*, 9864.

Park, C.Y., Oh, S.H., Kang, S.M., Lim, Y.S., and Hwang, S.B. (2009). Hepatitis delta virus large antigen sensitizes to TNF-alpha-induced NF-kappaB signaling. Molecules and cells *28*, 49-55.

Passon, D.M., Lee, M., Fox, A.H., and Bond, C.S. (2011). Crystallization of a paraspeckle protein PSPC1-NONO heterodimer. Acta Crystallogr Sect F Struct Biol Cryst Commun *67*, 1231-1234.

Passon, D.M., Lee, M., Rackham, O., Stanley, W.A., Sadowska, A., Filipovska, A., Fox, A.H., and Bond, C.S. (2012). Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. Proc Natl Acad Sci U S A *109*, 4846-4850.

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

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

Pelchat, M., and Perreault, J.P. (2004). Binding site of Escherichia coli RNA polymerase to an RNA promoter. Biochemical and biophysical research communications *319*, 636-642.

Peng, R., Dye, B.T., Perez, I., Barnard, D.C., Thompson, A.B., and Patton, J.G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. RNA *8*, 1334-1347.

Polson, A.G., Ley, H.L., 3rd, Bass, B.L., and Casey, J.L. (1998). Hepatitis delta virus RNA editing is highly specific for the amber/W site and is suppressed by hepatitis delta antigen. Molecular and cellular biology *18*, 1919-1926.

Ponzetto, A., Cote, P.J., Popper, H., Hoyer, B.H., London, W.T., Ford, E.C., Bonino, F., Purcell, R.H., and Gerin, J.L. (1984). Transmission of the hepatitis B virus-associated delta agent to the eastern woodchuck. Proc Natl Acad Sci U S A *81*, 2208-2212.

Ponzetto, A., Negro, F., Gerin, J.L., and Purcell, R.H. (1991). Experimental hepatitis delta virus infection in the animal model. Progress in clinical and biological research *364*, 147-157.

Ponzetto, A., Negro, F., Popper, H., Bonino, F., Engle, R., Rizzetto, M., Purcell, R.H., and Gerin, J.L. (1988). Serial passage of hepatitis delta virus in chronic hepatitis B virus carrier chimpanzees. Hepatology *8*, 1655-1661.

Poole, A.M., and Logan, D.T. (2005). Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome? Molecular biology and evolution *22*, 1444-1455.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. Cell *123*, 249-263.

Proteau, A., Blier, S., Albert, A.L., Lavoie, S.B., Traish, A.M., and Vincent, M. (2005). The multifunctional nuclear protein p54nrb is multiphosphorylated in mitosis and interacts with the mitotic regulator Pin1. Journal of molecular biology *346*, 1163-1172.

Pugnale, P., Pazienza, V., Guilloux, K., and Negro, F. (2009). Hepatitis delta virus inhibits alpha interferon signaling. Hepatology *49*, 398-406.

Query, C.C., Bentley, R.C., and Keene, J.D. (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell *57*, 89-101.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nature reviews Genetics *17*, 47-62.

Quiskamp, N., Poeter, M., Raabe, C.A., Hohenester, U.M., Konig, S., Gerke, V., and Rescher, U. (2014). The tumor suppressor annexin A10 is a novel component of nuclear paraspeckles. Cell Mol Life Sci 71, 311-329.

R Core Team (2013). R: A language and environment for statistical computing. RFoundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rajesh, C., Baker, D.K., Pierce, A.J., and Pittman, D.L. (2011). The splicing-factor related protein SFPQ/PSF interacts with RAD51D and is necessary for homology-directed repair and sister chromatid cohesion. Nucleic Acids Res *39*, 132-145.

Rajesh, P., Rajesh, C., Wyatt, M.D., and Pittman, D.L. (2010). RAD51D protects against MLH1-dependent cytotoxic responses to O(6)-methylguanine. DNA repair *9*, 458-467.

Rakowski, A.G., and Symons, R.H. (1989). Comparative sequence studies of variants of avocado sunblotch viroid. Virology *173*, 352-356.

Ranjith-Kumar, C.T., Zhang, X., and Kao, C.C. (2003). Enhancer-like activity of a brome mosaic virus RNA promoter. J Virol *77*, 1830-1839.

Rehm, C., Wurmthaler, L.A., Li, Y., Frickey, T., and Hartig, J.S. (2015). Investigation of a Quadruplex-Forming Repeat Sequence Highly Enriched in Xanthomonas and Nostoc sp. PLoS One *10*, e0144275.

Riva, M.A., Riva, E., Spicci, M., Strazzabosco, M., Giovannini, M., and Cesana, G. (2011). "The city of Hepar": rituals, gastronomy, and politics at the origins of the modern names for the liver. Journal of hepatology *55*, 1132-1136.

Rizzetto, M., Canese, M.G., Arico, S., Crivelli, O., Trepo, C., Bonino, F., and Verme, G. (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*, 997-1003.

Rizzetto, M., Purcell, R.H., and Gerin, J.L. (1980a). Epidemiology of HBV-associated delta agent: geographical distribution of anti-delta and prevalence in polytransfused HBsAg carriers. Lancet *1*, 1215-1218.

Rizzetto, M., Shih, J.W., and Gerin, J.L. (1980b). The hepatitis B virus-associated delta antigen: isolation from liver, development of solid-phase radioimmunoassays for delta antigen and anti-delta and partial characterization of delta antigen. Journal of immunology *125*, 318-324.

Robertson, H.D., Manche, L., and Mathews, M.B. (1996). Paradoxical interactions between human delta hepatitis agent RNA and the cellular protein kinase PKR. J Virol *70*, 5611-5617.

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

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends in biochemical sciences *21*, 327-335.

Romeo, R., Del Ninno, E., Rumi, M., Russo, A., Sangiovanni, A., de Franchis, R., Ronchi, G., and Colombo, M. (2009). A 28-year study of the course of hepatitis Delta infection: a risk factor for cirrhosis and hepatocellular carcinoma. Gastroenterology *136*, 1629-1638.

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

Ryu, W.S., Netter, H.J., Bayer, M., and Taylor, J. (1993). Ribonucleoprotein complexes of hepatitis delta virus. J Virol *67*, 3281-3287.

Saenger, W. (1984). Principles of Nucleic Acid Structure (Springer New York).

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One *7*, e30733.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nature reviews Genetics *8*, 424-436.

Sasaki, Y.T., and Hirose, T. (2009). How to build a paraspeckle. Genome Biol *10*, 227.

Sato, S., Cornillez-Ty, C., and Lazinski, D.W. (2004). By inhibiting replication, the large hepatitis delta antigen can indirectly regulate amber/W editing and its own expression. J Virol *78*, 8120-8134.

Schroeder, L.A., and deHaseth, P.L. (2005). Mechanistic differences in promoter DNA melting by Thermus aquaticus and Escherichia coli RNA polymerases. The Journal of biological chemistry *280*, 17422-17429.

Schwartz, S., Campbell, M., Nasioulas, G., Harrison, J., Felber, B.K., and Pavlakis, G.N. (1992). Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. J Virol *66*, 7176-7182.

Seeman, N.C., Rosenberg, J.M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A *73*, 804-808.

Seemann, S.E., Gorodkin, J., and Backofen, R. (2008). Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. Nucleic Acids Res *36*, 6355-6362.

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

Shelkovnikova, T.A., Robinson, H.K., Troakes, C., Ninkina, N., and Buchman, V.L. (2014). Compromised paraspeckle formation as a pathogenic factor in FUSopathies. Hum Mol Genet *23*, 2298-2312.

Showalter, S.A., and Hall, K.B. (2004). Altering the RNA-binding mode of the U1A RBD1 protein. Journal of molecular biology *335*, 465-480.

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

Sikora, D., and University of Ottawa. Dept. of Biochemistry Microbiology and Immunology. (2012). Hepatitis Delta Virus identification of host factors involved in the viral life cycle, and the investigation of the evolutionary relationship between HDV and plant viroids.

Sims, R.J., 3rd, Mandal, S.S., and Reinberg, D. (2004). Recent highlights of RNA-polymerase-II-mediated transcription. Current opinion in cell biology *16*, 263-271.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica et biophysica acta *1351*, 73-88.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annual review of biochemistry *72*, 449-479.

Song, B., Yeh, K.C., Liu, J., and Knipe, D.M. (2001). Herpes simplex virus gene products required for viral inhibition of expression of G1-phase functions. Virology *290*, 320-328.

Song, X., Sui, A., and Garen, A. (2004). Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: effects on steroidogenesis and oncogenesis. Proc Natl Acad Sci U S A *101*, 621-626.

Song, X., Sun, Y., and Garen, A. (2005). Roles of PSF protein and VL30 RNA in reversible gene regulation. Proc Natl Acad Sci U S A *102*, 12189-12193.

Song, X., Wang, B., Bromberg, M., Hu, Z., Konigsberg, W., and Garen, A. (2002). Retroviral-mediated transmission of a mouse VL30 RNA to human melanoma cells promotes metastasis in an immunodeficient mouse model. Proc Natl Acad Sci U S A *99*, 6269-6273.

Souquere, S., Beauclair, G., Harper, F., Fox, A., and Pierron, G. (2010). Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. Mol Biol Cell *21*, 4020-4027.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harbor perspectives in biology *3*.

Spitale, R.C., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2014). RNA structural analysis by evolving SHAPE chemistry. Wiley interdisciplinary reviews RNA *5*, 867-881.

Steuten, B., Hoch, P.G., Damm, K., Schneider, S., Kohler, K., Wagner, R., and Hartmann, R.K. (2014). Regulation of transcription by 6S RNAs: insights from the Escherichia coli and Bacillus subtilis model systems. RNA Biol *11*, 508-521.

Straub, T., Grue, P., Uhse, A., Lisby, M., Knudsen, B.R., Tange, T.O., Westergaard, O., and Boege, F. (1998). The RNA-splicing factor PSF/p54 controls DNA-topoisomerase I activity by a direct interaction. The Journal of biological chemistry 273, 26261-26264.

Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S., & Spector, D. L. (2009). MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Research, 19(3), 347–359.

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., Moriarty, A.M., Thornton, G.B., and Lanford, R.E. (1992). Production of infectious hepatitis delta virus in vitro and neutralization with antibodies directed against hepatitis B virus pre-S antigens. J Virol 66, 1241-1245.

Szybalski W, Kubinski H, Sheldrick P.(1966.Pyrimidine clusters on the transcribing strand of DNA and their possible role in the initiation of RNA synthesis. Cold Spring Harb Symp Quant Biol 31:123-7.

Takeda, R., Petrov, A.I., Leontis, N.B., and Ding, B. (2011). A three-dimensional RNA motif in Potato spindle tuber viroid mediates trafficking from palisade mesophyll to spongy mesophyll in Nicotiana benthamiana. The Plant cell *23*, 258-272.

Tavanez, J.P., Cunha, C., Silva, M.C., David, E., Monjardino, J., and Carmo-Fonseca, M. (2002). Hepatitis delta virus ribonucleoproteins shuttle between the nucleus and the cytoplasm. RNA *8*, 637-646.

Taylor, J., Mason, W., Summers, J., Goldberg, J., Aldrich, C., Coates, L., Gerin, J., and Gowans, E. (1987). Replication of human hepatitis delta virus in primary cultures of woodchuck hepatocytes. J Virol *61*, 2891-2895.

Taylor, J., and Pelchat, M. (2010). Origin of hepatitis delta virus. Future microbiology *5*, 393-402.

Taylor, J.M. (2009). Chapter 3. Replication of the hepatitis delta virus RNA genome. Advances in virus research *74*, 103-121.

Taylor, J.M. (2014). Host RNA circles and the origin of hepatitis delta virus. World journal of gastroenterology *20*, 2971-2978.

Taylor, J.M. (2015). Hepatitis D Virus Replication. Cold Spring Harbor perspectives in medicine *5*.

Teplova, M., Song, J., Gaw, H.Y., Teplov, A., and Patel, D.J. (2010). Structural insights into RNA recognition by the alternate-splicing regulator CUG-binding protein 1. Structure *18*, 1364-1377.

Thomas, M., Chedin, S., Carles, C., Riva, M., Famulok, M., and Sentenac, A. (1997). Selective targeting and inhibition of yeast RNA polymerase II by RNA aptamers. The Journal of biological chemistry *272*, 27980-27986.

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

Trotochaud, A.E., and Wassarman, K.M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. Nature structural & molecular biology *12*, 313-319.

Tseng, C.H., Cheng, T.S., Shu, C.Y., Jeng, K.S., and Lai, M.M. (2010). Modification of small hepatitis delta virus antigen by SUMO protein. J Virol *84*, 918-927.

Tsuda, K., Kuwasako, K., Takahashi, M., Someya, T., Inoue, M., Terada, T., Kobayashi, N., Shirouzu, M., Kigawa, T., Tanaka, A.*, et al.* (2009). Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3. Nucleic Acids Res *37*, 5151-5166.

Udayakumar, D., Bladen, C.L., Hudson, F.Z., and Dynan, W.S. (2003). Distinct pathways of nonhomologous end joining that are differentially regulated by DNA-dependent protein kinase-mediated phosphorylation. The Journal of biological chemistry *278*, 41631-41635.

Urban, R.J., Bodenburg, Y., Kurosky, A., Wood, T.G., and Gasic, S. (2000). Polypyrimidine tract-binding protein-associated splicing factor is a negative regulator of transcriptional activity of the porcine p450scc insulin-like growth factor response element. Molecular endocrinology *14*, 774-782.

Urban, R.J., Bodenburg, Y.H., and Wood, T.G. (2002). NH2 terminus of PTB-associated splicing factor binds to the porcine P450scc IGF-I response element. American journal of physiology Endocrinology and metabolism *283*, E423-427.

Usheva, A., and Shenk, T. (1996). YY1 transcriptional initiator: protein interactions and association with a DNA site containing unpaired strands. Proc Natl Acad Sci U S A *93*, 13571-13576.

Ussery, D., Soumpasis, D.M., Brunak, S., Staerfeldt, H.H., Worning, P., and Krogh, A. (2002). Bias of purine stretches in sequenced chromosomes. Computers & chemistry *26*, 531-541.

Velazquez, G., Guo, Q., Wang, L., Brieba, L.G., and Sousa, R. (2012). Conservation of promoter melting mechanisms in divergent regions of the single-subunit RNA polymerases. Biochemistry *51*, 3901-3910.

Verrijzer, C.P., and Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity. Trends in biochemical sciences *21*, 338-342.

Wagner, S.D., Yakovchuk, P., Gilman, B., Ponicsan, S.L., Drullinger, L.F., Kugel, J.F., and Goodrich, J.A. (2013). RNA polymerase II acts as an RNA-dependent RNA polymerase to extend and destabilize a non-coding RNA. The EMBO journal *32*, 781-790.

Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., and Chang, H.Y. (2011). Understanding the transcriptome through RNA structure. Nature reviews Genetics *12*, 641-655.

Wang, D., Pearlberg, J., Liu, Y.T., and Ganem, D. (2001). Deleterious effects of hepatitis delta virus replication on host cell proliferation. J Virol *75*, 3600-3604.

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

Wang, K.S., Choo, O.L., Weiner, A.J., Ou, J.H., Denniston, K.J., Gerin, J.L., and Houghton, M. (1987). The viroid-like structure of the hepatitis delta (delta) genome: synthesis of a viral antigen in recombinant bacteria. Progress in clinical and biological research *234*, 71-82.

Wang, K.S., Choo, Q.L., Weiner, A.J., Ou, J.H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L., and Houghton, M. (1986). Structure, sequence and expression of the hepatitis delta (delta) viral genome. Nature *323*, 508-514.

Wassarman, K.M. (2007). 6S RNA: a regulator of transcription. Molecular microbiology *65*, 1425-1431.

Wassarman, K.M., and Saecker, R.M. (2006). Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. Science *314*, 1601-1603.

Wedemeyer, H., and Manns, M.P. (2010). Epidemiology, pathogenesis and management of hepatitis D: update and challenges ahead. Nature reviews Gastroenterology & hepatology 7, 31-40.

Weeks, K.M., and Crothers, D.M. (1993). Major groove accessibility of RNA. Science 261, 1574-1577.

West, Jason A., et al. "The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites." Molecular cell 55.5 (2014): 791-802.

Westhof, E. (2015). RNA structure from deep sequencing. Nature biotechnology 33, 928-929.

Widenius M, Axmark D. Mysql Reference Manual. 1st ed. Paul DuBois. Sebastopol, CA USA: O'Reilly & Associates Inc., 2002.

Wiebusch, L., and Hagemeier, C. (1999). Human cytomegalovirus 86-kilodalton IE2 protein blocks cell cycle progression in G(1). J Virol *73*, 9274-9283.

Williams, M.a.D., T. (2013). Protein-Ligand Interactions. Methods in Molecular Biology *1008*.

Wolff, H., Brack-Werner, R., Neumann, M., Werner, T., and Schneider, R. (2003). Integrated functional and bioinformatics approach for the identification and experimental verification of RNA signals: application to HIV-1 INS. Nucleic Acids Res *31*, 2839-2851.

Wong, S.K., and Lazinski, D.W. (2002). Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. Proc Natl Acad Sci U S A *99*, 15118-15123.

Wu, H.N., Lin, Y.J., Lin, F.P., Makino, S., Chang, M.F., and Lai, M.M. (1989). Human hepatitis delta virus RNA subfragments contain an autocleavage activity. Proc Natl Acad Sci U S A *86*, 1831-1835.

Wu, J.C., Chen, C.L., Lee, S.D., Sheen, I.J., and Ting, L.P. (1992). Expression and localization of the small and large delta antigens during the replication cycle of hepatitis D virus. Hepatology *16*, 1120-1127.

Wu, J.C., Chiang, T.Y., Shiue, W.K., Wang, S.Y., Sheen, I.J., Huang, Y.H., and Syu, W.J. (1999). Recombination of hepatitis D virus RNA sequences and its implications. Molecular biology and evolution *16*, 1622-1632.

Wu, J.C., Hsu, S.C., Wang, S.Y., Huang, Y.H., Sheen, I.J., Shih, H.H., and Syu, W.J. (2005). "Defective" mutations of hepatitis D viruses in chronic hepatitis D patients. World journal of gastroenterology *11*, 1658-1662.

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

Wu, W., Yan, C., Gan, T., Chen, Z., Lu, X., Duerksen-Hughes, P.J., Zhu, X., and Yang, J. (2010). Nuclear proteome analysis of cisplatin-treated HeLa cells. Mutation research *691*, 1-8.

Xia, Y.P., Yeh, C.T., Ou, J.H., and Lai, M.M. (1992). Characterization of nuclear targeting signal of hepatitis delta antigen: nuclear transport as a protein complex. J Virol *66*, 914-921.

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M.M., and Handa, H. (2001). Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science *293*, 124-127.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Molecular and cellular biology *22*, 2918-2927.

Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S., and 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*, 863-875.

Yamaguchi, Y., Wada, T., and Handa, H. (1998). Interplay between positive and negative elongation factors: drawing a new view of DRB. Genes to cells : devoted to molecular & cellular mechanisms *3*, 9-15.

Yamazaki, T., and Hirose, T. (2015). The building process of the functional paraspeckle with long non-coding RNAs. Front Biosci (Elite Ed) *7*, 1-41.

Yarosh, C.A., Iacona, J.R., Lutz, C.S., and Lynch, K.W. (2015). PSF: nuclear busy-body or nuclear facilitator? Wiley interdisciplinary reviews RNA *6*, 351-367.

Yuan, X., Wu, J., Shan, Y., Yao, Z., Dong, B., Chen, B., Zhao, Z., Wang, S., Chen, J., and Cong, Y. (2006). SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway. Virology *346*, 74-85.

Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225-229.

Zawel, L., and Reinberg, D. (1993). Initiation of transcription by RNA polymerase II: a multi-step process. Progress in nucleic acid research and molecular biology *44*, 67-108.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nature cell biology *16*, 2-9.

Zhang, D.J., and University of Ottawa. Department of Biochemistry Microbiology and Immunology. (2013) Involvement of the Polypyrimidine Tract-Binding Protein-Associated Splicing Factor (PSF) in the Hepatitis Delta Virus (HDV) RNA-Templated Transcription.

Zhang, Q., Chen, C.Y., Yedavalli, V.S., and Jeang, K.T. (2013a). NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. MBio *4*, e00596-00512.

Zhang, W.W., Zhang, L.X., Busch, R.K., Farres, J., and Busch, H. (1993). Purification and characterization of a DNA-binding heterodimer of 52 and 100 kDa from HeLa cells. The Biochemical journal *290 (Pt 1)*, 267-272.

Zhang, Y., Zhang, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L., and Chen, L.L. (2013b). Circular intronic long noncoding RNAs. Mol Cell *51*, 792-806.

Zhang, Z., and Carmichael, G.G. (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell *106*, 465-475.

Zhong, N., Kim, C.Y., Rizzu, P., Geula, C., Porter, D.R., Pothos, E.N., Squitieri, F., Heutink, P., and Xu, J. (2006). DJ-1 transcriptionally up-regulates the human tyrosine hydroxylase by inhibiting the sumoylation of pyrimidine tract-binding protein-associated splicing factor. The Journal of biological chemistry *281*, 20940-20948.

Zhong, X., and Ding, B. (2008). Distinct RNA motifs mediate systemic RNA trafficking. Plant signaling & behavior *3*, 58-59.

Zolotukhin, A.S., Michalowski, D., Bear, J., Smulevitch, S.V., Traish, A.M., Peng, R., Patton, J., Shatsky, I.N., and Felber, B.K. (2003). PSF acts through the human immunodeficiency virus type 1 mRNA instability elements to regulate virus expression. Molecular and cellular biology *23*, 6618-6630.

Contribution of collaborators

The Infernal alignment was performed by Dr Martin Pelchat and the bio-informatic scripts were written by Mrs Lynda Rocheleau and Dr Martin Pelchat (chapter 3). The immunoprecipitation step of the RIP experiments of HDV RNA with PSF, p54 and PSP1 were performed by Ms Gabrielle Goodrum, under the guidance of Yasnee Beeharry (chapter 4.1). The qPCR to quantify NEAT1 and IL8 in 293-HDV cells were performed by Mr Christian John Imperiale under the guidance of Dr Dorota Sikora (chapter 4.2).

Appendices

Appendix I Review of R199G mutational assays

Nucleotide positions	Mutations	Detection of HDV	Plasmid/RNA transfection and detection method	Reference
1665-1669	CGCCC>AT ATA	Complete abolition of RNA replication	Plasmid transfection, Luciferase reporter gene assay	(Liao et al., 2012)
1657-1658	GG>AA	able to synthesize short RNAs but not monomeric HDV 68 RNA.	plasmid	(Liao et al., 2012)
1596-1598	AUG>UAG	loss of dimeric but not 65 monomeric HDV RNA synthesis	plasmid	(Liao et al., 2012)
1599-1678	deletion	Reduction promoter activity by 50%	plasmid	(Liao et al., 2012)
1579-1588 genomic	deletion	reduction in transcriptional 154 activity	plasmid	(Liao et al., 2012)
1579-1598 genomic	deletion	reduced approximately 50%	plasmid	(Liao et al., 2012)
1579-1598 antigenomic	deletion	complete loss of promoter 157 activity	plasmid	(Liao et al., 2012)
1596-1598 genomic	CAU>CUA	Monomeric but not dimeric HDV	plasmid	(Liao et al., 2012)
1596-1598 antigenomic	AUG>UAG	Monomeric but not dimeric HDV	plasmid	(Liao et al., 2012)
1665-1689	CGCCC>AT ATA	No genomic or antigenomic RNA detected	RNA	(Liao et al., 2012)
1657-1658 genomic	GG>AA	smaller than 247 monomeric HDV RNA fragments were detected	RNA transfection, Northern detection	(Liao et al., 2012)
1657-1658 antigenomic	GG>AA	genomic-sense HDV RNA	RNA	(Liao et al., 2012)

		monomer was detected		
1566-1620; 1655-18	Deletion; deletion	Strong band	Co-immunoprecipitation with RNAP II	(Greco-Stewart et al., 2007)
1566-1620; 1655-18 ; 1642- 1650	Deletion; Deletion ; AAAGAGGAG>UUUCUCCUC	No band	Co-immunoprecipitation with RNAP II	(Greco-Stewart et al., 2007)
1566-1620; 1655-18 ; 1626-1628 ; 1632-1637	Deletion ; Deletion ; CUC > GAG ; CUCUUU >GAGAAA	No band	Co-immunoprecipitation with RNAP II	(Greco-Stewart et al., 2007)
1566-1620; 1655-18 ; 1626-1628 ; 1632-1637 ; 1642- 1650; 1647	Deletion ; Deletion ; CUC > GAG ; CUCUUU >GAGAAA ; AAAGAGGAG>UUUCUCCUC ; Insertion UUA	No band	Co-immunoprecipitation with RNAP II	(Greco-Stewart et al., 2007)
1630	U>C	Antigenomic RNA accumulation (%) 25	Transfection of cDNA > detection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1630	U>G	Antigenomic RNA accumulation (%) 6	Transfection of cDNA > detection of RNA by Northern	(Gudima et al., 1999)
1630	U>A	Antigenomic RNA	Transfection of cDNA >	(Gudima et al., 1999)

		accumulation (%) 80	detection of RNA by Northern	
1629	U>UU	Antigenomic RNA accumulation (%) 9	Transfection of cDNA > detection of RNA by Northern	(Gudima et al., 1999)
1648	G>GU	Antigenomic RNA accumulation (%) 47	Transfection of cDNA > detection of RNA by Northern	(Gudima et al., 1999)
1631	A>G	100		
1629	U>A	Antigenomic RNA accumulation (%) 100	Transfection of cDNA > detection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1648	G>–	44	Transfection of cDNA > detection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
16291630	UU>– –	42	Transfection of cDNAdetection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1648	G> GUA	> 1%	Transfection of cDNA detection of	(Gudima et al., 1999)

			RNA by Northern (check mutation, RACE)	
1648	G > GUAA	> 1%	Transfection of cDNAdetection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1637	U > UCU	100	Transfection of cDNAdetection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1637	U > UCUCU	80	Transfection of cDNAdetection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1637, 1642	U > UG, U> UC	60	Transfection of cDNAdetection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1637, 1642	U > UGA, U > UUC	40	Transfection of cDNAdetection of RNA by Northern (check mutation, RACE)	(Gudima et al., 1999)
1638-1641	UCUGU > GUAUC	108	Transfection in Huh7 and quantification	(Wu et al., 1997)

			of RNA by Northern	
1638-1641	UCUGU > AUUCG	84	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1638-1641	UCUGU > CGACA	123	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1638-1641	UCUGU > UACAA	83	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1638-1641	UCUGU > UAGCU	43	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1638-1641	UCUGU > CCGUC	50	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1638-1641	Insertion UCUGU > CUUCUGU	60	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1638-1641	Insertion UCUGU > CUCUUCUGU	70	Transfection in Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1637	Insertion G	100	Huh7 and quantification of RNA by	(Wu et al., 1997)

			Northern	
1637	Insertion GA	16	Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1637	deletion	1	Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1636, 1637	deletion	0.1	Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1635,1636,1637	deletion	<0.06	Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1632-1637 <-> 1642- 1647	Stem flip	1.8	Huh7 and quantification of RNA by Northern	(Wu et al., 1997)
1622-1631	AGUACUCUUACU > AGUGCGGCCGCACU	No antigenomic RNA signal (G>AG and AG>G)	Linker-scanning mutations to keep secondary structure , transfection in Huh7, Northern Blot	Wang et al., 1997
8-17	CAAGUUCUUGA > CAAGCGGCCGCCGA	RNA replication slightly affected (G>AG and AG>G)	Linker-scanning mutations to keep secondary structure , transfection in Huh7, Northern Blot	(Wang et al., 1997)
1622-1631	AGUACUCUUACU > AGUGCGGCCGCACU	No antigenomic RNA signal Defective mRNA synthesis	Transfection of mutated RNA in Huh7, Northern Blot	(Wang et al., 1997)
8-17	CAAGUUCUUGA > CAAGCGGCCGCCGA	RNA replication (genomic > antigenomic) slightly affected	Transfection of mutated RNA in Huh7, Northern Blot	(Wang et al., 1997)

		Defective mRNA synthesis		
1664-1669	CCGCCC> UAUAUA	Almost no replication product 0.02	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1666 ; 1670	G > U, G > U	Low replication levels 0.4	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1657-1658	GG>AA	Almost no replication product 0.02	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1655	C > A	Level of replication comparable 0.8	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1668	C > A	Level of replication comparable 0.9	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1662	deletion	Higher levels of replication 1.2	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1655-1656	CU>UC	Higher levels of replication 1.4	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1654	G > A	Lower levels of replication 0.09	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
1633 ; 1625	U > A ; deletion	Lower levels of replication 0.09	COS-7 cells transfected with plasmid > detection by	(Beard et al., 1996)

			Northern Blot	
1662	deletion	Higher than wild-type 1.2	COS-7 cells transfected with plasmid > detection by Northern Blot	(Beard et al., 1996)
R199G			Transcription nuclear assay	(Beard et al., 1996)
1566-1679 genomic	deletion	Promoter activity lost	Plasmid transfected > CAT reporter gene	(Macnaughton and Lai, 1993)
1566-1649 genomic	deletion	No CAT activity	Plasmid transfected > CAT reporter gene	(Macnaughton and Lai, 1993)
1650-1679 genomic		Low CAT activity	Plasmid transfected > CAT reporter gene	(Macnaughton and Lai, 1993)
1679-1650 antigenomic	Essential for promoter activity	CAT activity (confirmed by human growth hormone gene reporter)	Plasmid transfected > CAT reporter gene	(Macnaughton and Lai, 1993)
479-1650 antigenomic	Lower activity than construct 1679-1650	Essential for CAT activity (confirmed by human growth hormone gene reporter)		(Macnaughton and Lai, 1993)
479-1655 antigenomic	Lower activity than construct 1679-1650	Essential for CAT activity (confirmed by human growth hormone gene reporter)		(Macnaughton and Lai, 1993)

Appendix II Review of the functions of PSF, p54 and PSP1 in the cell

Cellular pathway	Role of DBHS protein	References
DNA replication	PSF interacts directly with the DNA and fosters the annealing of two DNA molecules to form a double strand	(Akhmedov et al., 1995; Akhmedov and Lopez, 2000)
	PSF promotes the association of the complementary DNA to the double-stranded DNA to form a D-loop	(Akhmedov et al., 1995; Akhmedov and Lopez, 2000)
radioresistance	Use of PSF rescue mutants in HeLa cells knocked-down for PSF allowed for identifying that the N-terminal domain, and possibly also the RRM1 of PSF are responsible for this function	(Ha et al., 2011)
	Knockdown experiments of PSF in mouse cells unveiled the role of PSF	(Rajesh et al., 2010)
DNA double-strand break repair pathway	PSF interacts with RAD51D, a major protein in the repair pathway of double-stranded breaks The knockdown of both Rad51 and PSF in Mouse Embryonic Fibroblast cells resulted in increased cellular death.	(Rajesh et al., 2010)
	Following DNA double stranded break, the PSF and P54 complex, together with the Ku protein, are involved in the rejoining of the two DNA molecules.	(Bladen et al., 2005; Udayakumar et al., 2003)
	This role is conferred by PSF sequence situated in the N-terminal and the RRM1 region, rather than its ability of heterodimerize with its molecular partner p54. Following the DNA damage stimulus, the knockdown of PSF abrogated the recruitment of p54 to the DNA damage site, but the knockdown of p54 did not affect the recruitment of PSF. Furthermore, when the pool of p54 was limiting or when PSF was overexpressed in the cell, PSF recruited PSP1 to the DNA damaged sites	(Ha et al., 2011)
	The overexpression of PSP1 restored the double-strand break repair response in murine embryonic fibroblast knockout for	(Li et al., 2014)

	p54	
	PSP1 was also identified in a proteomic screen to identify protein interacting with ATR and ATM in response to DNA damage	(Matsuoka et al., 2007; Wu et al., 2010).
	Assays where cells were treated with the DNA damaging chemical cisplatin showed that PSP1 played a role in the DNA damage response	(Gao et al., 2014)
	The treatment of the DNA-damaging chemical methyl methanesulfonate resulted in an increase of PSP1 levels in the cell. The overexpression of PSP1 combined with the use of methyl methanesulfonate resulted in a reduction of the number of apoptotic cells compared to the cells only treated with methyl methanesulfonate alone	(Gao et al., 2016)
	In A341 cells, the heterodimer PSF/p54 interacts directly with the DNA-topoisomerase I and increase its activity by 16 fold	(Straub et al., 1998)
	Knockdown of p54 interfered with the DNA double-strand break repair pathway	(Mahaney et al., 2009)
splicing	PSF is essential for pre-mRNA splicing and p54 interacts with other splicing proteins	(Gozani et al., 1994; Patton et al., 1993; Peng et al., 2002)
	Crosslinking assays of a radiolabelled RNA to the proteins showed that p54 binds the 5' splice	(Kameoka et al., 2004)
	Since both PSF and p54 interacts with the C-terminal domain unit of the largest sub-unit of RNAP II and/ or the snRNPs, it was suggested that these proteins might be the link between the coupled transcription and splicing mechanisms	(Kameoka et al., 2004)
Transcriptional regulation	PSF and p54 directly bind the RNAP II (Emili A, 2002). A pre-incubation of purified PSF and p54 proteins with the CTD of the RNAP II results in an increase of RNA that immunoprecipitates with the CTD of the RNAP II	(Emili et al., 2002)
	Upon the silencing of PSF, there is a decrease in the expression of 298 cellular	(Song et al., 2005)

	genes and an increase in the expression of 184 cellular genes	
Hormonal transcriptional pathways	PSF and p54 are modulators of the transcription activity through human hormone receptor pathways	(Dong et al., 2007; Ishitani et al., 2003; Mathur et al., 2001; Shav-Tal and Zipori, 2002)
	In vitro immunoprecipitation assays showed that PSF, p54 and PSP1 bind to the activation function region of the androgen receptor	(Ishitani et al., 2003)
	PSF and p54 represses the transcription downstream of the progesterone and the androgen receptor pathway	(Dong et al., 2005; Dong et al., 2007)
	PSF directly binds the DBDs of the thyroid hormone receptor and the retinoid X receptor and acts as a transcriptional repressor	(Mathur et al., 2001)
	The N terminal region of PSF also binds to the IGFRE of the P450ssc gene and acts as a transcriptional repressor on the steroidogenesis	(Urban et al., 2000; Urban et al., 2002)
	PSP1, PSF and p54 form part of a complex of androgen receptor-binding proteins	(Ishitani et al., 2003)
	PSP1, p54 and PSF are present in the mouse Sertoli cells and might act as transcription regulators through a pathway downstream of an androgen-receptor	(Kuwahara et al., 2006)
Other physiological function	The DBHS are involved in the regulation of circadian cycle	(Brown et al., 2005; Duong et al., 2011)
	In a mouse model, PSP1 has been shown to have a role in spermatogenesis	(Myojin et al., 2004)
	p54 has a role in the pathway allowing for the regulation of the cyclic AMP-dependent glucocorticoid synthesis. A knockdown of p54 results in impairment to expand the cellular pool of cAMP	(Amelio et al., 2007; Lu and Shenk, 1999)
RNA transport and RNA stability	Molecular screen in embryonic mouse testis, *in vitro* immunoprecipitation and binding curves assays showed that PSP1 is a partner of importin *alpha* 2 *In vitro* immunoprecipitations and binding curves assays showed that PSP1 interacts with importin *alpha* 6	(Major et al., 2015)
	Frequently, prior to RNA termination, the pre-mRNA3' is cleaved, degraded and polyadenylated, PSF and p54 binds to the	(Kaneko et al., 2007)

	exonuclease XRN2, which is a major player in the degradation of 3'cleaved RNA. Knockdown against p54 interfered with the recruitment of XRN2 and therefore inhibited the transcription termination	
	PSF is involved in the retention of hyperedited mRNA	(Zhang and Carmichael, 2001)
	PSF, p54 and PSP1 are involved in the stabilization of the NEAT1 RNA	(Mao et al., 2011a; Naganuma and Hirose, 2013; Sasaki and Hirose, 2009)
	PSF, PSP1, p54 and Matrin 3 bind strongly to hyperedited RNAs, which allows for their sequestration in the paraspeckles and the regulation of the cellular RNA metabolism	(Chen and Carmichael, 2009; Zhang and Carmichael, 2001); Hundley and Bass, 2010)
	p54 stabilizes several phosphodiesterase mRNA by enhancing their contact with the exoribonuclease XRN2	(Lu and Sewer, 2015)
Cellular immunity	PSF and p54 bind to cellular DNA and regulate the transcription of proinflammatory cytokines mRNAs	(Buxade et al., 2008)
Cellular cycle	In mouse cells treated with the DNA damaging chemical mitomycin C, the knockdown of PSF resulted in a 13% rise of the number of cells in G2/M cell cycle as well as an increase of dead and aneuploid cells	(Rajesh et al., 2010)
	In cells knocked down for PSP1 and treated with methyl methanesulfonate, there were more apoptotic cells compared to the cells only treated with methyl methanesulfonate	(Gao et al., 2016)
	Knockdown of p54 also interfered with the DNA double-strand break repair pathway	(Mahaney et al., 2009)
	In mouse cells treated with the DNA damaging chemical mitomycin C, the knockdown of PSF resulted in a 13% rise of the number of cells in G2/M cell cycle as well as an increase of dead and aneuploid cells.	(Mahaney et al., 2009)
	In the mouse cells with a knockdown of PSF, there was a reduction of 8-10% cells in G1 phase and an increase of	(Rajesh et al., 2011)

	respectively 5 and 6% of cells entering S and G2/M cell cycle	
	In HeLa cells, the knockdown of PSP1 results in a deregulation of the cell cycle: more cells were in the G2/M phase and some cells had two or more nuclei, which is reminiscent of a mitotic catastrophe	(Gao et al., 2016)
	The upregulation of PSP1 in HeLa cells resulted in an increase of the number of cells entering the G2/M phase	(Gao et al., 2016)

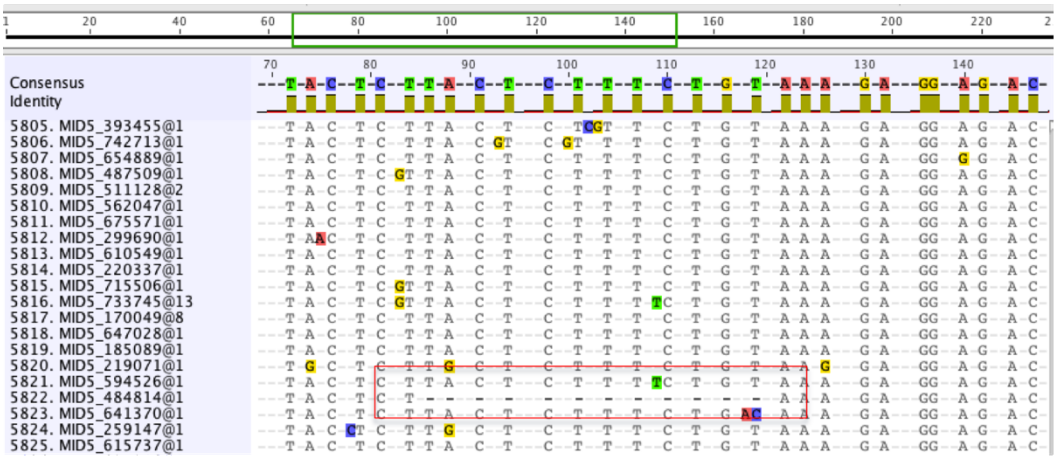
Appendix III Primer sequences

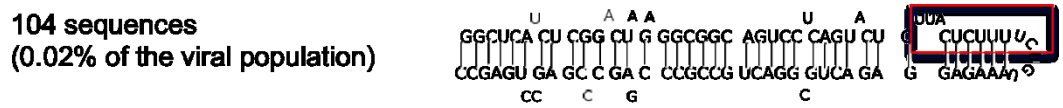
RNA species and polarity of the primer	**Primer sequence**
R199G Forward	5'GGAATTCTAATACGACTCACTATAGGGACTGCTCGAGGAT TCTTCTCTCC-3'
R199G Reverse	5'-ACATCCCCTCTCGGGTAC-3'
R199GForward-T7 promoter (T7 sequence underlined)	5'GAATTCTAATACGACTCACTATAGGGACTGCTCGAGGATCTCT TCTCTCC-3'
Deep-sequencing primer R199G Forward viral population	5'C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGATCAGA*C*A* C*GACTGCTCGAGGATCTCTTCTCTCCC-3'
Deep-sequencing primer R199G Reverse viral population	5'CCTATCCCCTGTGTGCCTTGGCAGTCTCAG[54]CCTCTCGGGTACT GATCCTCCCCCCGCGTCTCCTCG-3'.
Deep-sequencing primer R199G Forward control population	5'C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGACGAGT*G*C* G*TACTGCTCGAGGATCTCTTCTCTCCC-3'
Deep-sequencing primer R199G Reverse control population	5'CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCCTCTCGGGTACTG ATCCTCCCCCCGCGTCTCCTCG-3'.
HDV-ribozyme Forward	5'CCCTCGGTAATGGCGAATG-3'
HDV-ribozyme Reverse	5'CCCAGTGAATAAAGCGGGTT-3'
Beta-2 microglobulin Forward	5'GGCTATCCAGCGTACTCCAA-3'
Beta-2	5'TCACACGGCAGGCATACTC-3'

microglobulin Reverse	
NEAT1v2 Forward (Imamura K et al, 2014)	5’GATCTTTTCCACCCCAAGAGT-3’
NEAT1v2 Reverse (Imamura K et al, 2014)	5’CTCACACAAACACAGATTCCA-3’
IL8 Forward (Imamura K et al, 2014)	5’AGACAGCAGAGCACACAAGC-3’
IL8 Reverse	5’ATGGTTCCTTCCGGTGGT-3’

The phosphorothioate modifications are indicated by an asterix (*).

Appendix IV Deletion of nucleotides for 0.02% of the viral population sequences





(Upper figure) A part of the alignment of sequences of HDV population is represented and the red rectangle illustrates an example of a sequence with a deletion in the region of initiation of the transcription.
(Lower figure) In the HDV population replicating in 293 cells (including the unique sequences), 0.02% of the viral population are sequences with a deletion in the region of initiation of transcription (indicated by the red rectangle).

Appendix V Review of the characteristics of cellular double and single stranded DNA and RNA motifs that interact with PSF, p54 and PSP1

DBHS protein	Nucleic acid bound	Nucleic acid sequence and structure	Host	Reference
DRB of p54	DNA	purine rich motif **5'-AGGGA-3'** was found in a majority of sequences	long terminal repeats region in the enhancer element of intracisternal A, folded around histone octamers particles , in human cells	(Basu et al., 1997)
PSF	DNA	**5'-CTGAGTC-3' from sequence 5'-ATCCTGAGTCTGGGAGGGGCTGTGTGGGCC-3'** The mutation to 5'TTGGAGTC" or 5'CCTGAAAT3' abrogates the interaction with PSF	GC box of IGFRE of the porcine P450scc gene	(Urban et al., 2000)
PSF and p54	ss RNA	Enrichment in purines (50 to 67%) consensus sequence: **5'-UGGAGAGAGGAAC-3'**	Synthetic sequences	(Peng et al., 2002)
PSF, p54 or heterodimer of PSF-p54	ds RNA	RNA with a stem loop structure	purine rich sequence close to the 3' single stranded U5 snRNA, the stem 1b conserved among eukaryotes	(Peng et al., 2002)

PSF-p54	ss RNA	AU-rich elements mRNA (eg:proinflammatory cytokines mRNA such as TNF *alpha*)	In Jurkat T cells and 293HEK cells	(Buxade et al., 2008)
PSF	ss RNA	mRNA PAI-RBPI mRNA *alpha* catenin mRNA NF2 mRNA LKB1 mRNA tubulin6	mRNA from MDCK cells	(Figueroa et al., 2009a)
PSF and p54	RNA	rich in inosines and G	Mouse cells (MEF, NIH 3T3, C127I, mouse liver)	(Prasanth et al., 2005)
PSF and p54	RNA	rich in inosines and G	HIV-1	(Zolotukhin et al., 2003)
PSF and p54	RNA	rich in inosines and G	Mamma-lian cells	(DeCerbo and Carmichael, 2005)
Heterodimer PSF-p54	ds RNA	A-I enriched, edited RNAs	Synthetic RNAs and proteins from HeLa cells	(Zhang and Carmichael, 2001)
p54, unusual through three PIRs, PIR 1, PIR2 and PIR3	ss RNA	highly structured NEAT1	Mouse tissues	(Clemson et al., 2009; Murthy and Rangarajan, 2010)
Non - phosphorylated form of p54	ss RNA	poly(A), poly(C) and poly(U) synthetic RNAs, poly(G)	Synthetic RNAs	(Bruelle et al., 2011)
phosphorylated form of p54	ss RNA	poly(G)		(Bruelle et al., 2011)
Both	RNA	RNA rich in G, ssRNA with high structure	NEAT1	(Bruelle et al.,

phosphorylated and non-phosphorylated forms of p54				2011)
PSF	ss RNA		mRNA	(Figueroa et al., 2009a)
PSF	DNA		oncogenes	(Figueroa et al., 2009a)
PSF DRB	ss DNA	**5'-** **GCCTTCTGCAAAGAAGTCTTGCGCATC** **TTTTGTGAAGTTTATTTCTAGCTTTTTG** **ATGCTG-3'**	Mouse oncogene GAGE6	(Song et al., 2005)
PSF RRM	ss RNA	polypyrimidine tracts, pbt-A **5'-CAGCUGCCCUGCCUCCCACUCC-3'**	VL30 mouse retrotransposon	(Song et al., 2004; Song et al., 2005)
PSF RRM	ss RNA	pbt-B **5'-CAGCUUCUCAUUCCCCUGUCCCUCCCAUCC-3'**	VL30 mouse retrotransposon	(Song et al., 2004; Song et al., 2005)
p54	RNA	Hairpin RNA structure	EBV origin of replication in RNA	(Cao et al., 2015)
PSF	ss RNA	mRNA, "polypyrimidine tract at 3' of splice site" RNA highly structured and contains stem-loops structure in this region	HBV	(Heise et al., 2006) (Kidd-Ljunggren et al., 2000)
PSF	ss RNA		HIV	
PSF and p54	RNA	Highly structured INS	HIV RNA	(Nathans et al., 2009)
Heterodimer PSF/p54	RNA	enriched in INS which have a high A/U content (Schwartz et al., 1992) The viral genome contains hairpins and stem-loops domains	Gag mRNA of HIV	(Zolotukhin et al., 2003; Kula et al., 2013)
Heterodimer	RNA	highly structured and form hairpin and stem loops domains (Jain C and Belasco J G, 2001)	Rev-responsive	(Zolotukhin et al., 2003)

PSF/p54			element-containing mRNA	
PSP1	RNA	RRE (351 nt) RNA with high secondary structure, with a stem loop structure	HIV-1 sequence	(Mann et al., 1994; Naji et al., 2012)
PSP1	RNA	mRNA	293HEK eell line	(Baltz et al., 2012)
PSP1	RNA	Highly structured RNAs MALAT1 and NEAT1	Human cells (MCF7 and BJ cell lines)	(Clemson et al., 2009; West et al., 2014)

ss: single-stranded

ds: double-stranded

Appendix VI

Fig. 1. Schematic representation of the HDV RNA genome. The circular rod-like structure of the HDV genome is illustrated. The white arrow indicates the location of the HDAg ORF (i.e. initiating at nucleotide 1598). The Rz-G and Rz-AG boxes indicate the locations of both genomic and antigenomic polarities of the HDV ribozymes, respectively. A black arrow indicates the reported site of initiation of transcription from genomic HDV RNA (Beard et al., 1996; Gudima et al., 2000; Abrahem and Pelchat, 2008). The right terminal domain of genomic HDV RNA involved in viral replication is indicated by the gray rectangle (i.e. nucleotides 1541 to 60 of the genomic polarity). The sequences analyzed in this study correspond to nucleotides 1566 to 18 of the genomic polarity (i.e. between the locations indicated by the inverted fonts). The numbering is in accordance with (Kuo et al., 1988).

Several studies indicated a role for the right terminal domain of genomic HDV RNA in viral replication (Fig. 1; gray rectangle). In infected cells, the 5′ end of HDAg mRNA localizes in this region (i.e. position 1630; arrow on Fig. 1; (Gudima et al., 2000)). HDAg mRNA is post-transcriptionally processed with a 5′-cap and a 3′-poly (A) tail (Gudima et al., 1999, 2000), which suggests RNAP II involvement in the production of this mRNA. In vitro, this region acts as template to initiate antigenomic RNA synthesis, and the transcription reaction is inhibited by an antibody raised against the C-terminal domain of RNAP II (Abrahem and Pelchat, 2008). RNAP II forms an active pre-initiation complex on the right terminal domain of genomic HDV RNA, and this complex contains the same general transcription factors as those found on a typical DNA promoter (Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008). The TATA-binding protein, alone or within the TFIID complex, directly binds the RNA promoter, and was proposed to be required to nucleate the RNAP II pre-initiation complex (Abrahem and Pelchat, 2008). Notably, mutations affecting the secondary structure of this region were reported to decrease HDV RNA accumulation in cell culture (Beard et al., 1996; Wu et al., 1997; Gudima et al., 1999), lower RNAP II affinity (Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008), and affect RNAP II transcription initiation efficiency in vitro (Beard et al., 1996; Abrahem and Pelchat, 2008).

Although many studies have investigated DNA promoter recognition by RNAP II, little is known regarding how this enzyme recognizes an RNA template. Analogous to what is observed on DNA promoters (reviewed by (Baumann et al., 2010)), essential HDV RNA features (i.e. sequence and secondary structure) should be conserved and selected for during viral replication. By taking advantage of next-generation sequencing technology, the goal of this study was to investigate the positions of covariation and nucleotide conservation in a large population of heterogeneous HDV RNA sequences to define the selected features on the right terminal domain of genomic HDV RNA. We generated 2351 new HDV variants of this region derived from 473,139 sequences obtained by high-throughput 454 sequencing and originating from an HDV population replicating in a cellular system. We developed a pipeline to filter, align and analyze sequence conservation and covariation of this region from the population of sequences. Our results indicated the polymorphic nature of this segment of HDV, by showing that it accumulates as a population of different sequences. Despite sequence heterogeneity, our analyses revealed the conservation of the rod-like conformation of this region and identified conserved nucleotides at the tip of the rod-like structure, near the proposed transcription initiation site. These conserved features, which are also found on sequences from HDV variants isolated from various hosts, are likely important for HDV replication by RNAP II, and will be useful at identifying other RNA promoters for RNAP II.

Results

High-throughput sequencing of the right terminal region of genomic HDV RNA from a cellular system

To investigate the features of the right terminal domain of genomic HDV RNA involved in replication (Fig. 1; gray rectangle), we needed a system where the selective pressure was mainly on viral replication. We used the HDV replication system previously developed by Chang et al. (2005). In this system, 293 cells contain a replicating HDV RNA genome with a frame-shift deletion in the HDAg ORF, and allow the synthesis of HDAg-S under the control of a promoter inducible by tetracycline (Chang et al., 2005). Because a low level of HDAg-S is produced in the cells without induction, basal HDV replication is possible for several months, and HDV RNA genomes capable of replication are amplified upon tetracycline induction (Chang et al., 2005).

The 293-HDV cells were grown for more than a year without induction of HDAg-S expression to allow the accumulation of mutations on the HDV RNA genome compatible with viral replication (Fig. 2A). HDV RNA production was then induced with tetracycline to amplify functional or even ameliorated HDV genomes. Two days after induction, total RNA was extracted, reverse transcribed (RT) using random primers, and HDV cDNA was amplified by PCR. Because a 199 nt fragment from the right terminal region of genomic HDV, including ~60 nt of HDAg ORF, was previously reported to act as an RNA promoter for RNAP II (Beard et al., 1996; Abrahem and Pelchat, 2008), primers designed to specifically amplify this region were used (Fig. 1; gray rectangle), as described previously (Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008). The sample quality was verified by agarose gel electophoresis, and the identity of the sequence was confirmed by Sanger sequencing (data not shown). To control for mutations introduced during either the RT-PCR or the deep-sequencing protocol, a genomic HDV RNA with the same sequence (hereinafter referred to as reference sequence) was synthesized by in vitro transcription with T7 RNAP and similarly processed. Both populations were tagged with a different bar code during the PCR for multiplexing, mixed at a ratio of about 1:100 control:viral population, and sent for deep-sequencing using the 454 Roche technology. We obtained 2510 and 747,158 readings for the control and the viral population, respectively.

Refinement of the populations

As reported previously, readings obtained by the 454 Roche technology usually include unrelated sequences, are heterogeneous in length and contain base calling errors (Gorzer et al., 2010; Beerenwinkel et al., 2012). Consequently, we developed a pipeline to refine the readings by performing several filtering steps. To remove readings unrelated to HDV, we calculated identity scores by comparing each reading to both polarities of the reference sequence using ClustalW (Thompson et al., 1994). Most of the readings were of the expected length (i.e. ~200 nt), and were ~75% identical to the reference sequence (Fig. 2B and C). For the viral population, there was also a smaller cluster of readings of approximately 64 nt, but most of these readings had low identities to HDV with large variations in their identity scores (Fig. 2C). Inspection of several sequences in this population revealed that they were chimeras composed of HDV and unidentified sequences, which is consistent with the generation of chimeras caused by

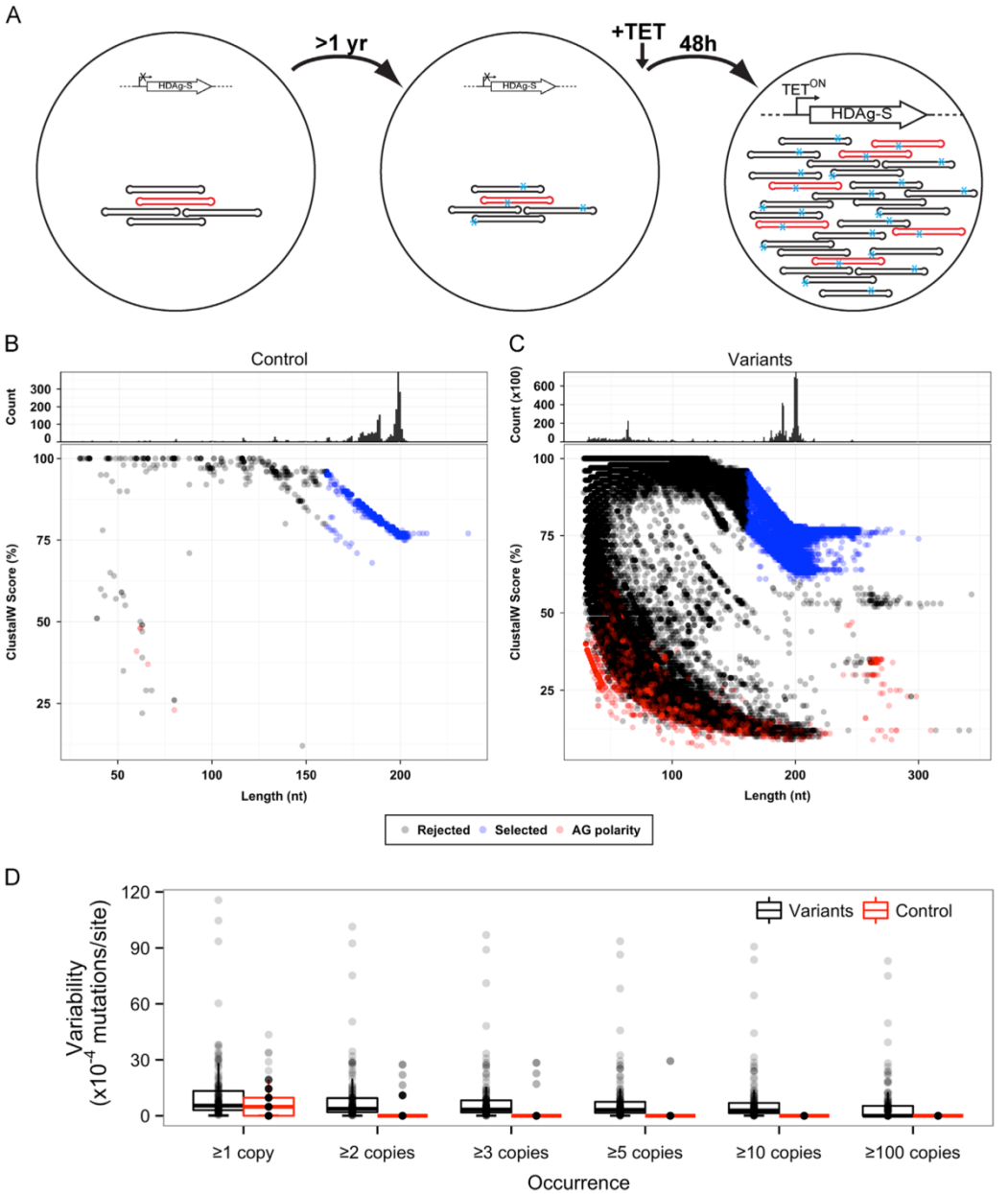


Fig. 2. Refinement of the sequence libraries generated by high-throughput sequencing of an HDV population. (A) Overview of the cellular system allowing HDV replication. A 293 cell line has been stably transfected with an HDV RNA genome containing a frame-shift deletion in the HDAg ORF and a plasmid allowing the production of the small antigen under the control of a promoter inducible by the addition of tetracycline. This cell line has been maintained for more than a year, allowing HDV replication at a basal level and the accumulation of mutations. Addition of tetracycline allowed HDV RNA production and amplification of functional or even ameliorated HDV genomes. (B) and (C) Filtering of the reading obtained by deep-sequencing using 454 technology according to the sequence length and percentage of identity to the reference sequence, for both the control (B) and the sequences amplified from 293-HDV cells (C). Top parts represent the number of sequences sorted according to their lengths. Bottom parts represent the percentage of identity of each reading to the reference sequence, as calculated by ClustalW (Thompson et al., 1994). Black and red dots indicate sequences with higher identities to the genomic and antigenomic polarity of the reference sequence, respectively. Sequences selected for further analysis are represented by blue dots. These sequences are at least 160 nt long and are at least 60% identical to the reference sequence. (D) Reduction of the background nucleotide variability by removal of sequences with low occurrence. Box plot representation of the nucleotide variability at each position calculated for sequences occurring at least 1, 2, 3, 5, 10 and 100 times. Black and red boxes indicate variants and control population, respectively.

non-specific amplification during deep-sequencing, as previously reported (Gorzer et al., 2010).

To obtain the information on the conservation of the secondary structure of this region, we only considered sequences of length longer than 160 nucleotides (Fig. 1; between nucleotides 1566 and 18) and at least 60% identical to the reference sequence (Fig. 2B and C; blue circles). Using these two constraints, 490,183 and 2070 sequences were selected for the viral population and the control, respectively. Furthermore, all of them were of genomic polarity, demonstrating the specificity of the primers used during amplification for this polarity of HDV RNA (Fig. 2B and C). Then, we performed pairwise alignment of each reading to the reference sequence, and calculated occurrences by clustering identical sequences (Fig. S1).

The cDNA amplification process and base calling error during deep-sequencing are known to generate apparent mutations that do not reflect selected variation (Beerenwinkel et al., 2012). To account for these errors, we calculated the nucleotide variability at each position, for both populations of sequences, and at different number of occurrences (Fig. 2D). Removal of sequences occurring less than three times greatly diminished nucleotide variability in the control without drastically affecting the number of different sequences in the viral population sample (Fig. 2D). Only seven positions varied in this subset of the control population, and none of these positions showed significant variability in the viral population (data not shown). The mutations in the control sample were likely generated by the additional PCR and/or transcription reaction used to generate the RNA species. Accordingly, only sequences occurring at least three times were kept for subsequent analysis. With these datasets, the overall nucleotide variations were 1.2×10^{-4} and 8.1×10^{-4} mutations/site for the control and the viral population, respectively. Altogether, we retained 1761 (85.07%) and 473,139 (96.52%) sequences representing 49 and 2351 different and recurring variants from the control and the viral population, respectively.

The right terminal region of genomic HDV RNA exists as a heterogeneous population in 293 cells

Based on the occurrence of the sequences, we found that one sequence was highly enriched and represented 76.3% of the variants obtained (i.e. 360,863 readings; Fig. S1). Interestingly, this sequence corresponded to the original HDV variant transfected into the cells (Chang et al., 2005). Despite of this, the number of different and recurring sequences obtained in the viral population sample suggested a larger sequence space generated during HDV replication. To evaluate the sequence space in our samples, neighbor-joining phylogenetic trees showing the genetic diversity of the different sequences composing both populations were generated. The trees were rooted on the reference sequence and plotted as circular dendrograms (Fig. 3). On the dendrograms, the size of the clusters in the trees is proportional to the occurrence of the sequences composing this cluster ($\log_2$ relationship). For the control, the clusters were phylogenetically close (Fig. 3, inset). The tree for the viral population from the 293 cells replicating HDV RNA was more heterogeneous, and numerous clusters with a high amount of sequences were phylogenetically distant from the reference sequence, indicating a larger sequence space. These results are in accordance with the accumulation of mutated variants during HDV replication, giving rise to a heterogeneous population (Wang et al., 1986; Chao et al., 1990, 1991).

Analysis of the variations found in the viral population

We calculated the nucleotide composition per position in order to determine the localization of position-specific variability and

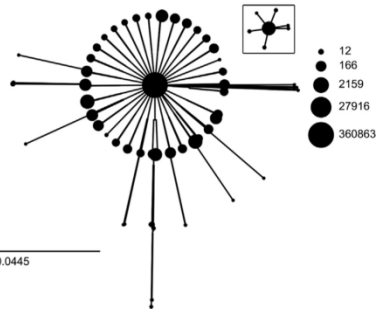


Fig. 3. Evaluation of the sequence space occupied by both the control population and the viral population replicating in 293 cells. Neighbor-joining phylogenetic trees rooted on the reference sequence were generated for both populations, and plotted as circular dendrograms. The sizes of the clusters are proportional to the occurrence of the sequences composing this cluster (examples of the $\log_2$ relationship between the surface of the circle and the number of sequences are found on the right side). The scale bar on bottom left indicates the distance as substitutions per site. The inset corresponds to the sequence space occupied by the control population, using the same scale as the viral population.

selective mutations. Variation was not homogeneous but varied according to the nucleotide position (Figs. 4 and S2). To discriminate between significant variability and background variations, we used four outlier tests (GESD, boxplot, medmad and shorth) to identify positions that appear to deviate from background variations (pooled as a "gray" zone on Fig. 4). Eleven positions showed significant variability in the four tests used, whereas the variability of 13 other positions was significant in at least one of the tests (Fig. 4; in red and blue, respectively). None of these 24 positions had significant variation in the control sequences. Because the variations were not homogeneous but fluctuated according to the nucleotide position, we recalculated the nucleotide variation rate. The variation rate of the viral population for these 24 positions was calculated to be 29.5×10^{-4} mutations/site, which is 24-fold the background variation rate calculated for the control (i.e. 1.2×10^{-4} mutations/site). For the other positions, the variation rate was 3.3×10^{-4} mutations/site, which is in the same order as background variation derived from the control. Noteworthily, the sequence located at the tip of the rod-like stucture (i.e. from position 1632 to 1557) was the most conserved. This conservation suggests that the sequence of this region might be required for the initiation to take place or for promoter recognition by the host transcription machinery. However, we cannot exclude the possibility that this motif might be associated with another activity unrelated to transcription initiation.

The analysis of the type of selected nucleotide changes revealed that 91.9% were transitions (either purine -> purine or pyrimidine -> pyrimidine) and 8.1% were transversions (purine -> pyrimidine or pyrimidine -> purine) (Fig. S2). Noteworthily, the highest nucleotide variation corresponded to a A->G transition located at position 1597. This selective mutation was observed in ~1% of the viral population (i.e. 4583 readings). This mutation is unlikely to be caused by experimental error since it was not found in the control population. This position is the second nucleotide of the anticodon CAU, which corresponds to the AUG initiation codon of HDAg on the HDV mRNA. Since HDAg-S, which is required for HDV replication, is provided *in trans* in the cellular system we used, a decrease of the selective pressure for the sequences able to produce the HDAg mRNA was expected. Interestingly, this mutation allows the conservation of the RNA secondary structure at this location by allowing Wobble base-pairing of the G with the U of the lower

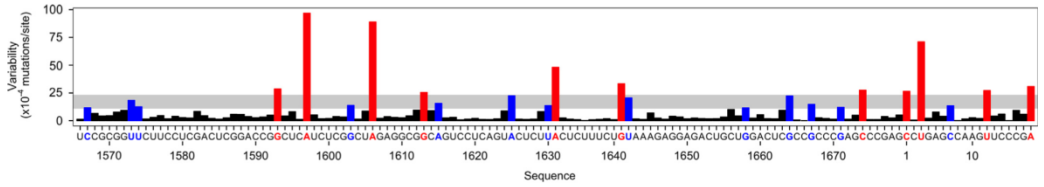


Fig. 4. Position-specific variability of the right terminal domain of genomic HDV RNA obtained from high-throughput sequencing. Representation of nucleotide variability for each position obtained from the refined alignment. The consensus sequence displayed corresponds to a region from nucleotides 1566 to 18 of the genomic polarity of HDV RNA. Four outlier tests were used to identify positions that appear to deviate from background variations, and their cut-offs was used to define the "gray zone". The calculated cut-offs were 15.58×10^{-4}, 23.33×10^{-4}, 10.79×10^{-4}, 11.19×10^{-4} mutations/site for GESD, boxplot, medmad and shorth, respectively. Blue and red indicate the position with significant variability in at least one or all four of the tests used, respectively. The U residue at location 1638 (gray on Fig. 5) was not used in the analysis due to high variability caused by the homopolymer effects during high-throughput 454 sequencing (Huse et al., 2007).

strand of genomic HDV RNA, suggesting the importance of the base pair at this location.

We next investigated conservation of the base pairs of this region. As a first step, we used the most energetically stable predicted secondary structure to assess base pair covariation. Then the secondary structure was manually adjusted based on the conservation derived from our dataset and is presented in Fig. 5A. The secondary structure derived from the conserved base pairs is also in accordance with a previously reported structure derived from in vitro nuclease mapping (Beard et al., 1996). We calculated the frequencies and compositions for each base pair across the alignment (Fig. S3). Sequences derived from the control allowed to establish a baseline for base pair variability of 2.1×10^{-4} mutations/pb (i.e. occurring 100 times). With this cut-off, we found that most of the nucleotide changes corresponded to transitions that enable the maintenance of the base pairs of either genomic or antigenomic polarities of HDV RNA (Fig. 5A). Interestingly, the majority of the variations are transitions generating G-U Wobble base pairs on antigenomic strand (Fig. 5A, CA and AC with a yellow background on genomic polarity), suggesting the importance of the secondary structure for this polarity. However, we cannot exclude the possibility that non-canonical C-A base pairs might also form on genomic strand.

Using the same approach, we calculated the frequencies and compositions of each bulge across the alignment occurring with variability of at least 2.1×10^{-4} mutations/site (i.e. occurring 100 times; Fig. 5B). We identified three bulges containing enriched mutations in their composition: U_{1599}/C_1C_2 -> U_{1599}/U_1C_2 (24.88×10^{-4} mutations/site; 1177 readings), A_{1606}/G_{1671} -> G_{1606}/G_{1671} (85.94×10^{-4} mutations/site; 4050 readings), and $U_{1629}U_{1630}A_{1631}$ -> $U_{1629}U_{1630}G_{1631}$ (46.01×10^{-4} mutations/site; 2177 readings). Our analysis also suggests the formation of a homopurine pair between A_{1606} and G_{1671}. Furthermore, the bulge at the initiation site always contains at least one uridine. Conservation of this uridine residue at this location is probably necessary for efficient initiation of complementary strand synthesis, since RNAP II is known to preferentially initiate transcription with purine residues (Baumann et al., 2010). Unfortunately, preliminary prediction of non-canonical base pairs within these bulges using the isostericity matrices was inconclusive due to the high conservation of this region (Leontis et al., 2002).

Conserved features of the right terminal stem-loop region of genomic HDV RNA in isolated variants

The previous analysis was performed on a viral population replicating in a specific cellular system. To assess the biological significance of our results and variations caused by the use of different hosts that might have different selection pressures, we analyzed both the positions of nucleotide conservation and covariation by extracting this region in sequences corresponding to HDV variants isolated from various hosts. Based on sequence analysis, the *Deltavirus* genus was previously classified into several major clades (Deny, 2006). We selected only the sequences from clade 1 since the viral population analyzed above was generated from a variant from this clade. Also, we decided to keep identical sequences in order to take into account selected sequence fitness. The sequences were extracted from the Subviral RNA database (Rocheleau and Pelchat, 2006), aligned using ClustalW (Thompson et al., 1994), and analyzed as above (Fig. S4).

Nucleotide comparison of the proposed initiation site for the transcription from this domain (i.e. nucleotide 1630; (Gudima et al., 1999; Abrahem and Pelchat, 2008)) and the surrounding nucleotides (i.e. nucleotides 1592 to 8 of the genomic polarity) in all the variants analyzed revealed a sequence heterogeneity pattern similar to what we observed in the viral population isolated from 293 cells. Specifically, the sequence is the most conserved at the tip of the rod-like structure with variation mostly upstream from the tip (Fig. 6A). We also calculated frequencies and compositions for each base pair across the alignment, as performed above (Fig. S5). The number of base pair variations was reduced as compared to the viral population in 293 cells, likely due to the small number of sequences available (i.e. 40 variants). Despite this, the majority of the variations are transitions allowing base pairs on either or both polarities of the HDV RNA genome, including generation of G-U Wobble base pairs on antigenomic strand (Fig. 6B, CA and AC with a yellow background on genomic polarity), analogous to our observation from the sequences derived from the viral population in 293 cells. In addition, several of the mutations observed in the bulges derived from the viral population were also found in these clade 1 variants, including the enrichment of purines at position 1606 and 1671 (Fig. 6B, blue rectangles).

Discussion

Previous studies indicated a role for the right terminal region of genomic HDV RNA in viral replication. This region includes the site of transcription initiation for HDAg mRNA, binds an active RNAP II pre-initiation complex, acts as template to initiate antigenomic RNA synthesis in vitro, and mutations affecting the rod-like conformation of this region decrease both RNAP II affinity/initiation and HDV RNA accumulation in cells (Beard et al., 1996; Gudima et al., 1999; Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008). Here, we took advantage of next-generation sequencing technology to generate 473,139 new HDV sequences of this region from a cellular system in which the selective pressure was mainly on viral replication.

Because HDAg-S could not be produced by the mutated HDV RNA genome in this replication system, but provided *in trans* by the cells, we were expecting reduced selective pressure on both

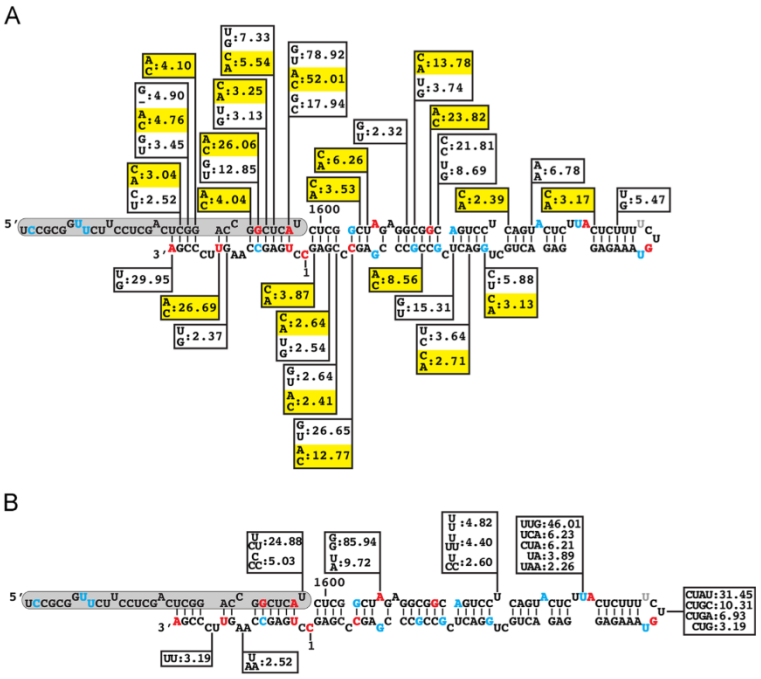


Fig. 5. Covariation analysis of the right terminal domain of genomic HDV RNA obtained from high-throughput sequencing. The covariation variability of every base pairs (A) and single-stranded region (B) was calculated and displayed on the consensus RNA secondary structure. Blue and red nucleotides indicate the position with significant variability, as determined in Fig. 4. Yellow background indicates transitions generating C-A on genomic HDV RNA. The gray rectangles represent the 5′-end of HDAg ORF. All numbers correspond to $\times 10^{-4}$ mutations/site. The gray U residue at location 1638 was not used in the analysis due to high variability caused by the homopolymer effects during high-throughput 454 sequencing (Huse et al., 2007).

the sequence and the secondary structure of the region corresponding to either the HDAg ORF or its promoter. However, our analysis indicates that both the sequence and the secondary structure of this segment of HDV are very conserved. This suggests that, in addition to its proposed role as promoter for HDAg transcription, this region is also involved in HDV replication. The highest nucleotide variation corresponded to a A- > G transition at position 1597, which was observed in ~1% of the viral population. This mutation disrupts the initiation codon of HDAg but is still predicted to allow base pairing with the opposite strand. The importance of a base pair at this location is in agreement with a recent study reporting that mutations disrupting the base pairing at this location hinder HDV replication (Liao et al., 2012).

Analysis of the sequences corresponding to the right terminal region of genomic HDV RNA revealed that this region is less heterogeneous than expected based on previous reports on isolated HDV variants (Wang et al., 1986; Chao et al., 1990, 1991). One of the sequences was highly enriched and represented 76.3% of the variants obtained (i.e. 360,863 readings), suggesting enhanced fitness for this sequence, which also corresponded to the original HDV variant transfected into the cells (Chang et al., 2005). Despite this, comparison of the sequence space between the viral and the control populations revealed that this region of the HDV RNA genome is heterogeneous in 293 cells, consistent with the notion that HDV RNA forms a population of different sequences due to the infidelity of a "DNA-dependent" RNAP acting on an RNA template (Wang et al., 1986; Chao et al., 1990, 1991). In total, 2351 different and recurring sequences were found in the sample derived from the viral population. However, due to the approach used, we were not able to distinguish between variations that occurred over the year of replication from those following HDAg-S induction by tetracycline. Interestingly, nucleotide variations for the viral population were not distributed evenly and 24 positions with higher variability were identified. The variation rate of these "hot spots" was calculated to be 29.5×10^{-4} mutations/site and account for the larger sequence space observed for the viral population. We cannot completely exclude the possibility that some of the sequence diversity observed might have been artificially generated during the protocols used. However, it is unlikely due to the short length of the cDNA fragment, and because throughout our pipeline, we used a control sample of the same sequence to establish cut-offs to account for the error-rate due to the experimental steps of the reverse-transcription, the PCR amplification and the deep-sequencing. Based on the control, we calculated that an overall variation rate of 1.2×10^{-4} mutations/site might be due to the protocols used.

Although it reflects the mutation rate during HDV replication in a non-physiological cellular system (i.e. 293 cells) where the antigen is provided in *trans*, our calculated variation rate for the viral population (i.e. overall 8.1×10^{-4} mutations/site) is in accordance with mutation rates calculated for other RNA viruses (i.e. 10^{-3}–10^{-5} substitutions/nt), which have a high polymerase error-rate, giving rise to a heterogeneous population (Domingo et al., 2012). It is also one order of magnitude higher than what is reported for RNAP II when acting on DNA templates (i.e. 10^{-5} substitutions/nt; (Cramer, 2004)). This suggests that RNAP II has

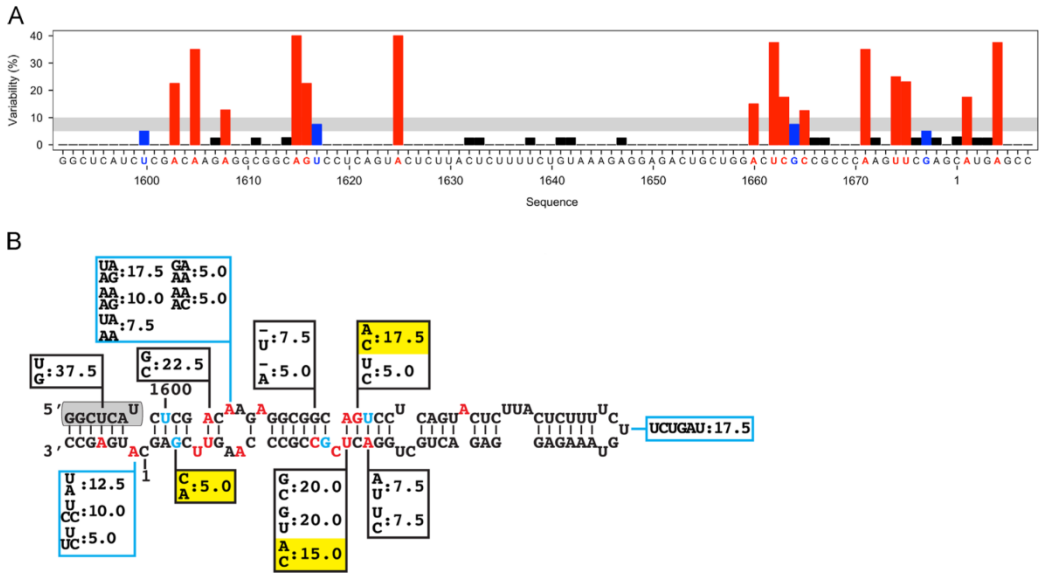


Fig. 6. Conserved RNA features of the right terminal region of genomic HDV from clade I variants. (A) Representation of nucleotide variability for each position obtained from the secondary structure alignment of 40 clade I HDV variants extracted from the Subviral RNA Database (Rocheleau and Pelchat, 2006). The consensus sequence displayed corresponds to a region from nucleotides 1592 to 8 of the genomic polarity of HDV RNA. (B) The covariation variability of every base pair was calculated and displayed on the consensus RNA secondary structure. Yellow background indicates transitions generating C-A on HDV RNA of genomic polarity. The gray rectangles represent the 5′-end of HDAg ORF. The blue rectangles indicate covariation variabilities in single-stranded regions. All numbers correspond to percentage of mutations/site. Blue and red indicate the position with variability of at least 5% and 10%, respectively.

an increased mutation rate when acting on HDV RNA, which is also supported by studies showing that HDAg-S might accelerate forward translocation of the polymerase at the cost of fidelity (Yamaguchi et al., 2001). However, we cannot exclude the possibility that the observed variations might also be due to the activity of another protein. Previous analysis of the sequences of a few HDV RNA genome variants in the same cellular system estimated a variation rate of 2.1% changes/nucleotide/year reported for the complete HDV RNA genome, and attributed most of the mutations to adenosine deaminase acting on RNA (ADAR) activity (Chang et al., 2005). Interestingly, in this study, the sequence of the right terminal region was conserved. Our estimated variation rate is also lower than what was reported on complete HDV RNA sequences from sequential isolates (i.e. $2–3 \times 10^{-3}$ changes/nucleotide/year; (Weiner et al., 1988; Lee et al., 1992)), and what was calculated for viroids, small single-stranded circular RNA genomes similar to HDV but replicating in plants (i.e. 2.5×10^{-3} changes/site/replication cycle; (Gago et al., 2009)). However, the mutation frequencies calculated for viroids were derived from samples isolated in the context of a natural infection, with more selective pressure from their host.

More importantly, we found that the sequence at the tip of the rod-like structure of this region is very conserved in both sequences derived from the viral population in 293 cells (Fig. 4) and clade 1 variants (Fig. 6A). This region is composed of a stretch of pyrimidines upstream of the terminal loop, which is matched on the opposite strand by a region containing almost exclusively purines, allowing the conservation of the rod-like structure of this region. This is consistent with previous reports on both a decrease of HDV accumulation in cells and reduced RNAP II interaction by the inversion of the strands of the tip region (i.e. “flip” mutant), suggesting that in addition to the structure, sequence conservation is important for viral replication (Wu et al., 1997; Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008). It is also possible that the sequence of either or both strands serves as a binding site for either polypurine or polypyrimidine binding proteins. One candidate protein is PSF, a polypurine binding protein we recently reported to bind this region, and which is also known to associate with RNAP II (Emili et al., 2002; Greco-Stewart et al., 2006).

Most of the nucleotide changes were upstream of the tip and our results indicate that they were selected to maintain the rod-shaped secondary structure of either polarity of this region of HDV RNA. This suggests that the secondary structure of these regions is important for HDV replication/transcription and is in accordance with previous experiments in which mutagenesis disturbing the secondary structure of this region affected both HDV accumulation in cells and RNAP II binding (Beard et al., 1996; Wu et al., 1997; Gudima et al., 1999; Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008). Although, several specific nucleotides within this region were reported to be essential for high level of HDV accumulation in cells, in most of the cases the mutations introduced could also disrupt base pairing. Interestingly, our high-throughput sequencing of this region indicates that most of the selected nucleotide changes corresponded to transitions to maintain the secondary structure of the antigenomic polarity (i.e. generating G-U on antigenomic strand), suggesting the involvement of the antigenomic strand of this region in HDV replication. In support of this hypothesis, both strands of this region associate with RNAP II (Greco-Stewart et al., 2007), and a small 5′-capped HDV RNA of genomic polarity corresponding to this region was identified during a screening for small RNAs in cells replicating HDV (Haussecker et al., 2008). This small HDV RNA might represent a transcription product from antigenomic RNA. Additionally, an HDV cDNA fragment corresponding to this region was

reported to have bidirectional promoter activity, although it was never confirmed using RNA fragments (Macnaughton et al., 1993; Liao et al., 2012).

Conclusion

In this study, we used next-generation sequencing technology to define selected features on the right terminal domain of genomic HDV RNA from HDV variants isolated in cellular system in which the selective pressure was mainly on viral replication. We analyzed both the sequences and the secondary structural implication by investigating nucleotide conservation and positions of covariation in a dataset composed of 473,139 sequences representing 2351 new HDV variants for this region. We also corroborated our finding with sequences from HDV variants isolated in various hosts. Our analysis suggests a precise RNA secondary structure for this region and indicates the conservation of the nucleotides at the tip of the rod-like structure. Both features might be important for HDV replication, likely through the recruitment of RNAP II, and are in accordance with previous mutagenesis on this region of the HDV RNA genome. We also developed a pipeline to filter, align and analyze the sequences, which could be a useful strategy for future high-throughput sequencing analysis of sequence conservation and base pair co-variation in an RNA population. More importantly, the selected features identified in this study might be useful in identifying other RNA promoters for RNAP II, including in human RNAs.

Materials and methods

Cell culture and HDAg-S induction

The 293-HDV cells are 293 cells stably transfected with a plasmid encoding HDAg-S under the control of tetracycline and an HDV RNA genome deficient in HDAg-S production, and were kindly provided by John Taylor (Chang et al., 2005). The 293-HDV cells were grown at 37 °C with 5% CO_2 in DMEM supplemented with 10% calf serum (CS), hygromycin and blasticidin. Viral replication was induced upon addition of 1 μg/ml of tetracycline, and two days later the total RNA was extracted with Trizol (Invitrogen) according to the manufacturer's recommendations.

In vitro transcription of the control population

To serve as a control population to account for mutations introduced during the protocols used, a genomic HDV RNA, with the same sequence as the variant originally transfected into the cellular system used (Chang et al., 2005), was synthesized by in vitro run-off transcription using T7 RNAP (New England Biolabs; Pickering, Ontario, Canada; NEB), as previously described (Greco-Stewart et al., 2007; Abrahem and Pelchat, 2008). To generate the cDNA for the transcription reaction, PCR amplification was performed on a plasmid encoding a dimer of the HDV genome (pHDVd2) with both sense (5′–GAATTCTAATACGACTCACTATAGGG^{1541}ACTGCTCGAGGATCTCTTCTCTCC^{1564}-3′; underlined nucleotide sequence indicates T7 promoter) and antisense (5′-^{60}ACATCCCCTCTCGGGTAC^{43}-3′) oligonucleotides. After transcription, the DNA template was digested with DNase I (NEB) for 30 min at 37 °C and the RNA was fractionated by 7M urea denaturing polyacrylamide gel electrophoresis (PAGE) in 1XTBE buffer (100 mM Tris–borate, pH 8.3, 1 mM EDTA). The band corresponding to the control RNA was visualized by UV shadowing, excised, and eluted overnight in 500 mM ammonium acetate, 0.1% SDS. The RNA was then precipitated in ethanol, resuspended in H_2O, desalted by Sephadex G-50 columns (GE Healthcare), and precipitated in ethanol. The purified control RNA was resuspended in H_2O, quantified by spectrophotometry at 260 nm and stored at −20 °C.

Reverse-transcription and PCR

Both the control population and total RNA from 293-HDV cells were reverse transcribed with random primers according to the manufacturer's instructions (Biorad). cDNAs were then used as templates for PCR amplifications with Deep Vent polymerase (NEB). For both cDNAs, the antisense primer used was 5′–CCTATCCCCTGTGTGCCTTGGCAGTCTCAG^{54}CCTCTCGGGTACTGATCCTCCCCCCGCGTCTCCTCG^{19}-3′. The sense primers were C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGACGAGT*G*C*G*T^{1541}ACTGCTCGAGGATCTCTTCTCTCCC^{1565}-3′ and C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGATCAGA*C*A*C*G^{1541}ACTGCTCGAGGATCTCTTCTCTCCC^{1565}-3′ for the control and the viral population, respectively (the * indicate phosphorothioate modifications). The PCR products were purified from a 1.5% agarose gel (Qiagen), and the identity of the cDNAs was confirmed by Sanger sequencing (StemCore facilities, Ottawa Hospital Research Institute). One microgram of the viral population DNA and 10 ng of the control DNA were pooled, and sent for deep-sequencing using the Roche 454 GS FLX Titanium platform (McGill sequencing facilities, Genome Quebec). Raw sequencing data from both samples were deposited on the Sequence Read Archive of NCBI [SRA: SRR765851, SRR765852].

Analysis of HDV variants from deep-sequencing

For each sequence, the name of the sequence, the composition in nucleotides, the length and the sequencing quality score were stored in a database. The percentage of identity of each sequence to both polarities of the reference HDV sequence was calculated with ClustalW 2.1 (Thompson et al., 1994) and stored in the database. A cut-off of 160 nt of length and 60% identity was used to select the sequences for alignment with Mosaik 1-3.0 from the Marth laboratory (http://bioinformatics.bc.edu/marthlab/Mosaik). In house Perl-scripts were used to cluster the sequences based on identity, and to obtain statistics on nucleotide composition. In house R-scripts were used to analyze the correlation between the variability of the sequences and the number of identical sequences, and analyze both nucleotide composition and covariation. An in house R-script was used to detect hot spots of variability by using cutoff generated by selecting both the minimum and maximum of four outlier detection procedures included in the R package Parody (i.e. ("GESD", "boxplot", "medmad" and "shorth") (http://www.bioconductor.org/packages/release/bioc/html/parody.html). Neighbor-joining phylogeny of the sequences was performed with the R package APE (Thompson et al., 1994; Paradis et al., 2004), and the trees were drawn using a modified radial.phylog R-script from the package ADE4 (Dray and Dufour, 2007). Secondary structure prediction was performed with Mfold (Zuker, 2003).

Analysis of HDV variants from various hosts

The HDV sequences were taken from the Subviral RNA Database (http://subviral.med.uottawa.ca/; (Rocheleau and Pelchat, 2006)). The sequences were first aligned with ClustalW 2.1 (Thompson et al., 1994) and neighbor-joining phylogeny of the sequences was performed with the R package APE (Thompson et al., 1994; Paradis et al., 2004). The sequences clustering with known clade 1 variants were extracted and realigned with ClustalW 2.1. In house R- and PERL-scripts were used to analyze both the composition and nucleotide variation from the alignment, as performed with the HDV sequences generated from deep-sequencing.

Acknowledgments

This work was funded by a grant from the Natural Science and Engineering Research Council of Canada (NSERC Canada) awarded to M. Pelchat. The authors wish to acknowledge Alfredo Staffa of the genotyping platform of the McGill University and Genome Quebec Innovation Centre for his technical assistance. The authors would like to thank Professor Earl G. Brown, Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Canada, for giving valuable suggestions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.12.017.

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 Res. 36, 5201–5211.
Baumann, M., Pontiller, J., Ernst, W., 2010. Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. Mol. Biotechnol. 45, 241–247.
Beard, M.R., MacNaughton, T.B., Gowans, E.J., 1996. Identification and characterization of a hepatitis delta virus RNA transcriptional promoter. J. Virol. 70, 4986–4995.
Beerenwinkel, N., Gunthard, H.F., Roth, V., Metzner, K.J., 2012. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol. 3, 329.
Borderia, A.V., Stapleford, K.A., Vignuzzi, M., 2011. RNA virus population diversity: implications for inter-species transmission. Curr. Opin. Virol. 1, 643–648.
Casey, J.L., 2012. Control of ADAR1 editing of hepatitis delta virus RNAs. Curr. Top. Microbiol. Immunol. 353, 123–143.
Chang, F.L., Chen, P.J., Tu, S.J., Wang, C.J., Chen, D.S., 1991. The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus. Proc. Natl. Acad. Sci. U.S.A 88, 8490–8494.
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. J. Virol. 79, 8182–8188.
Chang, J., Nie, X., Chang, H.E., Han, Z., Taylor, J., 2008. Transcription of hepatitis delta virus RNA by RNA polymerase II. J. Virol. 82, 1118–1127.
Chao, Y.C., Chang, M.F., Gust, I., Lai, M.M., 1990. Sequence conservation and divergence of hepatitis delta virus RNA. Virology 178, 384–392.
Chao, Y.C., Lee, C.M., Tang, H.S., Govindarajan, S., Lai, M.M., 1991. Molecular cloning and characterization of an isolate of hepatitis delta virus from Taiwan. Hepatology 13, 345–352.
Cramer, P., 2004. Structure and function of RNA polymerase II. Adv. Protein Chem. 67, 1–42.
Deny, P., 2006. Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? Curr. Top. Microbiol. Immunol. 307, 151–171.
Domingo, E., Sheldon, J., Perales, C., 2012. Viral quasispecies evolution. Microbiol. Mol. Biol. Rev. 76, 159–216.
Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.
Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B. J., Ingles, C.J., 2002. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA 8, 1102–1111.
Filipovska, J., Konarska, M.M., 2000. Specific HDV RNA-templated transcription by pol II in vitro. RNA 6, 41–54.
Fu, T.B., Taylor, J., 1993. The RNAs of hepatitis *delta* virus are copied by RNA polymerase II in nuclear homogenates. J. Virol. 67, 6965–6972.
Gago, S., Elena, S.F., Flores, R., Sanjuan, R., 2009. Extremely high mutation rate of a hammerhead viroid. Science 323, 1308.
Gorzer, I., Guelly, C., Trajanoski, S., Puchhammer-Stockl, E., 2010. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. J. Virol Methods 169, 248–252.
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, 68–78.
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, 35–44.
Gudima, S., Dingle, K., Wu, T.T., Moraleda, G., Taylor, J., 1999. Characterization of the 5′ ends for polyadenylated RNAs synthesized during the replication of hepatitis *delta* virus. J. Virol. 73, 6533–6539.
Gudima, S., Wu, S.Y., Chiang, C.M., Moraleda, G., Taylor, J., 2000. Origin of hepatitis delta virus mRNA. J. Virol. 74, 7204–7210.
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. Nat. Struct. Mol. Biol. 15, 714–721.
Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M., 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 8, R143.
Kuo, M.Y., Chao, M., Taylor, J., 1989. Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. J. Virol. 63, 1945–1950.
Kuo, M.Y., Goldberg, J., Coates, L., Mason, W., Gerin, J., Taylor, J., 1988. Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. J. Virol. 62, 1855–1861.
Lee, C.M., Bih, F.Y., Chao, Y.C., Govindarajan, S., Lai, M.M., 1992. Evolution of hepatitis delta virus RNA during chronic infection. Virology 188, 265–273.
Leontis, N.B., Stombaugh, J., Westhof, E., 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. 30, 3497–3531.
Liao, F.T., Hsu, L.S., Ko, J.L., Lin, C.C., Sheu, G.T., 2012. Multiple genomic sequences of hepatitis delta virus are associated with cDNA promoter activity and RNA double rolling-circle replication. J. Gen. Virol. 93, 577–587.
Macnaughton, T.B., Beard, M.R., Chao, M., Gowans, E.J., Lai, M.M., 1993. Endogenous promoters can direct the transcription of hepatitis delta virus RNA from a recircularized cDNA template. Virology 196, 629–636.
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, 387–390.
Moraleda, G., Taylor, J., 2001. Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. J. Virol. 75, 10161–10169.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Rocheleau, L., Pelchat, M., 2006. The Subviral RNA Database: a toolbox for viroids, the hepatitis delta virus and satellite RNAs research. BMC Microbiol. 6, 24.
Ryu, W.S., Bayer, M., Taylor, J., 1992. Assembly of hepatitis delta virus particles. J. Virol. 66, 2310–2315.
Sureau, C., Moriarty, A.M., Thornton, G.B., Lanford, R.E., 1992. Production of infectious hepatitis delta virus in vitro and neutralization with antibodies directed against hepatitis B virus pre-S antigens. J. Virol. 66, 1241–1245.
Taylor, J.M., 2009. Replication of the hepatitis delta virus RNA genome (Chapter 3). Adv. Virus Res. 74, 103–121.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Wang, K.S., Choo, Q.L., Weiner, A.J., Ou, J.H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L., Houghton, M., 1986. Structure, sequence and expression of the hepatitis delta (delta) viral genome. Nature 323, 508–514.
Weiner, A.J., Choo, Q.L., Wang, K.S., Govindarajan, S., Redeker, A.G., Gerin, J.L., Houghton, M., 1988. A single antigenomic open reading frame of the hepatitis delta virus encodes the epitope(s) of both hepatitis delta antigen polypeptides p24 delta and p27 delta. J. Virol. 62, 594–599.
Wu, T.T., Netter, H.J., Lazinski, D.W., Taylor, J.M., 1997. Effects of nucleotide changes on the ability of hepatitis delta virus to transcribe, process, and accumulate unit-length, circular RNA. J. Virol. 71, 5408–5414.
Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M.M., Handa, H., 2001. Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science 293, 124–127.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Feb 18, 2016

This is a License Agreement between Yasnee Beeharry ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Yasnee Beeharry
License number	3812051076844
License date	Feb 18, 2016
Licensed content publisher	Elsevier
Licensed content publication	Virology
Licensed content title	Conserved features of an RNA promoter for RNA polymerase II determined from sequence heterogeneity of a hepatitis delta virus population
Licensed content author	Yasnee Beeharry,Lynda Rocheleau,Martin Pelchat
Licensed content date	February 2014
Licensed content volume number	450
Licensed content issue number	n/a
Number of pages	9
Start Page	165
End Page	173
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Role of paraspeckle proteins and RNA promoter structure in Hepatitis Delta Virus replication
Expected completion date	Mar 2016
Estimated size (number of pages)	300
Elsevier VAT number	GB 494 6272 12

Appendix VII - Curriculum Vitae

Yasnee BEEHARRY

DIPLOMAS

- **Ph.D. Biochemistry**, September 2010-present, University of Ottawa, Canada.
  (Transfer from M.Sc. to Ph.D. in February 2012)
  Graduate courses completed
  -Fall 2010: Current Topic RNA Molecular Biology
  -Fall 2011: Advanced Topics in Nutrition and Regulation of Metabolism
  -Winter 2013: Special Topics in Microbiology and Immunology
  -Winter 2014: Biology and Pathogenesis of HIV Infection

- **M.Sc. Molecular and Cell Biology**, June 2010, University of Pierre and Marie-Curie/ Pasteur Institute, Paris, France (with honours)-Specialization in Fundamental Virology.
- **Licence Sciences and Technologies**, **(B.Sc. equivalent),** June 2008, University of Pierre and Marie-Curie, Paris, France (with honours)-Specialization in Life sciences.
- **Baccalauréat (High school diploma equivalent)**, June 2005, Lycée Labourdonnais, Curepipe, Mauritius (with high honours)-Specialization in Science, concentration in Mathematics.

RESEARCH EXPERIENCE

PUBLICATIONS

1) **Yasnee Beeharry**, Lynda Rocheleau and Martin Pelchat. (2014) Conserved features of an RNA promoter for RNA polymerase II determined from sequence heterogeneity of a hepatitis delta virus population. Virology 450, 165-173

2) Dorota Sikora, Da Jiang Zhang, Teodora Bojic T, **Yasnee Beeharry**, Ali Tanara, Martin Pelchat. (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. doi: 10.1371/journal.pone.0054832

3) Teodora Bojić, **Yasnee Beeharry**, Da Jiang Zhang, and Martin Pelchat (2012) 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 93:1591-1600.

4) Lark L. Coffey, **Yasnee Beeharry**, Antonio V. Bordería, Hervé Blanc, and Marco Vignuzzi. (2011) Arbovirus high fidelity variant loses fitness in mosquitoes and mice. Proc Natl Acad Sci U S A. vol. 108 no. 38 16038-16043.

5) Stéphanie Beaucourt, Antonio V. Bordería, Lark L. Coffey, Nina F. Gnädig, Marta Sanz-Ramos, **Yasnee Beeharry**, Marco Vignuzzi (2011) Isolation of Fidelity Variants of RNA Viruses and Characterization of Virus Mutation Frequency. Journal of Visualized Experiments (doi: 10.3791/2953)

6) Book chapter. **Yasnee Beeharry** and Martin Pelchat. "Subversion of RNA Processing Pathways by the Hepatitis delta Virus" in "RNA processing". 2011.

ARTICLE Hepatitis Delta, a fascinating virus. Yasnee Beeharry. The Partner, winter 2013, issue 48.

FELLOWSHIP

-Excellence Scholarship, University of Ottawa, 2014-2015
-Ontario Graduate Scholarship Program Research Award, 2014-2015

CONFERENCES (* presenting author):

1) July 2014: Canadian Society of Microbiologists- Student Symposium Competition (international) Montreal, QC, Canada. Oral presentation.
Beeharry Y*, Rocheleau L and Pelchat M. High-throughput sequencing of a Hepatitis Delta Virus population indicates a conserved RNA structure for an RNA promoter for RNAP II

2) July 2014: International Union of Microbiological Societies Congresses. (international) Montreal, QC, Canada. Poster presentation.
Beeharry Y*, Rocheleau L and Pelchat M. - High-throughput sequencing of a Hepatitis Delta Virus population indicates a conserved structure for an RNA promoter for RNAP II

3) Sept 2013: Riboclub (international). Mount Orford, QC, Canada. Oral presentation.
Beeharry Y, Rocheleau L and Pelchat M*. Conserved features of an RNA promoter for RNA polymerase II determined from high-throughput sequencing of a population of hepatitis delta virus.

4) Sept 2013: Riboclub (international). Mount Orford, QC, Canada. Poster presentation.
Beeharry Y*, Rocheleau L and Pelchat M. Conserved features of an RNA promoter for RNA polymerase II determined from deep-sequencing of a Hepatitis Delta Virus population

5) June 2013: Canadian Society of Microbiologists (national). Ottawa, ON, Canada. Poster presentation.
Beeharry Y*, Rocheleau L and Pelchat M. Selection for the rod-like secondary structure of the right terminal domain of the hepatitis delta virus RNA genome.

6) Sept. 2012: Riboclub (international). Mount Orford, QC, Canada. Poster presentation.
Beeharry Y* and Pelchat M. Structural features conserved in the HDV promoter recognized by RNA Polymerase II.

7) June 2012: RNA Society (international). Ann Arbor, MI, USA. Poster presentation.
Beeharry Y* and Pelchat M. Characterization of HDV RNA promoter features required for recognition by RNA polymerase II.

8) Sept. 2011: Riboclub (international). September 17-21, 2011. Mount Orford, QC, Canada. Poster presentation.
Beeharry Y* and Pelchat M. Secondary structure of an HDV-derived RNA promoter required for recognition by RNA Polymerase II.

9) March 2011: ACFAS (national). Sherbrooke, QC, Canada. Oral presentation.
Beeharry Y* and Pelchat M. Caractérisation du promoteur d'ARN nécessaire à l'interaction avec l'ARN polymérase II.

MEETINGS/SEMINARS

1) January 2015: Poster Day in Computational Biology and Bioinformatics (institutional)
2) August 2014: Group for Research in Inflammation and Pathogenesis Poster day (institutional)
3) June 2012: Riboclub seminars (national). Oral presentation.
4) May 2012: RNA Journal club (institutional). Oral presentation.
5) February 2012: Biochemistry Department of University of Ottawa seminar day (institutional). Oral presentation.
6) May 2011: Biochemistry Department of University of Ottawa poster day (institutional). Poster presentation.
7) November 2011: RNA Journal club. Oral presentation.
8) March 2011: RNA Journal club. Oral presentation.
9) February 2011: Work in Progress Seminar (institutional). Oral presentation.

AWARDS

1) 2013/09 -Biochemistry travel award University of Ottawa (400$)
2) 2013/05 -Biochemistry PhD poster prize, 3rd place (50$)
3) 2012/06 -Biochemistry travel award University of Ottawa (500$)
4) 2012/06 -Graduate studies travel award (300$)
5) 2011/09 - Biochemistry travel award University of Ottawa (500 $)
6) 2001/05 -Orthography award, honours- Indian Ocean regional Championship
7) 1997/05 - Orthography award, high honours - Alliance française, Ile Maurice

WORK EXPERIENCE

- September 2010-October 2016: laboratory research work on Hepatitis Delta Virus- Department of Microbiology, Immunology and Biochemistry, University Of Ottawa (Ottawa, Canada)
- December 2009- June 2010: laboratory research work on Chikungunya Virus- Pathogenesis and Viral Populations unit, Pasteur Institute (Paris, France)
- March 2009- June 2009: laboratory research work on Rotavirus-Virology unit, Trousseau Hospital affiliated to Pierre and Marie Curie University (Paris, France)
- June 2008- August 2008: laboratory research work -Bunyaviruses and Arboviruses unit, Pasteur Institute (Paris, France)
- June 2007-August 2007: laboratory research work - Microbiology and Virology units, Candos Hospital (Candos, Mauritius)

RESEARCH REPORTS

- June 2010: Master 2 thesis and oral defense. "Generation of tools to study RNA viruses quasispecies on the Chikungunya virus model."
- June 2010: Redaction of a review and oral presentation (Pasteur Institute fundamental virology seminars class). Group work on the theme of "Neurotropic viruses as tools to study homeostasis of the central nervous system".
- June 2009: Master 1 thesis and oral defense. "Elaboration of a reverse genetic system to study Rotaviruses."

LABORATORY TECHNIQUES

- Molecular biology: cloning, mutagenesis, DNA/RNA/protein extraction, RT-PCR, q-PCR, polyacrylamide/agarose gels, transcription *in vitro*, coimmunoprecipitation, Western-Blot, immunodepletion
- Cell biology: Cell culture, transfection, RNA silencing, immunofluorescence, virus infection and isolation
- Mosquito husbandry and infection
- BSL2 and BSL3 containment facilities
- Radioactivity handling (P-32)

SUPERVISORY EXPERIENCE

- Sept 2013-Dec 2013: Teaching Assistant: Proctor/corrector for Molecular Biology class for 3rd year undergraduate students in Biochemistry/Biology, University of Ottawa
- November 2010-2015. During my PhD training, I also supervised undergraduate students, introducing the cellular and molecular biology techniques and teaching how to troubleshoot their experiments.
- September 2008-June 2009: Physics, Chemistry and Mathematics Tutor for highschool students, employed by Complétude Society

COMPUTATIONAL SKILLS

- Programming basics in Unix, PERL, R
- Basic computer skills: Microsoft office applications (Word, Excel, Power-Point, Share Point), Adobe illustrator, GIMP, Photoshop, Inkscape
- Bio-informatic tools for sequencing and RNA structure analysis: Blast, Clustal, Mfold, Vienna, Sequencher, Geneious, Bioedit, ApE

MISCELLANEOUS

Volunteer work

- April 2012-December 2014: Involvement in Partners in Research activities, a Canadian charity foundation to promote science among Canadian students.
- April 2012- December 2014: Help in translation (English to French) of the website for Sanofi Biogenius Canada competition, of the PIR newsletter "The Partner" and of the monthly website of PIR, "Virtual Researcher On Call"
- 2011: Evaluator for a competition organized by «Partners in Research» and «Let's talk Science» (scientific projects of highschool students)
- September 2008-June 2009: Involvement in AFEV, Student organization involved in the society to struggle against inequality and poverty (Paris, France). Enrolment in a big sister program.

Languages

- Creole (mother tongue)
- French (primary language)
- English (bilingual proficiency)
- Spanish (basic)