



ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



**Universitat Autònoma
de Barcelona**

Department of Biochemistry, Molecular Biology and
Biomedicine

Doctoral thesis

**Studying Hepatitis B and Hepatitis Delta
viruses quasispecies by Next-generation
sequencing: Conservation, variability, and
interaction**

Thesis author

Beatriz Pacín Ruiz

Thesis directors

Dr. Francisco Rodríguez Frías Dra. Maria Francesca Cortese

Thesis tutor

Dr. Francisco Rodríguez Frías

Barcelona, 2023

Agradecimientos

Después de dos años y poco breves pero intensos, aquí cierro esta etapa, que para mí representa un enorme aprendizaje y crecimiento a nivel científico, pero sobre todo personal. Quiero empezar dando las gracias a mis directores de la tesis Paco y a Francesca, por haberme acogido y formado en esta etapa, por haber confiado en mí y haberme ayudado a construirme a nivel laboral y personal.

Quiero agradecer la disciplina de Chari, gracias a ti soy un poco más organizada que cuando llegue, a David y a Marta por ayudarme en la resolución de mis problemas teóricos y técnicos. A la Dra. Maria Buti y a todo su equipo, siempre es un placer aprender de vosotros. A la Dra. Ariadna Rando, Unai, Gerard y Adri por contribuir con vuestro granito de arena en esta tesis. Además, quiero agradecer a mis Delta babysitters; Sara y Adriana, por toda vuestra ayuda y consejos. A mi ángel de la guarda aquí en Vall d'hebron al que cada día echo de menos; mi cucu Irene. A Gerardini, por alegrarme los días con tu energía radiante a pesar de que estes dormido de lunes a viernes. Además, me gustaría agradecer al equipo de la Dra. Gloria González del CIMA de Pamplona, por haberme acogido cálidamente durante mi estancia y por el trabajo estupendo que ha dado lugar a parte de esta tesis, y al Dr. Gracián Camps, un placer hacer equipo contigo.

Quiero aprovechar para dar la enhorabuena, además de las gracias, a todo mi T-Tesis team; Meri, Adri, Elena, Mar y Aurora. A mis secres favoritos Xose y Manolo, que cada día tenéis alguna diferente preparada y nos llenáis de vida y alegría. Y gracias a toda la gente que me habéis llenado de buena vibra durante esta etapa: las tietas, Ferrán, Merche, Nestor, Chantal... a toda la gente del VHIR; Pamela, Pablo, Ray, Tao, el team de HIV, y a todos los estudiantes que nos han acompañado.

Y no puedo estar más agradecida de haber encontrado la mejor compañera en este bonito camino en el que a veces encontramos piedras y otras pedrolas, confesaré que me lo he pasado pipa trabajando contigo. En esta aventura Fast and Furious, puedo decir que he encontrado a mi Bandolera Bryan O'conner dentro, y a mi compi de E.colis fuera, gracias AMIGA Selene.

El escritor Albert Espinosa describió la existencia de "personas amarillas", como el color del sol, que llegan a nuestra vida para iluminarnos y darnos calor. He tenido la suerte de encontrarme a lo largo de mi vida a seres de luz y que me han ayudado en mi camino. Quiero empezar agradeciendo a mi cuadrilla de Bilbao: Oihane, Amondarain, Ana, y Patri, y especialmente a María, Iris y Enara, gracias por conocerme como si me hubierais parido y cuidarme cada día desde la distancia, eskerrik asko guztiengatik. A los que hicisteis de Pamplona mi hogar durante 4 años, mi ejemplo de superación; Amaia, eres una luchadora y por ello te admiro. A Laura

y a Bollito por llenarme de alegría cada día. Y a mi inspiración y mi guía siempre que necesito encontrar mi norte: Reyes. Gracias a la conversación en segundo de del Dr. José Luis Vizmanos que me hizo reflexionar y centrarme para conseguir mis objetivos.

A las personas amarillas que hacen de Barcelona mi hogar. Nerea y Teti, habéis sido las mejores compañeras de piso que podría haber escogido. Ingriciti, te agradezco de manera especial toda tu ayuda incondicional, tienes ahora y siempre mi mano para ti. A mi esposo nº1, Joselu, por despertarme cada sábado transformado en mani manitas y por compartir conmigo este camino.

Volia agrair al millor regal que m'ha donat el bàsquet a la vida; Inés, Carlota, Carla i Laia, moltes gràcies per ser la meva familia i per ensenyar-me els veritables valors del compromís d'equip i de la amistat, gràcies per cuidar-me cada dia. I per descomptat no em podia deixar a la Blanca, a la Collado i a l'Alba, una altra meravella que m'ha regalat el bàsquet aquest any.

Pau, me has enseñado entre otras cosas a aprender a ordenar mi cabeza, pero, sobre todo, que quién la sigue la consigue. Gracias por llenarme cada día de nuevos estímulos y aprendizajes, zoriontsu egiten nauzu.

Gracias de corazón a mi familia, que me ha visto andar, crecer, caerme, levantarme, llorar y reír a carcajadas, y por el apoyo incondicional. A mi abuelo Chema, mi gran referente, y al tito Carlos, por cuidar de nosotras siempre. A mi hermana, mi madre y a mi abuela, me habéis enseñado que siempre se puede sacar fuerza para nadar a contracorriente, mis tres leonas luchadoras. A mi fader Josian, espero que estés orgulloso, te adoro. No tendré nunca palabras suficientes para agradeceros todo lo que hacéis cada día por mí y todo lo que os quiero.

¡Muchas gracias a todos de corazón!

“Necesitas caos en tu alma para dar a luz a una estrella danzante”

Friedrich Nietzsche

A las tres mujeres de mi vida y a mi padre

Index

Abbreviations.....	1
Abstract.....	7
Resumen	9
Resum.....	11
Chapter 1.Introduction.....	14
1. Hepatitis B virus infection as a global health problem.....	15
1.1. HBV: more than a small virion.	16
1.2. HBV genome organization	18
1.2.1. Precore/core ORF.....	20
1.2.2. P ORF.....	21
1.2.3. S ORF.....	23
1.2.4. X ORF.....	26
1.3. HBV life cycle	29
1.4. HBV infection by a clinical and therapeutic point of view	34
2. Hepatitis delta virus: an overlooked infection.....	39
2.1. HDV virion: a viroid-like particle	41
2.2. HDV life cycle.....	43
2.3. HDV self-cleaving property: ribozyme.....	46
2.4. The multifunctional HDAG protein	48
2.5. Current therapy against HDV.....	53
3. HDV-mediated HBV inhibition	56
4. Viral quasispecies.....	58
4.1. HBV variability.....	60
4.2. HDV variability	62
5. Studying viral quasispecies: next generation sequencing (NGS)	65
5.1. Cluster generation.....	66
5.2. Paired end sequencing	70
5.3. Sequencing by synthesis (SBS)	70
Chapter 2.Hypothesis and objectives	74
Study 1. Inspecting HDV ribozyme conservation and comparison of quasispecies variability and evolution between ribozyme and 5'-HDAG.....	75
Hypothesis.....	75
Objectives	77

Study 2. <i>In vivo</i> study of the involvement of HDV in HBV variability: the role of type I interferon pathway.....	78
Hypothesis.....	78
Objectives	79
Chapter 3. Materials	81
1. Primers and probes	82
2. Commercial kits	84
2.1. Serological Markers.....	84
2.2. DNA and RNA extraction	85
2.3. RNA and DNA quantification	87
2.4. DNA and RNA amplification and transcription	88
2.5. Nucleic acid purification.....	89
2.6. NGS library preparation	91
2.7. Mutagenesis, plasmid purification and transfection.....	92
3. Cell lines	93
Chapter 4. Study 1. Inspecting HDV ribozyme conservation and comparison of quasispecies variability and evolution between ribozyme and 5'-HDAG.....	95
1. Methods.....	96
1.1. Patients and samples.....	96
1.2. HDV RNA extraction	97
1.2.1. HDV RNA: WHO Standard Curve and Viral Load Estimation	98
1.3. Preparation of the library for the analysis by NGS of the ribozyme and 5'-HDAG 99	
1.3.1. HDV ribozyme amplification	99
1.3.2. 5'-HDAG amplification and library preparation	102
1.3.3. Quantification, normalization of the amplicons and library production	104
1.3.4. Preparation of the library.....	104
1.3.5. Preliminary amplification of pooled amplicons libraries.....	105
1.3.6. Preparation of the master pool and sequencing	106
1.4. Sequencing quality filters and bioinformatic analysis.....	106
1.4.1. Quality filters	106
1.4.2. Study of the QS conservation	109
1.4.3. Analysis of HDV quasispecies: incidence-based diversity indices	110
1.4.4. Analysis of HDV quasispecies: abundance-based indices	111
1.4.5. Analysis of HDV quasispecies: functional indices.....	112
1.4.6. Variability in HDV quasispecies and single nucleotide variations....	113
1.5. Analysis of the effects of the observed mutations on viral expression <i>in vitro</i>	114

1.5.1.	Introduction of the mutations in pCMV-HDV-1.2X plasmids	115
1.5.2.	Sanger sequencing.....	117
1.5.3.	Culture of Huh7 cells and <i>in vitro</i> testing.....	118
2.	Results.....	119
2.1.	Patients selected in the study of ribozyme conservation and variability	119
2.2.	Conservation of the quasispecies in the ribozyme region	120
2.3.	Ribozyme Quasispecies (QS) evolution and variability during patient's follow-up	123
2.4.	Analysis of Mutations.....	125
2.5.	Effects of the mutation on the viral expression <i>in vitro</i>	128
2.6.	Ribozyme quasispecies variability and evolution compared to the 5'-HDAG	130
2.6.1.	Clinical characteristics of the patients included in the study	130
2.6.2.	Viral quasispecies complexity between ribozyme and 5'-HDAG.....	131
2.6.3.	Quasispecies evolution rate between ribozyme and 5'-HDAG	132
Chapter 5. Study 2. <i>In vivo</i> study of the involvement of HDV in HBV variability: the role of type I interferon pathway.....		
1.	Methods.....	136
1.1.	Superinfection mouse model	136
1.2.	Analysis of HBV expression in serum	138
1.3.	Analysis of the intrahepatic HBV expression: pgRNA.....	139
1.4.	Next-generation sequencing analysis of HBV and HDV quasispecies variability in the liver of superinfected IFN α / β -R WT or KO mice.....	142
1.4.1.	Amplification of the regions of interest in HBV genome.....	143
1.4.2.	Amplicon purification	146
1.5.	Sequencing quality filters and bioinformatic analysis.....	147
1.5.1.	Rare Haplotype Load study.....	147
1.5.2.	Analysis of single nucleotide variations	147
2.	Results.....	149
2.1.	HBV expression in superinfected mice	149
2.1.1.	Serological evaluation of HBV replication	149
2.1.2.	Intrahepatic HBV expression	151
2.2.	Analysis of the intrahepatic HBV RNA quasispecies in SRT at 7 and 21 dpi	153
2.3.	Analysis of the intrahepatic HBV X gene quasispecies at 7 and 21 dpi....	161
2.4.	INF-I dependent HDV variability.....	170

Chapter 6. Discussion	177
Study 1. Inspecting HDV ribozyme conservation and comparison of quasispecies variability and evolution between ribozyme and 5'-HDAG.....	178
Study 2. <i>In vivo</i> study of the involvement of HDV in HBV variability: the role of type I interferon pathway.....	184
Chapter 7. Limitations of the studies.....	195
Chapter 8. Conclusions.....	199
Chapter 9. Futures perspectives.....	204
Chapter 10. Bibliography	207
Chapter 11. Manuscript	242
Chapter 12. Appendix. Publications and Congresses	257

Abbreviations

A

Aa	Amino acid
AAT	α -1 antitrypsin
AAV	Adeno-associated vector
ADAR	Adenosine deaminase acting on RNA
AGL	Antigenic Loop
AID	Activation-induced cytidine deaminase
ALT	Alanine aminotransferase
APOBEC	Apolipoprotein B mRNA editing enzyme
ARM	Arginine-rich motifs
AST	Aspartate aminotransferase

B

BCP	Basal core promoter
BLV	Bulevirtide

C

Cas9	CRISPR-associated protein 9
cccDNA	Covalently closed circular DNA
CCD	Coiled-coil domain
CHB	Chronic Hepatitis B
CHD	Chronic Hepatitis D
CpAms	C protein assembly modulators
CRISPR	Clustered regularly interspaced short palindromic repeats

D

DMEM	Dulbecco's Modified Eagle Medium
DR	Direct repeat
dsDNA	Double-stranded DNA
dslDNA	Double-stranded linear DNA

E

EAlb	Albumin enhancer
EASL	European Association for the Study of the Liver
EDTA	Ethylenediaminetetraacetic acid
enh	Enhancer
ER	Endoplasmic reticulum

F

FBS	Fetal bovine serum
FRET	Fluorescent resonance energy transfer

G

GalNAc	N-acetyl galactosamine
--------	------------------------

H

HBc	Hepatitis B core protein
HBcAg	Hepatitis B core Antigen
HBcrAg	Hepatitis B core-related Antigen
HBeAg	Hepatitis B e Antigen
HBV	Hepatitis B Virus
HBVtg	Hepatitis B Virus transgenic
HBx	Hepatitis B X protein
HCC	Hepatocellular carcinoma
HCV	Hepatitis C Virus
HDAg	Hepatitis Delta Antigen
HDV	Hepatitis Delta Virus
HGS	Gini-Simpson index
hIFN β	Human IFN β
hIP10	Human Interferon gamma induced protein 10kDa
HIV	Human immunodeficiency virus
HLH	Helix-loop-helix
Hsp70	Heat shock protein 70
HSPG	Heparan Sulfate Proteoglycan

I

IC	Information Content
IFB	Intermediate filament protein
IFIT1	Interferon-induced protein with tetratricopeptide repeats 1
IFN	Interferon
IRF	Interferon regulatory factors
ISG	IFN stimulated gene
ISO	Isoprenylation
ITR	Inverted terminal repeats

K

kDa	Kilodalton
KO	Knock-out

L

L-HBsAg	Large Hepatitis B surface Antigen
L-HDAg	Long Hepatitis Delta Antigen
LNF	Lonafarnib
LTR β	Toll-like receptor β

M

M-HBsAg	Medium Hepatitis B surface Antigen
MAVS	Mitochondrial antiviral-signaling protein
MDA5	Melanoma differentiation associated gene 5
MF	Mutation Frequency
MI	Monoinfected
MPTC	Master percentaje
mRNA	messenger RNA
MSTR	Master reads
MVB	Multivesicular bodies
MxA	Myxovirus resistance protein 1

N

NAPs	Nucleic acid polymers
NEF	Negative Elongation Factor
NES	Nuclear export signal
NF-kB	Nuclear factor Kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
NLS	Nuclear import signal
NPC	Nuclear pore complexes
nt	Nucleotide
NUC	Nucleotide analogues

O

OBI	Occult HBV infection
ORF	Open reading frames

P

PCR	Polymerase chain reaction
PegIFN- α	Pegylated IFN α
pgRNA	Pre-genomic RNA
PKC	Protein kinase C
PML	Promyelocytic leukemia protein
PRE	Post-transcriptional regulatory element
PRR	Pattern Recognition Receptors

Q

QS	Quasispecies
----	--------------

R

rAAV	Recombinant deno-associated vector
RCA	Rolling Circle Amplification
rcDNA	relaxed circular DNA
RHL	Rare Haplotype Load
RISC	RNA-induced silencing complex
RNP	Ribonucleoprotein
RT	Reverse transcriptase

S

S-HBsAg	Small Hepatitis B surface Antigen
S-HDAg	Short Hepatitis Delta Antigen
sgRNA	Sub-genomic RNA
SI	Superinfected
SMCHD1	Structural maintenance of chromosomes flexible hinge domain containing 1
Sn	Shannon entropy
SNV	Single nucleotide variations
STAT1	Signal transducer and activator of transcription 1

T

th	T helper
TM	Transmembrane
TP	Terminal protein

V

VAS	Viral assembly sequence
-----	-------------------------

U

UPGMA	Unweighted pair group method with arithmetic mean
-------	---

W

WHO	World Health Organization
WT	Wild-type

Abstract

The Hepatitis B virus (HBV) affects around 296 million people worldwide. Of them, more than 5% are coinfecting with Hepatitis delta virus (HDV), causing the most severe viral hepatitis. HDV is a viroid-like virus since it presents a catalytically active RNA portion known as ribozyme, and a region encoding for the unique viral protein, called delta antigen (HDAG). HDV is a satellite virus of HBV and strongly inhibits the expression of its helper. The mechanism behind this inhibition isn't fully known yet, but the activation of type I interferon (IFN) and its stimulated genes such as mutagenic enzymes (APOBEC/AID or ADAR1) might play a role.

Of note, both viruses are characterized by a high degree of genetic variability and circulate as a dynamic mix of closely related variants called quasispecies (QS). The next-generation sequencing (NGS) allows for a deep analysis of the complete viral QS, supporting the study of the QS evolution and the detection of newly generated and infrequent haplotypes. This thesis aims to analyze HDV and HBV QS variability through NGS and it is composed of two studies focused on:

- Studying QS conservation and evolution from chronic hepatitis delta (CHD) patients in the ribozyme and comparing it with the 5'HDAG region.
- Inspecting the impact of HDV in HBV variability and the role of type I IFN in an HDV-superinfected HBV-transgenic mouse model with the IFN α / β receptor knock-out (KO).

In the first study, two longitudinal serum samples were collected from 25 CHD patients. The QS was analyzed by NGS (*Miseq, Illumina*) to determine ribozyme QS conservation and its evolution and compare this last with the 5'-HDAG. The most relevant mutations observed in the ribozyme were also tested *in vitro*. The ribozyme QS was overall conserved, with a hyperconserved region (nt 715-745). No differences in QS were observed over time. Thirteen mutations were detected, three with higher frequency (T23C, T69C and C64 deletion). Of them, C64 deletion reduced 1log the HDV replication *in vitro*. Compared to the 5'HDAG, the latter one was more complex and variable. In terms of evolution, the viral QS evolved differently depending on the region analyzed.

In the second study, wildtype (WT) or KO for the IFN α / β receptor HBVtg mice were injected with adeno-associated vectors (AAV) containing HDV (superinfection) or luciferase (monoinfection). HBV expression was analyzed by quantifying the circulating HBV DNA and RNA and the intrahepatic RNA (pgRNA). The QS of the intrahepatic viral RNA was analyzed by NGS at 7- and 21-days post-infection (dpi) in the *X* and *SRT* regions for HBV and the HDV 5'HDAG. In this model, HDV

determined an inhibition in circulating HBV DNA and RNA, and limitedly impact pgRNA. The SRT QS was more complex at 7dpi than at 21dpi, different from the *HBX*. Multiple SNVs were observed at both times, especially C→T transitions (typical of APOBEC/AID deaminases) in all the experimental conditions. Among the non-common variations, some of them mainly affected superinfected mice and seemed not associated with the activity of specific mutagenic enzymes. HDV QS was more complex in WT than KO mice, especially at 21dpi, with a higher frequency of G→A transitions typical of ADAR1.

In the study focused on HDV, we identified a hyperconserved region of the ribozyme as a possible target for gene therapy, and we observed that 5'-HDAG QS evolved more rapidly than the ribozyme. The second study showed, we saw that the mutagenic enzymes could contribute to the HBV and HDV variability. Of note, some mutations, especially in superinfected mice could occur in a type I IFN-independent way, thus suggesting that other pathways or even the cellular RNA polymerase might be involved.

Resumen

El virus de la hepatitis B (VHB) afecta a unos 296 millones de personas, más del 5% coinfectadas con el virus de la hepatitis delta (VHD), causando la hepatitis vírica más grave. El VHD presenta un dominio de RNA auto-catalítico, la ribozima, y una región que codifica para la única proteína viral, denominada antígeno delta (HDAG). El VHD es un virus satélite del VHB e inhibe fuertemente la expresión de su “helper”. No se conoce completamente el mecanismo de esta inhibición. La activación del interferón tipo I (IFN-I) y la activación de enzimas mutagénicas (APOBEC/AID o ADAR1) podrían ser relevantes.

Ambos virus presentan alta variabilidad genética y circulan como una mezcla dinámica de variantes estrechamente relacionadas, denominadas quasiespecies (QS). La secuenciación masiva (NGS) permite un análisis complejo de la QS, además del estudio de su evolución y detección de haplotipos minoritarios. Esta tesis busca analizar la variabilidad de la QS del VHD y del VHB mediante NGS. Compuesta por dos estudios principales:

- Estudio de la conservación y evolución de la QS de pacientes con hepatitis crónica delta (CHD) en la ribozima y su comparación con la región 5'HDAG.
- Inspeccionar el impacto del VHD en la variabilidad del VHB y el papel del IFN-I en un modelo de ratón VHB-transgénico superinfectado, con el receptor del IFN α/β noqueado (IFN α/β R-KO).

En el primer estudio, se recogieron dos muestras longitudinales de suero de 25 pacientes con CHD y se analizó la QS mediante NGS (Miseq, Illumina) para determinar la conservación de la ribozima, su evolución y comparar esta última con el 5'-HDAG. Se analizaron *in vitro* las mutaciones más relevantes observadas en la ribozima. La QS de la ribozima estaba globalmente conservada, con una región hiperconservada (nt715-745). No se observaron diferencias en el QS con el tiempo. Se detectaron trece mutaciones, tres con mayor frecuencia (T23C, T69C y delección C64). La delección C64 redujo 1log la replicación del VHD *in vitro*. En comparación con el 5'HDAG, éste era más complejo y variable. En términos de evolución, la QS viral evolucionó de forma diferente según la región analizada.

En el segundo estudio, los ratones HBVtg wild-type (WT) o IFN α/β R KO fueron inyectados con vectores adenoasociados (AAV) con VHD (superinfección) o luciferasa (monoinfección). La expresión del VHB se analizó cuantificando el ADN y el ARN circulantes y el ARN intrahepático (ARNpg). La QS del ARN viral intrahepático se analizó mediante NGS a los 7 y 21 días post-infección (dpi) en las regiones *X* y *SRT* del VHB y el 5'HDAG del VHD. El VHD determinó una inhibición en

el ADN y el ARN circulantes del VHB, pero limitada en el ARNpg. La QS del SRT fue más compleja a los 7dpi que a los 21dpi, a diferencia del HBX. Se identificaron múltiples SNVs en ambos tiempos, especialmente C→T transiciones (típicas de APOBEC/AID) en todas las condiciones experimentales. Entre las variaciones no comunes, algunas afectaban principalmente a los ratones superinfectados y no parecían estar asociadas con la actividad de enzimas mutagénicas. La QS del VHD fue más compleja en los ratones WT que en los KO, especialmente a los 21dpi, con mayor frecuencia de transiciones G→A típicas de ADAR1.

En el estudio centrado en el VHD identificamos una región híperconservada de la ribozima como posible diana de terapia génica, y observamos que la QS del 5'-HDAG evolucionaba más rápido que la ribozima. En el segundo estudio, vimos que las enzimas mutagénicas podían contribuir a la variabilidad del VHB y el VHD. Destacamos que algunas mutaciones, especialmente en ratones superinfectados, podrían producirse independientemente del IFN-I, sugiriendo que podrían estar implicadas otras vías o incluso la ARN polimerasa celular.

Resum

El virus de l'hepatitis B (VHB) afecta uns 296 milions de persones arreu del món, més del 5% coinfectades amb el virus de l'hepatitis delta (VHD), causant l'hepatitis vírica més greu. El VHD presenta un domini d'ARN auto-catalític, la ribozima, i una regió que codifica per l'única proteïna viral, l'antigen delta (HDAG). El VHD és un virus satèl·lit del VHB i inhibeix la seva expressió. El mecanisme d'aquesta inhibició encara no es coneix totalment. L'activació de l'interferó tipus I (IFN-I) i l'activació d'enzims mutagènics (APOBEC/AID o ADAR1) podrien desenvolupar un paper important.

Ambdós virus presenten una elevada variabilitat genètica i circulen com una barreja dinàmica de variants estretament relacionades entre sí, denominades quasiespecies (QS). La seqüenciació massiva (NGS) permet un anàlisi complex de la QS així com l'estudi de la seva evolució i detecció d'haplotips minoritaris. Aquesta tesis té com a objectiu analitzar la variabilitat de la QS del VHD i del VHB mitjançant NGS. Està composta de dos estudis principals:

- Estudi de la conservació i evolució de la QS de pacients amb hepatitis crònica delta (CHD) a la ribozima i la comparació amb la regió 5'HDAG.
- Inspeccionar l'impacte del VHD en la variabilitat del VHB i el paper del IFN-I en un model de ratolí VHB-transgènic superinfectat, amb el receptor del IFN α / β inactivat (IFN α / β R-KO).

En el primer estudi es van recollir mostres longitudinals de sèrum de 25 pacients amb CHD i es va analitzar la QS mitjançant NGS (Miseq, Illumina) per a determinar la conservació de la ribozima i la seva evolució i comparar-la amb el 5'-HDAG. Es van analitzar *in vitro* les mutacions més rellevants observades en la ribozima. La QS de la ribozima estava globalment conservada, amb una regió hiperconservada (nt715-745). No es van observar diferències en la QS al llarg del temps. Es van detectar tretze mutacions, tres amb major freqüència: T23C, T69C i la deleció C64. Aquesta última va reduir 1log la replicació del VHD *in vitro*. En comparació amb el d'HDAG, aquest últim era el més complex i variable. La QS viral va evolucionar de forma diferent dependent de la regió analitzada.

En el segon estudi, els ratolins HBVtg wildtype (WT) o IFN α / β R-KO van ser injectats amb vectors adenoassociats (AAV) amb VHD (superinfecció) o luciferasa (monoinfecció). L'expressió del VHB es va analitzar quantificant l'ADN i l'ARN circulants així com l'ARN intrahepàtic (ARNpg). La QS del ARNpg es va analitzar mitjançant NGS als 7-21 dies post-infecció (dpi) en les regions X i SRT del VHB i el 5'HDAG del VHD. El VHD va determinar una inhibició de l'ADN i l'ARN circulants del VHB, però limitada en el ARNpg. La QS del SRT va ser més complexa als 7dpi que als

21 dpi, a diferència del HBX. Es van identificar múltiples SNVs en ambdós temps, especialment transicions C→T (típiques de APOBEC/AID) en totes les condicions experimentals. Entre les variacions no comunes, algunes afectaven principalment als ratolins superinfectats i no semblaven estar associats amb l'activitat d'enzims mutagènics. La QS del VHD va ser més complexa en els ratolins WT que els KO, especialment als 21dpi, amb major freqüència de transicions G→A típiques de ADAR1.

En l'estudi centrat en el VHD vam identificar una regió hiperconservada de la ribozima com a possible diana de teràpia gènica, i vam observar que la QS del 5'-HDAG evolucionava més ràpid que la ribozima. En el segon estudi, vam veure que els enzims mutagènics podien contribuir a la variabilitat del VHB i el VHD. Cal destacar que algunes mutacions, especialment en ratolins superinfectats, podrien produir-se de forma independent del IFN-I, suggerint que podrien estar implicades altres vies o inclús l'ARN polimerasa circular.

Chapter 1

Introduction

1. Hepatitis B virus infection as a global health problem

In 1965, Blumberg and colleagues identified an antigen in the blood of an Australian aboriginal that could react with the blood of an American haemophilic patient [1]. Four years later this protein was linked to hepatitis and to the Hepatitis B virus (HBV) surface antigen (HBsAg) [2,3]. In 1970, Dane and colleagues identified HBV viral particles (so called Dane Particles) by electron microscopy [4]. The complete viral genome was first sequenced in 1979 by Galibert and partners [5]. Although a lot had been learned since then in both clinics and basic virology, there is still a lot to explore.

Hepatitis B virus belongs to *Hepadnaviridae* family, genus *Orthohepadnavirus*. This virus has an exclusive tropism for the liver. Similar viruses have been identified in other mammals and birds. All of them have a very similar genomic arrangement and share an essential step of the viral replication: the reverse transcription of the viral RNA to produce a double-stranded DNA molecule [6].

HBV is found in body fluids such as blood, saliva, urine, semen, or vaginal fluid of infected people. The transmission route can change depending on the geographical area. In most high-income countries the infection occurs at adulthood, and it is spread mainly through sexual contact, parenterally (sharing of needles) or other direct contacts with contaminated body fluids [7]. Differently, in highly endemic area, the infection mainly occurs at early age, especially from infected mother to the child at birth (perinatal transmission) or through the exposition to infected blood from infected child to another [8,9].

From 1992, the HBV vaccination has been included in most of the national vaccination programs. Nonetheless, the World Health Organization (WHO) estimates that around 296 million people worldwide are living with chronic hepatitis B, and that in 2019 it caused 820,000 deaths, with 1.5 million new infections per year (WHO- last update in June 2022) [9]. Of note, the infection is heterogeneously worldwide distributed, with highly endemic areas such as

Southeast Asia, China, Taiwan, Alaska, and sub-Saharan regions, regions with intermediate endemicity such as Mediterranean countries, Eastern Europe, Central Asia, Middle East, and South America and low endemic regions such as United States of America, Western Europe, Australia, and Japan [10,11] (Figure 1). Considering this high prevalence, the WHO considers the HBV infection a global health problem.

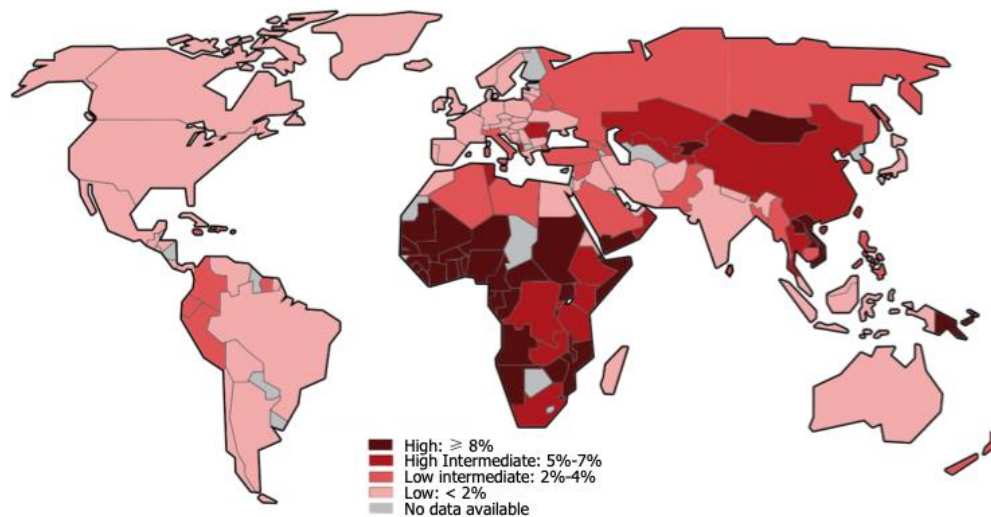


Figure 1. Hepatitis B virus global distribution. In this world map it is indicated the estimated distribution of chronic HBV infected patients. The prevalence (%) of HBV infected is represented in red scale, from light red for the less affected areas to dark red, to the more affected areas. Map obtained from Jefferies et al. 2018 and data extracted from Schweitzer et al. 2015 [12,13].

This concept is even more reinforced if we consider that the total number of HBV infected people is probably underestimated, especially in vulnerable populations such as marginalized groups, immigrants, prisoners, or high-risk groups such as intravenous drug users (IDUs) [14].

1.1. HBV: more than a small virion.

Although HBV is a small virus, it causes a complex infection characterized by the release of different viral particles, a intricate life cycle, and the expression of several viral proteins that play key roles in both viral replication and cellular activity, thus strongly contributing to the pathogenesis of the liver disease.

HBV is a small DNA virus of 40-42 nm in diameter. The Dane particle (Figure 2) is composed by an icosahedral nucleocapsid consisting of 180-240 core proteins (HBc) that holds the viral genome and polymerase and some host proteins (specific isoforms of protein kinase C (PKC), cellular protein nucleophosmin (B23), Hsp70 and others) [15–17]. The nucleocapsid is covered by a lipidic envelope consisting of three different-in-length surface proteins that share the C-terminal end and differ in the length of the N-terminal end: the Large (L-HBsAg), the medium (M-HBsAg) and Small (S-HBsAg) [18]. Of note, this is not the unique viral particles identified in serum of infected patients.

During viral replication, the empty subviral particles (SVPs), composed by just the viral surface antigens, are released with an excess between 1,000 to 100,000-fold related to the complete virions. They acquire a spherical (25 nm) or filamentous (22nm) form with variable length [19] and differ in viral proteins composition, being the first constitute of S-HBsAg, M-HBsAg and hardly L-HBsAg, and the second especially abundant of the large protein [20,21]. The proportion of L-HBsAg also affects the SVPs secretion pathway. The filamentous SVP share the same secretion of the complete virions (the host cell endosomal sorting complexes required for transport -ESCRT- components and the multivesicular bodies - MVBs), whereas the spheres appear to be secreted through the constitutive host cell secretion pathway [22,23].

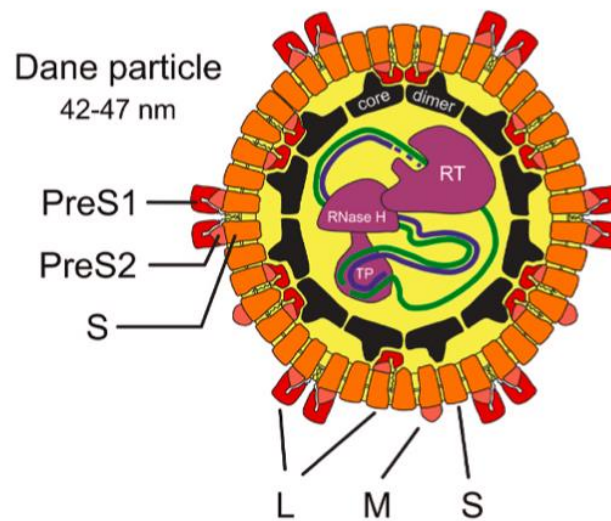


Figure 2. Structure of HBV Dane particle. HBV is provided of an envelope and present the three viral surface proteins (L, M and S) that differ for the composition of the three domains (S, PreS1 and preS2). The envelope protects a capsid composed by dimers of the core protein (black), that inside contains the polymerase protein with its three domains (in violet, the Retrotranscriptase (RT), RNase H and the terminal protein (TP)) and the HBV relaxed circular DNA with its complete negative strand (green) and incomplete positive strand (purple) [24].

In addition of the SVPs, other viral particles have been described, including those provided of all the viral proteins but containing viral RNA instead of DNA. Although the mechanism behind their production is still unclear and they are released at lower concentrations related to the classical Dane particles [25], the RNA-containing viral particles are considered a direct surrogate of the expression of the viral intracellular minichromosome [26].

1.2. HBV genome organization

The HBV genome is formed by a partially relaxed, circular, non-covalently closed double-stranded DNA molecule (rcDNA) of approximately 3.2 kb (Figure 3). The two strands are asymmetric, as the negative strand (3'→ 5') or minus strand is 3200 bp long and contains the full viral DNA sequence, and the positive strand or plus strand (5'→3') includes just two thirds of the genome size. It was proposed that the disparity in the length of the two strands of the rcDNA might be due to a limitation of the dNTPs within the nucleocapsid during the retrotranscription step [27]. Of

note, although HBV is a DNA virus, the genome is produced through a retro-transcription step by using RNA as template called pre-genomic-RNA (pgRNA). This process will be deeply explained later in this thesis (section 1.3 of this chapter). The end of both strands is flanked by two 11 nt-long regions, called directed repetitive DR1 (at 5'-end of the minus strand) and DR2 (at 3'-end of the plus strand), both essential for the viral replication and the maintenance of the rcDNA circularity [28,29]. Moreover, the viral genome also comprises two enhancer sequences (Enh1 and Enh2) that contribute to promote viral transcription [30].

Another feature that characterizes HBV DNA is its high degree of overlapping. Four open reading frames (ORF) have been described, each one overlapping to a different extent with other ORF. This feature allows the virus to hold a relevant amount of genetic information into a small virion. The four ORFs encode for the different structural and functional proteins involved in viral replication, as detailed below.

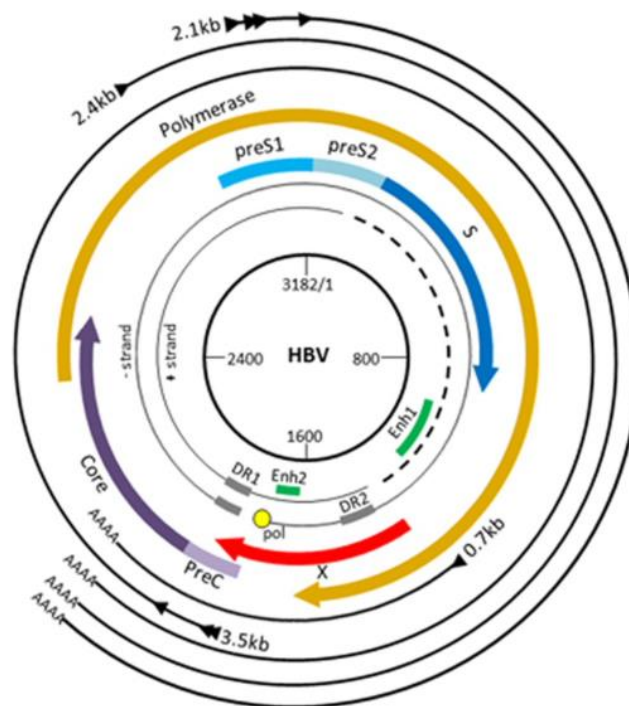


Figure 3. HBV genome organization. HBV genome is formed by a partially double stranded non-covalently closed DNA (rcDNA) with a complete negative strand (- strand, in grey with continuous line) and an incomplete positive strand (+ strand, dashed line). The four ORF are represented in colored arrows. The two enhancers are reported in green (Enh1 and Enh 2), whereas the two direct repeats (DR1 and DR2) in grey. Once inside the cells, the rcDNA is repaired and forms a covalently closed DNA (cccDNA) that serve as template for viral transcripts. The different mRNAs and their length are indicated in black line [31].

1.2.1. Precore/core ORF

The precore/core ORF has two in-frame starting codons, one at position 1814 and the second at 1901 in HBV genome and translates for two proteins. If translation begins at the position 1901, the core protein (HBcAg of 21 kDa) is produced [32], whereas, if translation starts from position 1814, a soluble non-structural protein of 25 kDa (212-214 aa in-length), known as HBV e antigen (HBeAg) is produced [33,34]. Notably, the Enhancer II and the Basal core promoter (BCP, nt 1742- 1849) reside within the core promoter. The BCP contains major nuclear binding sites, important for the accurate initiation of the pgRNA than the precore mRNA [35].

The HBcAg is composed by 183 or 185 aa depending on the genotype, and its function is not just limited to being the structural unit of the HBV inner capsid. It is also essential for the formation and maintenance of the HBV cccDNA [36], and for the encapsidation of the pgRNA together with the viral polymerase. It can be divided in two main domains, the N-terminal (149 or 151 aa -long depending on the genotype) coordinates Hbc self-assembly into dimers, and the C-terminal end, differently, encompasses the last 34 aa and it contains multiple Arginine-rich motifs for the encapsidation of the pgRNA and seven 7 highly conserved serine and threonine that can dynamically be phosphorylated or the phosphorylated thus being crucial for HBV life cycle [37–39].

Considering the key role of HBcAg in viral replication, new antiviral molecules targeting HBV encapsidation have been proposed, such as the C protein assembly modulators (CpAMs), that interact with Hbc dimers increasing the rate of formation of deficient capsids lacking pgRNA [40].

The HBeAg, instead, has immunomodulatory activity. It has been associated with reduced antigen presentation, the downregulation of the Kupffer cells and macrophages co-stimulatory molecules and the inhibition of T helper (Th) lymphocytes maturation [41,42]. This immunomodulatory effect might also affect the immune response and chronification of HBV infection in those neonates infected through vertical transmission, since this protein can cross the placenta and establish

Th-specific tolerance [43]. At clinical level, HBeAg is a marker of viral replication, infectivity, inflammation, disease severity and response to antiviral therapy [42,44]. As HBcAg is an important stimulator of the immune system, it presents several epitope regions for B and T lymphocytes, mutations in this region could hinder the recognition of the virus by cytotoxic T cells [45,46]. Mutations have been described in the BCP and CP, which could affect the production of HBeAg, necessary for the synthesis of anti-HBe, which are usually associated with immune control of the infection, for example the double mutation A1762T and G1764A at BCP reduces up to 70% the production of HBeAg [47,48].

1.2.2. P ORF

The P ORF encodes for the viral 90 kDa polymerase and covers the 70% of the viral genome. The viral polymerase is a multifunctional protein provided of three enzymatic properties: retrotranscriptase (RT), DNA-dependent polymerase and a RNase-H activity [34].

By a structural point of view, the polymerase consists of four main domains: the terminal protein (TP, encoded by the nt 2307-2862), the spacer (SP, from nt 2863-129), the reverse transcriptase (RT, nt 130-1161) and the RNase H (RH, 1162-1623) [49]. The TP domain is essential for HBV replication since it allows binding to the pgRNA thanks to a conserved tyrosine residue, leading the synthesis of the negative strand (-DNA) during retrotranscription [50]. The spacer (SP) is about 150 amino acids-long and seems to serve as a junction between the different domains [51]. The RT domain acts as reverse transcriptase and DNA-dependent polymerase to synthesize both negative and positive chains of the rcDNA. The RNase H acts by degrading pgRNA during rcDNA synthesis and assists during the encapsidation of the complex formed by pgRNA, viral polymerase and chaperones [49].

The RT polymerase is divided into fingers, palm and thumb domains as predicted for HIV (Figure 4) [52]. The center of this structure presents the catalytic domain of the enzyme by which DNA strand extension occurs; this region can be divided into 7 catalytic core domains (A-G). Inside the C domain (aa rt200-210) resides the

hyper-conserved tyrosine-methionine-aspartate-aspartate (YMDD) motif, which is involved in the polymerization activity [53]. Mutations in this domain and selected during antiretroviral therapy could compromise the replicative efficiency of the viral polymerase [54], and have also been reported to cause antiviral response failure [55].

When looking the secondary structure of the protein, the HBV RT adopts the right-hand shape of a DNA polymerase with fingers, palm, and thumb subdomains (Figure 4). The finger and thumb subdomains are primarily composed of α -helices, whereas the palm regions mainly comprise α -helices and β -sheets. The catalytic core domains are distributed mainly in the palm and secondarily in the finger subdomains (Figure 4).

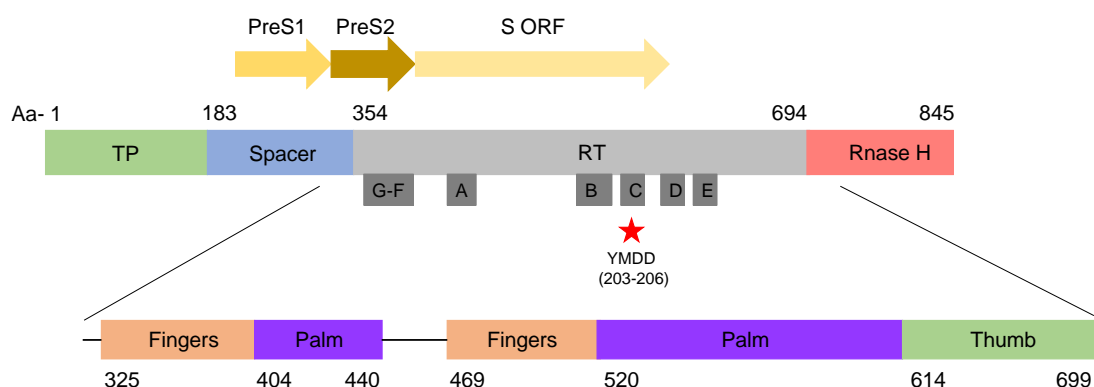


Figure 4. Schematic representation of HBV polymerase. This figure represents polymerase gene domains detailing polymerase retrotranscriptase domain (RT) from aa 325 to 699. There are indicated the three types of structure of RT and the amino acids on which they fall; fingers (orange, aa 325-403 and 469-519), palms (purple, aa 404-440 and 520-613) and the thumb (green, aa 614-699). Also, there are represented the PreS1, PreS2 and S ORF overlapping regions with arrows in different yellow tones. Catalytic domains from A to F are represented in dark grey. YMDD domain is highlighted with a red star. Figure modified from Das k, et al. 2001 and Clark D, et al. 2015 [52,56].

Of note, the polymerase is the specific target of the most common used antiviral strategy, which is based on nucleos(t)ide-analogues (NUCs). These drugs act as chain terminators because, due to their similarity to natural nucleos(t)ide, they are inserted into the nascent strand, prematurely blocking further elongation of the viral DNA genome [57]. The most administered NUCs are the nucleotide analogues Tenofovir (TDF) and Tenofovir Alafenamide (TAF), and the nucleoside analogue

Entecavir (ETV) [58] . Other available analogues are Lamivudine, Telbivudine, Adefovir, and Besifovir Dipivoxil (only administered in Korea) [59].

In view of the above, several mutations can be selected in this genomic region due to treatment and cause resistance to NUC-based treatment [52,60]. Moreover, considering the functional role of the polymerase in the viral life cycle, mutations in this gene could strongly impact the viral replication. As example, the mutation rtM204V/I/S has been associated with LMV resistance. Notably, *in vitro* studies shows that rtM204I/V mutants replicate less efficiently than wild-type HBV [54,61]. Is also important to consider that around 70% of the polymerase gene overlaps with the ORF of the surface proteins, which it means that mutations acquired during antiretroviral therapy could also affect the HBV envelope proteins. This field will be deeply discussed in the next sub-paragraph.

1.2.3. S ORF

The S ORF includes the preS1, preS2 and S genes and encodes for the three different-length surface antigens: the large HBsAg or L-HBsAg (protein of 39kDa, produced from the preS1, preS2 and S mRNA), the medium HBsAg or M-HBsAg (protein of 33 kDa, obtained from the preS2 and S) and the small HBsAg or S-HBsAg (protein of 24 kDa, encoded by the S gene) (Figure 5). These proteins are also the major component of the SVPs that are normally produced during HBV infection (section 1.3 of this chapter). The S domain contains a large hydrophilic portion (from the amino acid 99 to 169) called major hydrophilic region (MHR). It comprises 5 MHR sub-regions (MHR1-5) according to their conformation and 8 highly conserved cysteines that are observed in all HBV subtypes and that are involved in HBsAg antigenicity and viral secretion [62]. The 'a' determinant (between amino acids 124 and 147) is also an essential part of the MHR and is the major antigenic region of the three surface antigens [63].

HBV envelope proteins are integral proteins that anchor to the membrane thanks to the TM regions of the S domain (Figure 6). The N-terminal domain is where the protein inserts into the membrane and forms the TM1 (aa 8-22). The second

transmembrane domain (TM2, aa 80-92), directs the translocation of the protein into the ER lumen. The C-terminal region (aa 170-226) has not yet been fully characterized but it includes the two transmembrane domains (TM3/4) (Figure 5) [17,64,65].

In addition to their length, the three surface proteins differ in the post-translational modifications they undergo (Figure 5). All the proteins share a glycosylation in the asparagine 146 (N146) in a loop (aa 99-169 of the S domain) oriented towards the endoplasmic reticulum (ER) lumen. The M-HBsAg has a further N-glycosylation at the N-terminal end of the preS2 domain (N4), which is also exposed in the ER lumen. Related to the other proteins, the L-HBsAg has a dual orientation of the preS1, that can be exposed in the ER lumen or in the cytosol. This different topology is especially due to the TM1, whose function is to translocate the protein to the lumen. In the cytosolic orientation of the preS1, the TM1 is not used, so the preS2 domain is displayed in the cytosol and the N4 cannot be glycosylated, thus allowing the myristylation of the glycine in position 2 of the preS1 [66,67]. Studies suggest that in the 50% of the L-HBsAg a posttranslational translocation occurs in which TM1 is integrated into the membrane and the region of preS1 and preS2 is exposed to the ER lumen in a kind of maintained balance between types of large isoform [68–71]. Notably, these two forms of LHBS could have different functions: in mature virions the preS1-preS2 domain is exposed to the outside of the viral particle, and it is essential for the interaction with the cell, whereas the preS1-preS2 domain that is exposed to the cytoplasmic face has an essential role interacting with the nucleocapsid in viral morphogenesis [72,73].

Of note, the myristylation of the PreS1 is a condition required for the viral entry. A 47-nucleotide preS1 mimic peptide, called Bulevirtide (BLV) (before known as Myrcludex B) has been proposed as therapeutic strategy against HBV infection (currently in clinical trial), acting as competitive antagonist for the cellular receptor involved in viral entry. A combination of NUCs with Myrcludex could potentially limit the *de novo* infection and prevent hepatocytes re-infection [74].

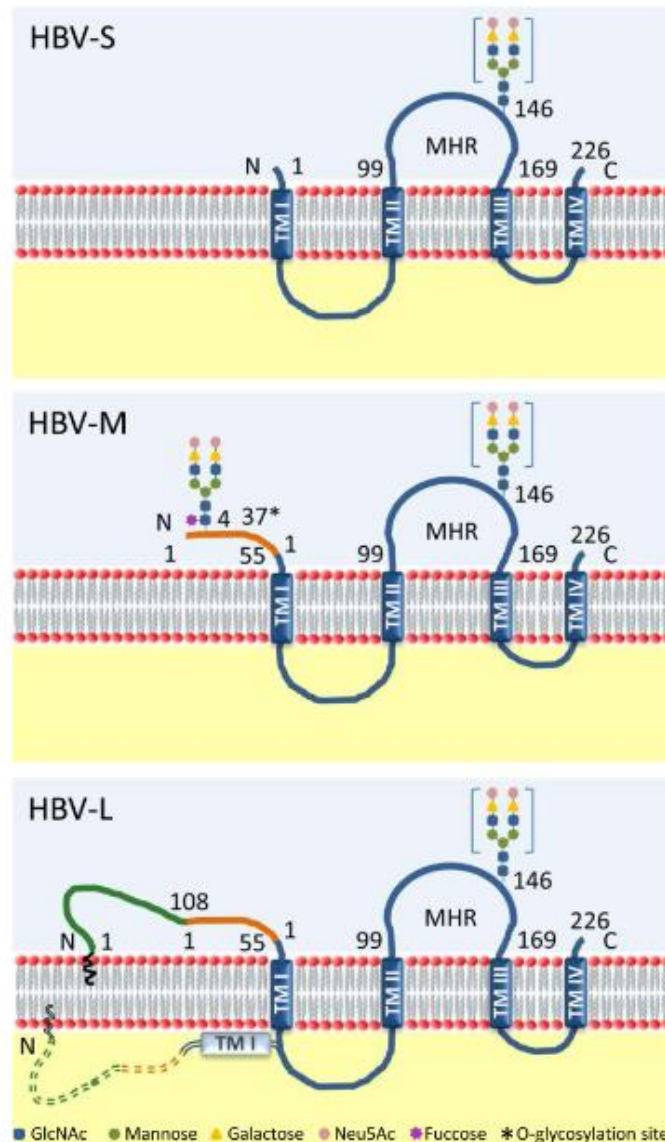


Figure 5. Schematic representation of the three HBV surface proteins. The S domain, which is shared by all of them has four transmembrane domains (TM1-TM4). This configuration can change in HBV large protein (HBV-L), where the TM1 can be transmembrane and the preS1-preS2 domain exposed to the ER lumen or can be intra-cytosolic. The extremes of each sub-domain are reported together with the site of the N-glycosylation (N146 in all the proteins and N4 in HBV-M). The M isoform can also be involved in O-glycosylation (site highlighted through an asterisk [75]).

As mentioned above, mutations in this ORF, which are associated with viral clearance and vaccine escape, could affect viral replication, viral particle release and even liver disease progression. To cite an instance, the mutation G145A in the “a” determinant is one of the main vaccine immune escape mutations, whereas

deletions in preS1 and preS2 were usually reported associated to hepatocarcinoma (HCC) and its recurrence [76,77].

Another point to note about the S ORF is that its sequence completely overlaps with the P ORF, which means that mutations selected by the immune response in one case, or by antiviral treatment in the other, could introduce changes in the overlapping region. For example, mutations in the polymerase gene associated with Lamivurtide resistance such as rtV173L, rtL180M and rtM204I, seems to introduce amino acid changes in the surface gene (sI195M and sW196S) [78]. In some cases, the amino acid change selected in one ORF, might generate a change in the overlapping region that compensates the first variation, thus maintaining the viral fitness. This is the case of the substitutions sP120T and sG145R in the “a” determinant, that introduce in the finger sub-domains of the RT the rtT128N and rtW153Q mutations, that can stabilize the enzyme functionality of those strains resistant to the treatment [79,80]. Similarly, changes in the P ORF could introduce substitutions in the S ORF that improve the viral fitness, such as the rtS878T selected after ETV/TDF-based treatment that inserts a premature stop codon in the HBsAg (C69*) that increases the viral transcription, generates immune escape, and promotes the release of exosomes-mediated virions [81].

1.2.4. X ORF

The X ORF is the smallest HBV open reading frame and encodes for the homonymous 17 kDa protein HBx. Its main function is to transactivate the viral expression by interacting with different host proteins involved in the regulation of cellular transcription [82]. Thanks to its property, HBx can not only promote viral expression, but also interfere with various cellular processes, making it the most important viral protein involved in carcinogenesis [83].

The x protein is 154 amino acids-long and it is formed by two main functional domains (Figure 6) [84]. The transactivation domain resides in the C-terminal domain (from the amino acid 50 to 154) while the N-terminal domain (aa 1-50) is

the negative regulatory domain fundamental to repress HBx-mediated transactivation [85].

The transactivation domain is essential for HBx activity, as previously reported by Tang et al, who demonstrated the loss of transactivation activity when this region was truncated [86]. Between the amino acids 88 and 100 of HBx there is a region that is conserved throughout the *Hepadnaviruses*, able to bind the DDB1 (Damage specific DNA binding protein 1) with CUL4-ROC1 (CRL4) ubiquitin ligase E3 that promotes the degradation of the Smc5/6 (Structural maintenance of chromosomes 5 and 6) complex. This complex acts by interfering with the acetylation of the histones H3 and H4, thus inhibiting the transcription of extrachromosomal elements such as cccDNA [82,87]. The interaction of HBx with the DDB1 hijacks the DDB1 to target the Smc5/6 and promotes the degradation of this restriction factor, thus enhancing HBV replication [88].

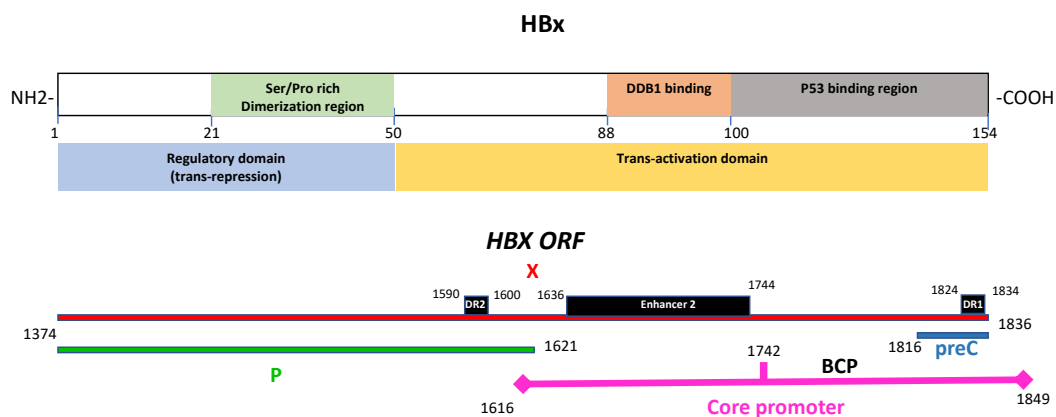


Figure 6. HBV X protein and ORF. The HBx protein covers a key role in HBV replication. It can be divided in two main domains: the regulatory domain (aa 1-50) and the transactivating domain (aa 51-154). The regulatory domain encompasses the Ser/Pro-rich dimerization region, whereas in the transacting domain reside both the DDB1 binding domain (aa 88-100) and the p53 binding domain (101-154). By a nucleotide point of view (in red), the HBX ORF overlaps with a portion of the polymerase (in green nt 1374-1621), with the core promoter and BCP (in pink, nt 1616-1849, with the BCP in position 1742-1849), and the precore region (in blue, nt 1816-1836). In the X ORF also reside some transcriptional regulator domains, such as the DR2 (nt 1590-1600), Enhancer 2 (1626-1744) and DR1 (1824-1834).

The negative regulatory domain also has a serine- and proline-rich region (aa 21-50) which is essential for the HBx protein dimerization [89]. In addition, the HBx is subjected to several posttranslational modifications such as acetylation,

phosphorylation, and disulfide bond formation, whose biological significance is still unclear [31,90]. Additionally, nine cysteine residues have been found highly conserved throughout the protein [91]. Notably, by a nucleotide point of view, the HBX ORF overlaps with other regions, such as a portion of the P ORF, the core promoter (including the Enh 2 and the BCP) and the precore region (Figure 6).

As mentioned above, the HBx protein, thanks to its ability to interact with different host proteins, can strongly modulate different cellular pathways (Figure 7), such as cell cycle progression, acute immune response, protein degradation pathways, genetic stability, apoptosis, and the regulation of genes involved in tissue organization [89].

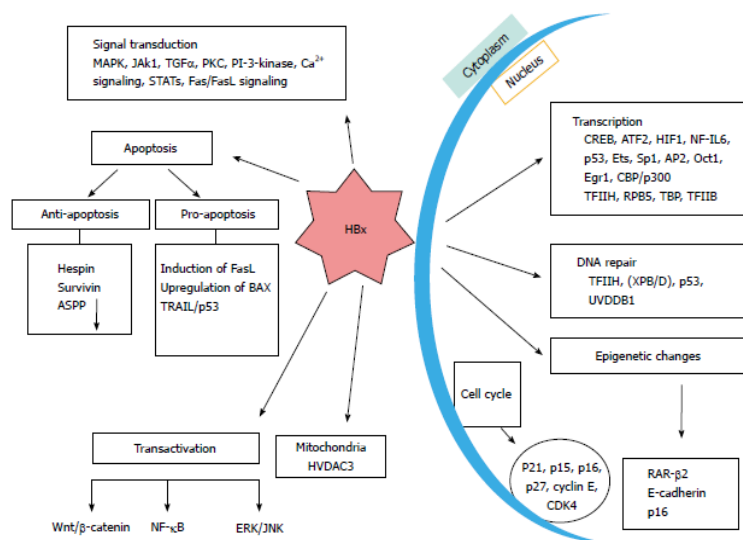


Figure 7. HBx modulation of cellular signaling pathways. The diagram shows the different cellular pathways that can be modulated by HBx and the interacting nuclear and cytoplasmic host factors [92].

Considering the tight interaction between HBx and cellular activity and homeostasis, and its key role in viral replication, amino acid changes in the viral protein may strongly contribute to the progression of liver disease [93]. For example, the mutations in HBX G1386A/C (corresponding to the V5M/L mutation in the protein), C1653T (H94Y), T1753V (I127V), and HBx C-terminal deletions or insertions have been associated with liver disease severity [93]. Substitutions in specific positions of the C-terminus of the HBx such as the 120, 121, 130 and 131

can promote the ability of the viral protein to interfere with cellular processes like the DNA repair process or cell cycle progression thus contributing to the development of HCC [92,94,95]. Mutations in the HBx can also impact viral replication. A deletion of 8bp (nt 396-404) has been associated with reduced secretion of HBV virions and HBsAg and may be involved in occult hepatitis [93]. Another example is the genotype-specific pattern of mutations (A12S/P33S/P46S/T36D), that were described by our group, that was mainly observed in low-replicative clinical stages and were also associated with the reduction of the expression of HBV *in vitro* [96].

1.3. HBV life cycle

Hepatitis B virus enters the hepatocytes after crossing the space of Disse. A double interaction is established between the arginine and lysine residues of the antigenic loop (AGL) of the pre-S1 of the L-HBsAg with the Heparan Sulfate Proteoglycans (HSPGs) of the hepatocyte membrane [97]. This initial attachment facilitates the interaction between the preS1 (between aa 2 and 48) and the Na⁺-taurocholate co-transporting polypeptide (NTCP) in the basolateral membrane of the hepatocyte [98,99]. This second interaction is specific, and the NTCP transport is considered the cellular receptor of HBV. Two different ways of viral entry have been proposed: by endocytosis and subsequent release of the nucleocapsids from the vesicles or by fusion of the cell plasma membrane and the viral envelope [100] (Figure 8).

After this step, the nucleocapsids with the rcDNA are released into the cytoplasm. These particles are targeted to the nuclear pore complexes (NPCs) and cross the nuclear membrane through the microtubules by active transport. This is possible by the recognition of an NLS (nuclear import signal) domain present in the HBc protein that allows the cross through the nuclear basket, a cage-like structure on the nuclear side of the NPC [101].

The rcDNA is released into the nucleoplasm by a still unknown mechanism and the HBc capsids are uncoupled. As previously commented in section 1.2 of this chapter, the positive strand of rcDNA is one-third shorter than the complete strand, so it must

undergo a repair process led by the cellular DNA repair complex. Different cellular factors and enzymes are involved in the repair and ligation of this strand [102]. At the end of this process, a double-stranded, covalently closed viral DNA is obtained, which, thanks to viral and cellular proteins, gives rise to a minichromosome known as cccDNA [36]. As it happens for the cellular chromosomes, the expression of the cccDNA is regulated by epigenetic modifications such as DNA methylation and histone acetylation that involve multiple host factors such as the general control nonderepressible 5 (GCN5) and Yin Yang 1 (YY1), the activator of transcription (STAT) 1 and 2, histone deacetylase (HDAC) 1 and 11, sirtuin (SIRT) 1 and 3, protein arginine methyltransferase (PRMT) 1 and 5, enhancer of zeste homolog 2 (EZH2), etc. [36,103].

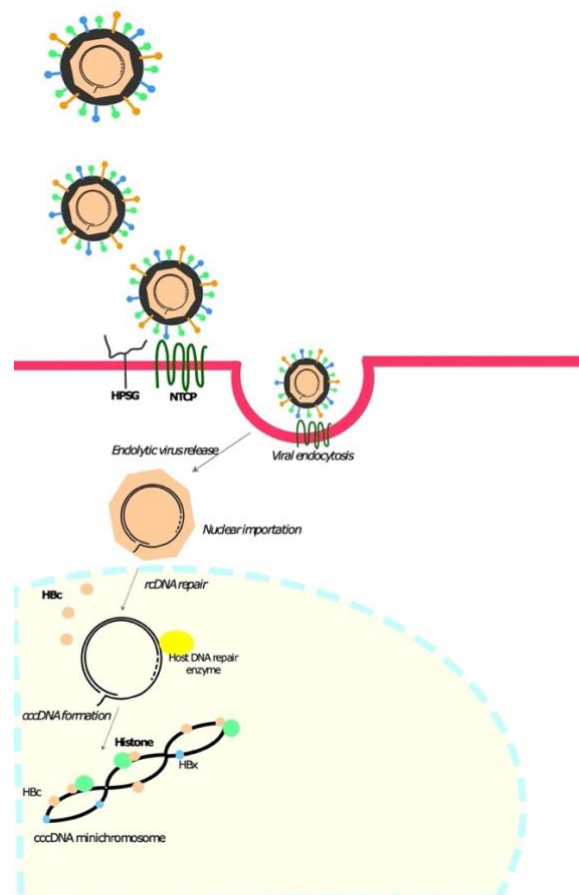


Figure 8. HBV entry into the hepatocyte. HBV entry into the hepatocytes through the interaction, first with the HPSG, and then the NTCP receptor. It is not still known if the virus enters by membrane fusion or endocytosis. In this figure, HBV is transported and released inside the cells through endocytic vesicles. Once inside, the rcDNA and core proteins are delivered into the nucleus where it is repaired by host DNA repair enzymes, circularized, and then coupled with viral (HBC, or core protein) and to cellular proteins (histones) thus giving rise to the circular covalently closed viral DNA (cccDNA). The viral x protein (HBx), although it is not a structural element, could transactivate the cccDNA transcriptional activity. HSPGs: Heparan Sulfate Proteoglycans; NTCP: Na⁺-taurocholate co-transporting polypeptide; rcDNA: relaxed circular DNA; HBC: core antigen or HBcAg. Figure modified from Garcia-Garcia S *et al.* 2021, Expert Review of Molecular Diagnostics [55].

The transcription of the cccDNA is carried out by the host polymerase II, giving rise to the various messenger RNAs required for viral replication and the production of viral proteins. HBx promotes this step by acting on cccDNA epigenetic modifications, for example recruiting histone deacetylase [87,104], or interacting with long non-coding RNA (like the DLEU2) [105], or with the YY1 factor [106]. Another role of the HBx is also to promote the degradation of the Smc5/6 restriction factor through its interaction with the DDB1 ligase (as previously cited in section 1.2.4 of this chapter, entitled “X ORF”). Related to cccDNA transcription, a direct interaction between cccDNA and host transcription factors have been also reported, such as the CCAAT/enhancer-binding protein (C/EBP), which binds a site in the Enh2 region [107]. Several other transcription factors can bind specific sites in the cccDNA promoter and enhancer, such as hepatocyte nuclear factors (HNF1, 3 and 4), activator protein 1 (AP-1), TATA binding protein (TBP), cAMP response element binding protein (CREB) and others [108].

Notably, all the HBV transcripts contain the viral post-transcriptional regulatory element (PRE) which includes three conserved RNA structures: the alpha (α), beta (β), and epsilon (ϵ) stem loops. The ϵ loop is located at the 5'-end of the pgRNA (nt 1847- 1907) and is essential for the viral replication and encapsulation signaling. The other two loops are in the 3'-end, in position 1292-1321 and 1411-1433 for respectively the α and β loops. Both participate in nuclear export, belonging to one of the main nuclear export domains (SEP2) [109].

By a functional point of view, the mRNAs produced can be schematically distinguished based on their downstream expression in respectively sub-genomic (sgRNA) or pre-genomic RNA (pgRNA) (Figure 9).

The pgRNA serves as a template for the synthesis of new rcDNA molecules by retro-transcription and thanks to the epsilon loop (or encapsidation signal), forms a ribonucleoprotein complex consisting of the viral RNA, the core proteins, and the viral polymerase. This complex will be packaged into a newly formed viral capsid [110,111]. At this step, a partial dephosphorylation of the phosphorylable residues

present in the C-terminal domain of the HBc proteins is required for an optimal encapsidation [112].

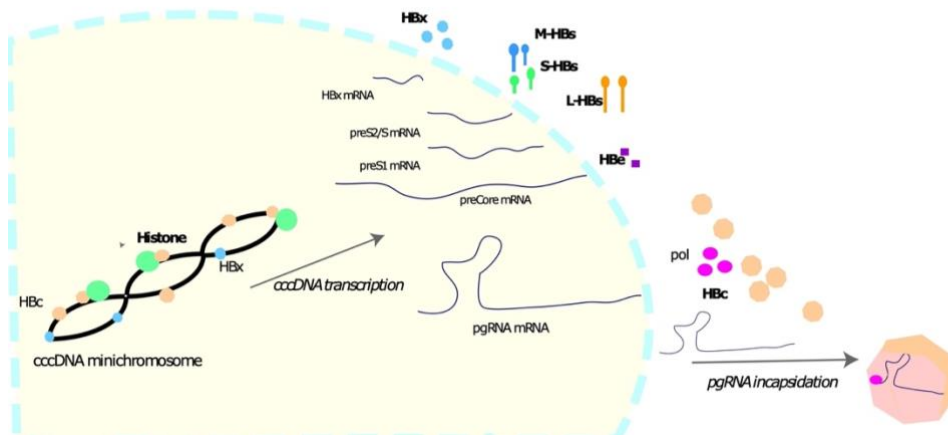


Figure 9. HBV cccDNA transcription. Transcription of the cccDNA gives rise to pregenomic RNA (pgRNA) and different sub-genomic RNAs: HBx, preS1, PreS2/S and preCore mRNAs. These mRNAs will give rise to the translation of the viral proteins: HBx, surface proteins (M-HBs, S-HBs, L-HBs), HBe, HBc and the viral polymerase. The pgRNA, differently, will be encapsidated, together with the viral polymerase, by the new formed capsid to be retrotranscribed. Figure modified from Garcia-Garcia S *et al.* 2021, Expert Review of Molecular Diagnostics [55]

Once the pgRNA and polymerase are wrapped into the capsid, the retrotranscription step begins. First, the polymerase PT region generates a 4 nt primer (TGAA) complementary to the 5' epsilon loop protrusion of the pgRNA and translocates it to the DR1 region located at the 3' end of the pgRNA [111,113]. This will promote the retrotranscription of the pgRNA, thus obtaining the negative (minus strand). The RNaseH activity of the same polymerase degrades the pgRNA, except for a sequence of 11-16 nt at the 5' end of the pgRNA [114]. This fragment, containing the DR1 sequence, is translocated to the DR2 present in the new produced DNA strand (-) and acts as priming for the synthesis of the positive DNA strand (plus strand) carried out by the DNA polymerase activity of the viral polymerase. The synthesis of this strand is incomplete due to limited availability of dNTPs within the capsid. The double stranded DNA can, so, acquire the classical circular conformation thanks to the terminal redundancies of each strand that juxtaposes the DR1 and DR2 [49,115], or a linear conformation (linear double-stranded DNA - dsDNA) caused by the failure of the translocation of the 11-16 nt RNA primer to the DR2 thus giving rise to viral particles carrying linear dsDNA. This dsDNA could be integrated into the host

genome thus forming the pool of integrated HBV DNA [115,116]. Of note, the integrated DNA is incomplete and can mainly support the expression of HBsAg [117].

In addition of its role as rcDNA template, the pgRNA allows the translation of the viral core protein (HBc or HBcAg) and the polymerase (P). The other subgenomic mRNAs encode for the different viral proteins. Of note, due to the circular organization of the rcDNA and its overlapping, all the mRNAs share the same terminal sequence (corresponding with the *HBX* sequence).

The envelope proteins are co-translationally inserted into the endoplasmic reticulum (ER) membrane. Part of these proteins are translocated through the ER lumen and are secreted as 22 nm subviral envelope particles (SVPs) or 42 nm infectious virions or Dane particles (both particles are deeply explained in section 1.1 of this chapter). In this last case, the mature capsid pass through the multivesicular bodies (MVB) acquiring the envelope. This process of vesicle formation is mediated by the ESCRT (endosomal sorting complex required for transport) machinery, and the virions are finally released into the bloodstream [100,118] (Figure 10).

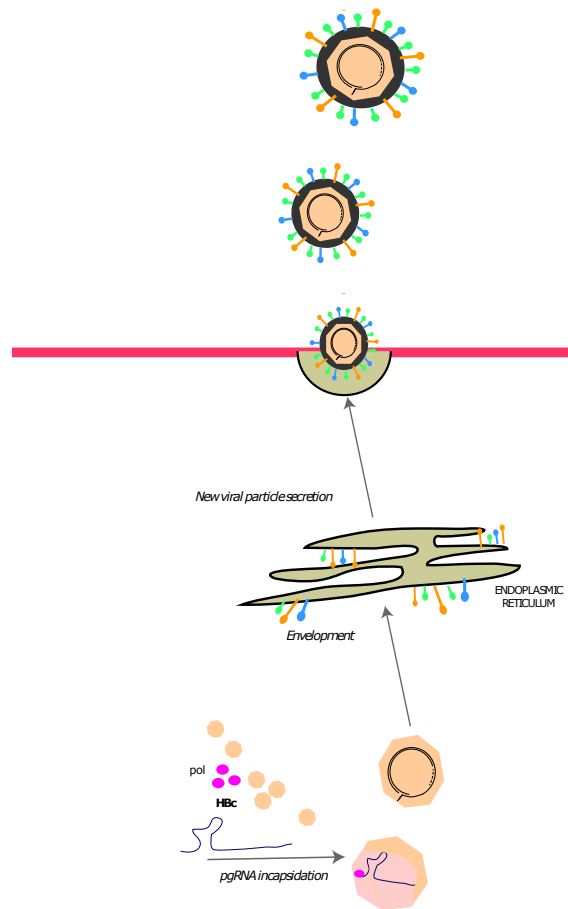


Figure 10. Formation and secretion of the HBV viral particle. The polymerase retrotranscribes the pgRNA thus obtaining the rcDNA (first the complete negative and the incomplete positive DNA strand). Once retrotranscribed, the mature capsids go through the endoplasmic reticulum (ER) where they acquire the plasma membrane modified with HBsAg surface proteins. The complete viral particles are now released into the bloodstream, being available to infect the neighboring cells. Figure modified from Garcia-Garcia S et al. 2021, Expert Review of Molecular Diagnostics [55].

1.4. HBV infection by a clinical and therapeutic point of view

The incubation period between the infection and the onset of symptomatology ranges from 30 to 180 days during an acute infection (as reported by las WHO report, in June 2022). Most of the time, the infection is self-limited, although the HBV cccDNA can be detected in the liver of acute infected patients, determining the so-called occult infection (OBI). This condition is defined as the persistence of replicative HBV DNA in the liver (cccDNA minichromosome) or HBV DNA in the blood in HBsAg-negative people [119,120].

Notably, 10% of the people infected in the adulthood, and 90% of the children infected at birth, develop a chronic infection, defined as the presence of HBsAg in serum for more than six months [7]. According to the European Association for the Study of the Liver (EASL), four clinical stages can be distinguished depending on both virological (HBsAg, HBeAg/anti-HBe, HBV DNA), biochemical (ALT) and disease markers (fibrosis markers) [57] (Figure 11).

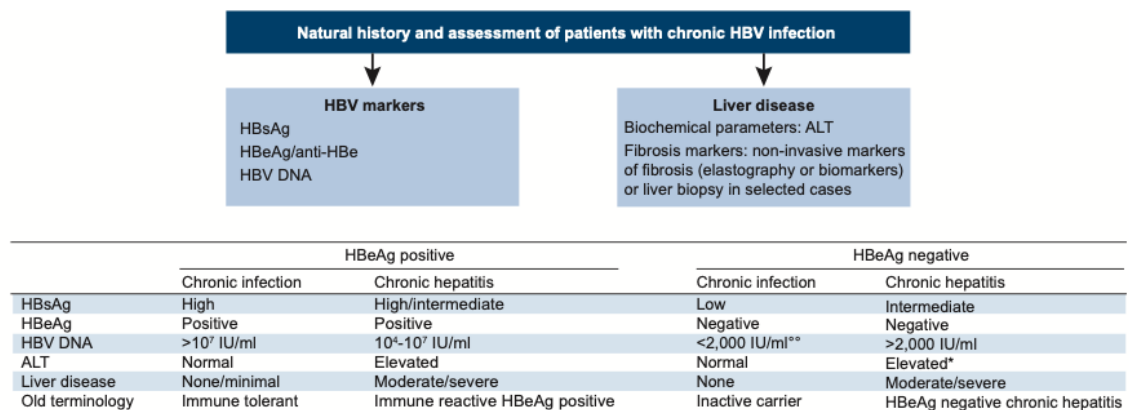


Figure 11. EASL-defined clinical stages for chronic HBV infection. The figure shows the virological (HBsAg, HBeAg and HBV DNA), biochemical (ALT) and clinical (liver disease) markers to define the four clinical stages of chronic HBV infection and the cut-off values suggested by the European Association for the Study of the Liver. ALT: Alanine Aminotransferase [57].

Of note, the different clinical stages can influence the evolution of the liver disease, and chronic hepatitis B patients (CHB) usually show higher risk of developing cirrhosis and HCC (Figure 12).

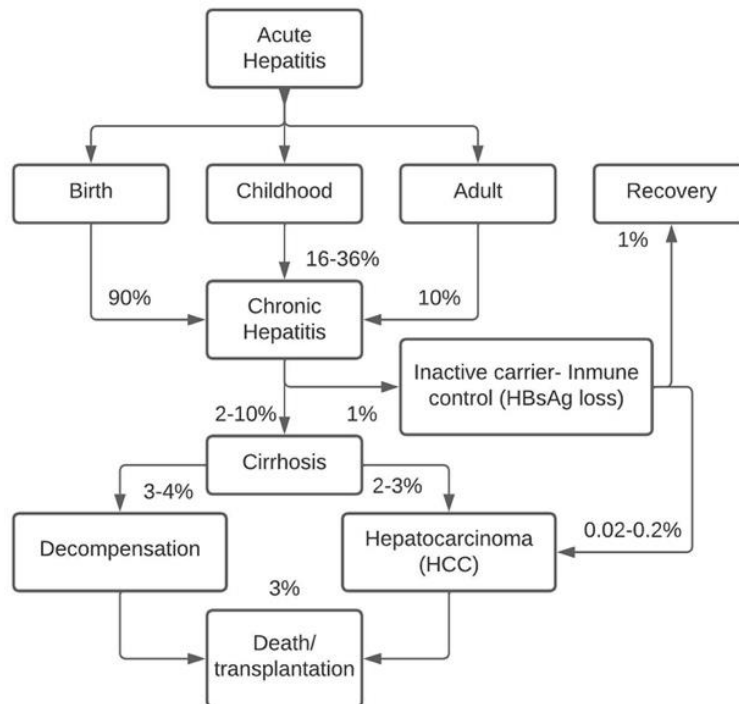


Figure 12. HBV infection and liver disease progression. The diagram shows the liver disease evolution and the risk of progression and the clinical consequences depending on the different clinical stages. Data obtained from Pollicino et al. 2021 and Flattovich et al. 2008. [120,121]

According to the EASL guidelines, patients with chronic hepatitis, HBV DNA > 2000 IU/mL, elevated transaminases (ALT higher than the upper limit) and presenting at least moderate necroinflammation should be treated, independently of the HBeAg state [57]. The current antiviral strategy is based on NUCs (as cited in section 1.2.2 of this chapter, entitled “P ORF”). This strategy enables the optimal control of HBV replication (with undetectable HBV DNA), and 1-5% of the treated patients experiences a clearance of the HBsAg in the serum/plasma with or without the development of anti-HBs, defined functional cure [122]. However, just the 30% percentage of these patients presented a durable suppression of the viremia, especially after treatment interruption [123]. This happens because the antiretroviral therapy does not act on the HBV cccDNA which persists into the nucleus of the infected cells, being source of the expression of the viral mRNAs and contributing to viral relapses and/or disease progression. It has been reported that in Asian patients, the cumulative risk of developing HCC at 5 and 10 years (even in absence of detectable HBV DNA) has been estimated to be between 1.6 and 5.9%

respectively [124]. In Caucasians, this risk is lower but still consistent (<1% in non-cirrhotic and 1.8-5.2% in cirrhotic patients) [125]. To try to guarantee a better control of HBV infection, new therapeutic approaches are currently in development: the entry inhibitor Buluvertide (or Myrcludex), the capsid inhibitors, inhibitors of the release of the HBsAg, gene silencing (silencing RNA or antisense oligonucleotides) and gene editing agents [126] (Figure 13). On the other hand, the new strategies are also focused on the acting on the immune response immune system activators, monoclonal antibodies, checkpoint inhibitors and T cell-targeted therapies [126].

A gene therapy strategy based on gene silencing could be extremely helpful to control this infection, since it could inhibit viral replication and interfere with the liver disease by directly acting on HBV. Small interfering RNAs (siRNAs) are valuable tools to inhibit gene expression since they are recognized and processed by the DICER and RNA-induced silencing complex (RISC) promoting the degradation of the mRNA target [127,128]. These siRNA are usually conjugated with N-acetyl galactosamine (GalNAc) to guarantee a liver-specificity, since the GalNAc receptor is highly expressed in the liver and promotes rapid endocytosis [129]. Antisense nucleotides (ASOs), differently, are single-stranded oligonucleotide and show a different mechanism of action related to the siRNA, since they promote the degradation of the RNA target through the RNase H [130].

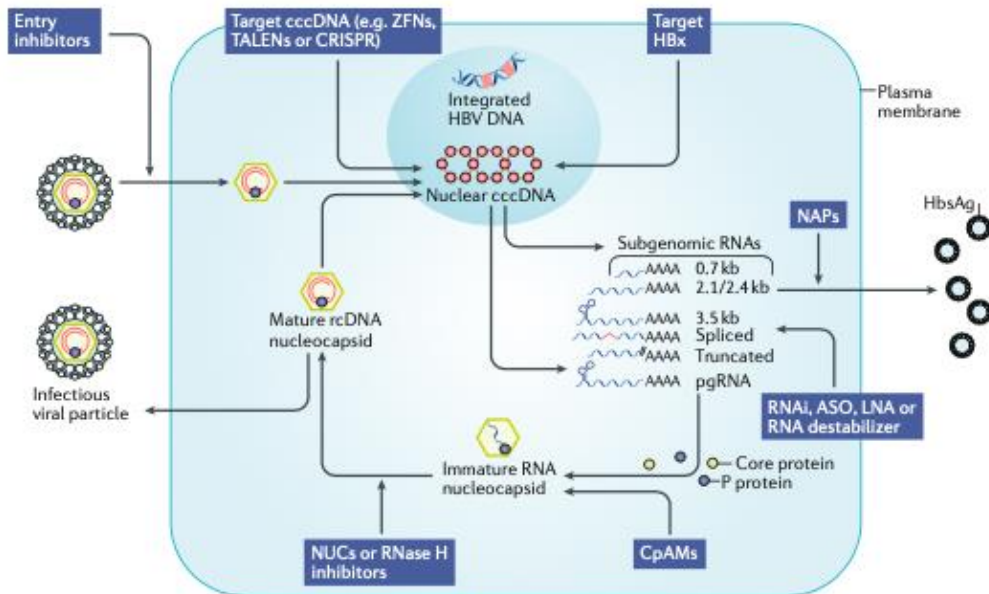


Figure 13. Novel therapies against HBV. This illustration shows the direct treatment lines for HBV inhibition. It shows the entry inhibitors, Nucleotide analogues (NUCs) or RNaseH inhibitors, CpAMs, surface protein inhibitors (NAPs) and new gene therapy strategies, such as RNA interference (RNAi), Antisense oligonucleotides (ASO), LNA or RNA destabilizers at cytoplasmic level. In addition, new strategies at nuclear level such as targets against HBx or cccDNA based on ZFNs, TALENs or CRISPR tools are shown [131].

Improving patient follow-up to define the optimal time to stop antiretroviral therapy, especially in patients with undetectable HBV DNA, is also of paramount importance. To this end, two surrogate markers of cccDNA have been studied: circulating HBV RNA and hepatitis B core-related antigen (HBcrAg) [132]. Circulating HBV RNA is a normal product of the viral replication, it could be an optimal surrogate for cccDNA activity as it is produced exclusively from the cccDNA minichromosome [26]. HBV RNA levels vary according to the clinical stage of the patient and could be a direct predictor of HBeAg seroconversion [26,133,134]. It can be detected even after years of successful treatment and its undetectability could be associated with low HBsAg (<100 IU/mL) [135].

Another marker is the core-related antigen (HBcrAg), which simultaneously measures three viral proteins sharing a 149 aa-long sequence: the HBcAg, the HBeAg and a 22 kDa truncated pre-core protein (p22) [136,137]. This marker has important clinical applications, studies have linked increased HBcrAg levels to the development of HCC or HBV reactivation after NUCs therapy suppression [138,139]. In addition, in patients with occult HBV, it may also predict the risk of HBV reactivation [140].

2. Hepatitis delta virus: an overlooked infection

The hepatitis delta virus (HDV) is a satellite virus of HBV since it needs its helper virus to produce new infectious viral particles. The unique specific viral protein, known as delta antigen (HDAg) was discovered by Dr. Rizzetto in chronic HBV-positive patients with a very severe form of hepatitis [141], and the virus was identified as a new hepatitis virus (called HBV-dependent hepatitis Delta or HDV) in 1980 [142]. The genome was inspected six years later. This HBV satellite virus was classified as the only member of the *Deltaviridae* family and the Deltavirus genus [143].

It has been reported that around 5% of people with chronic HBV infection have HDV co-infection, accounting for 15 to 20 million affected patients worldwide [144], with a different prevalence depending on the studied continent (Figure 14). HDV infection is first diagnosed by detecting anti-HDAg antibodies. However, this test is usually conducted in HBV patients with a more severe liver disease. In fact, a recent study performed at the Vall d'Hebron Hospital (Barcelona, Spain) showed that the incidence of HDV could be about 5 times higher than expected if the anti-HDAg is analyzed in all HBV-positive new diagnoses [145]. This is even more important considering that the HBV infection is still poorly studied especially in low-income countries [144]. A recent metanalysis reported that Mongolia had the highest national prevalence of HDV infection among the HBsAg carriers (36.9%) and a prevalence >10% was also estimated in Republic of Moldova, Western and Middle Africa [146]. In Pakistan there is a region known as HDV “belt” where 30-50% of the HBsAg-positive patients are HDV coinfecting [147]. Of note, a lot of data is still not available specially in especially in some regions of Africa, South America, and Asia [148]. Since it is still an overlooked infection, it is difficult to evaluate the real global prevalence of HDV infection.

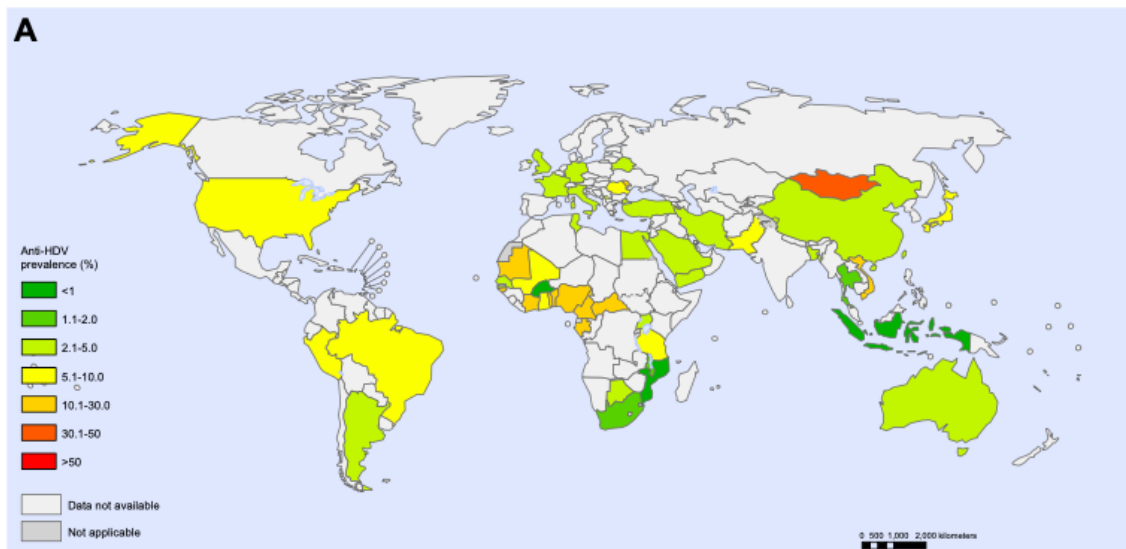


Figure 14. Percentage (%) of the anti-HDV global prevalence. In this world map it is indicated the prevalence of HDV (positive for the anti-HDAg) among the HBsAg-positive patients. The color scale indicates the percentage of anti-HDV positive population (from less prevalent, in green, to higher prevalent in red). The white and grey colors indicate those regions where data is not available or difficult to obtain. Figure extracted from Stockdale et al. 2020 [146].

Notably, the presence of HDV in the context of HBV infection has serious clinical implications (Figure 15). HDV can infect simultaneously with HBV (which is defined as coinfection) or in a context of CHB (superinfection) [149]. The coinfection of HDV with HBV in adults usually results in the clearance of both viruses, and just a little percentage of these patients develops a chronic infection. In contrast, it has been estimated that 90% of the superinfected patients develops a persistent HBV/HDV infection with a risk of developing liver cirrhosis, HCC and even liver failure of 50-70% within 5-10 years [147]. Moreover, the risk of developing a fulminant hepatitis is greater in HDV/HBV patients than in HBV monoinfected or in presence of other viral hepatitis [150,151].

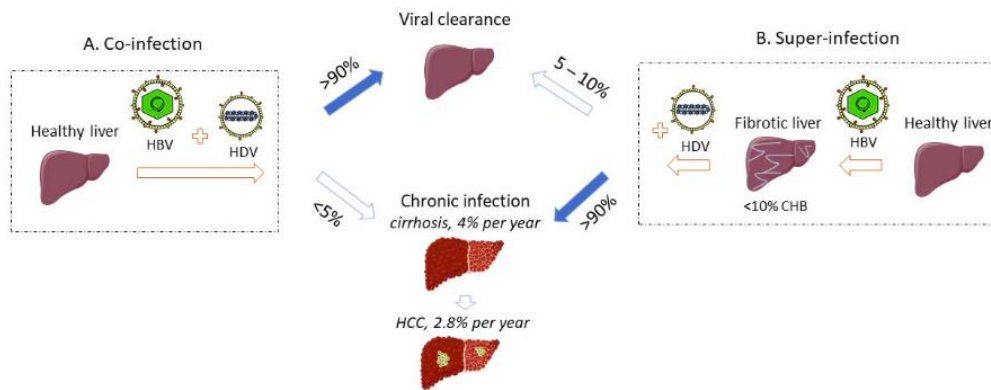


Figure 15. HDV-HBV coinfection and superinfection. The picture shows the clinical evolution of the two possible scenarios of HDV infection (coinfection and superinfection). The incidence of each clinical condition is reported, including the incidence per year of cirrhosis and HCC [152].

Notably, although HDV needs its helper virus, the replication of HBV is strongly inhibited. A large reduction in plasma HBV DNA is usually observed, and a reduction in both intrahepatic HBV rcDNA and cccDNA has also been reported [153]. This inhibition of HBV does not affect levels of HBsAg, a component of the HDV virion, and correlates with HDV-RNA [154]. Nevertheless, the mechanism behind the inhibition of HBV in the context of HBV/HDV coinfection is still unknown.

2.1. HDV virion: a viroid-like particle

HDV is a viroid-like virus. Viroids are naked infectious agents formed by a non-coding single-stranded circular RNA molecule, which can present branched or rod-like structures and are commonly known to infect plants. Viroids are believed to be ancestral, left over from the ancient world of RNA, which is believed to have existed before the emergence of DNA and proteins. [155,156]

Although differently mammals infecting *Deltaviruses* have been reported, HDV is the only member of this genus able to infect humans [155]. Like plant viroids, HDV is composed by a circular single-stranded RNA molecule provided of a catalytically active domain known as ribozyme, whose function will be extensively explained later in the present chapter. However, differently from the classic viroids, HDV genome is 4-7-fold higher than viroids (~246 bp in the smallest viroid and 1.6 kb in

HDV) [155,156]. Other element that differentiates HDV from classical viroids, is that HDV can produce its own specific viral protein called delta antigen (HDAG). This last exists in two different-in-length isoforms, produced by the same ORF: the short (S-HDAG) and the large (L-HDAG) delta antigens [157].

Hepatitis delta virus is the smallest virus present among mammalian viruses, being about 36 nm in diameter [158]. The delta virion consists of a ribonucleoprotein (RNP) complex formed by a copy of the viral RNA coupled to the S- and L-HDAG (Figure 16). The RNP complex is then enveloped by a plasma membrane modified with the three forms of the HBV surface antigen (as explained in this chapter in section 1.1)[142]. HDV is a satellite virus of the HBV since it lacks the surface antigens necessary to leave and infect the hepatic cells.

The HDV genome is a single stranded circular RNA of negative polarity ($3' \rightarrow 5'$), with a size of 1,668-1,697 nucleotides depending on the genotype. Its sequence has a high G/C content (around 60%) and a self-complementarity of 74% which gives rise to a stable rod-like secondary structure, resembling genome plant viroid's structure [159,160].

Despite its "simple" structure, by a virological point of view, HDV infection is quite complex since it takes advantage of the host machinery to promote the expression of its own antigen and to promote its life cycle. In the next section of my thesis, I will describe all the features that make HDV such a sophisticated virus.

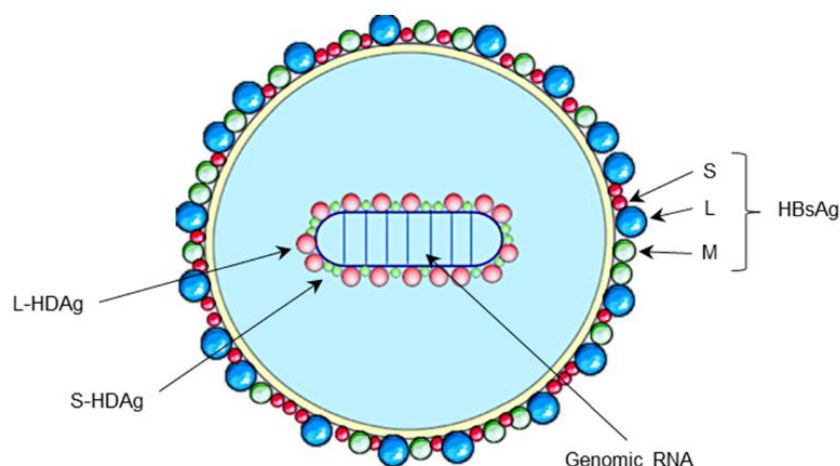


Figure 16. Hepatitis delta virus virion. HDV virion is formed by a ribonucleoprotein complex formed by the viral genomic RNA coupled to both the short and large delta antigen (S- and L-HDAg), and an envelope presenting the HBV surface antigens (S-, M and L-HBsAg) [161].

2.2. HDV life cycle

Since HDV shares with HBV the surface antigens, their mechanism of entry is similar (section 1.3 of this chapter, entitled “HBV life cycle”). However, another system of transmission has been recently described. HDV can spread by cell division, meaning that the viral genome molecules can be transferred from the mother to the daughter cells during cell division, thus promoting HDV spread by an entry-independent mechanism [162] (Figure 17). This may explain clinical studies in which HDV persists after one year in patients with liver transplantation and anti-HBsAg immunoprophylaxis, and after six weeks in HDV monoinfected mice with transplanted PHHs [163–165].

Once inside the hepatocyte, the RNP complex reaches the nucleus thanks to the nuclear localization sequences (NLS) of the HDAg (between aa 69 to 89) and enters the nucleoplasm through the interaction with the cellular importins that recognize specific sequences for transport through the nuclear pores [166,167]. This is another point of difference between HDV and classical viroid since the latter can transport itself to the nucleus after infecting the plant cell cytoplasm [166,168].

Once inside nucleus, the genomic RNA is transcribed by the cellular RNA polymerase II to produce the mRNA encoding the HDAg, which is essential to promote the viral replication.

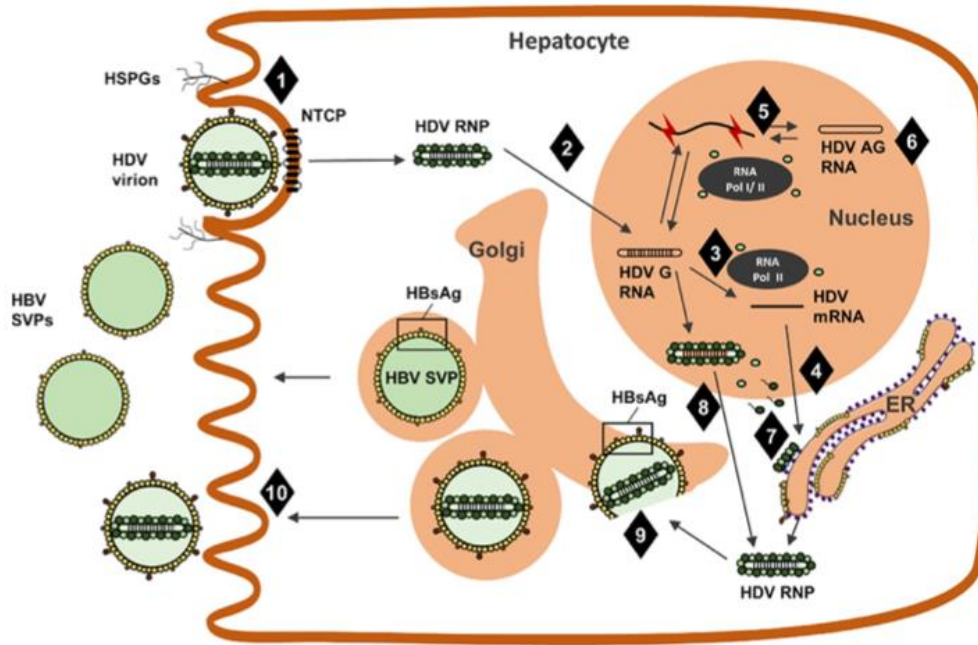


Figure 17. HDV replication cycle. HDV enters into the hepatocyte through the interaction of the L-HBsAg with HSPGs and the NCTP receptor (step 1). The ribonucleoprotein complex (RNP) reaches the nucleus (step 2), and it is used as template to produce that mRNA (step 3) that will be translated into the HDAg proteins, both S- and L-HDAg (step 4). This step is led by the cellular RNA polymerase II. In addition to the expression of mRNAs, the genome replicates by forming concatemers of HDV antigenome (through the action of the cellular RNA pol I) from which the antigenome molecules (5'->3') are released. The antigenome is then used to produce, thanks to the cellular RNA polymerase II, a new concatemer of genome molecules (steps 5 and 6). Both genomes and antigenomes are produced by a Rolling Circle Amplification (RCA) and the individual monomers are released from the concatemers by a ribozyme, a portion of RNA with enzymatic activity. The L-HDAg proteins produced during the steps 3 and 4 are subjected to several post-translational modifications, especially the farnesylation which is necessary for viral assembly (step 7). Both the S- and L-HDAg are coupled to the viral genome thus forming new RNPs (step 8) that are exported to the cytoplasm and, thanks to the interaction of L-HDAg with HBsAg, acquires the envelope (step 9). The new viral particles can so be released into the bloodstream and infect new hepatocytes [169].

HDV genome replicates through a rolling circle amplification (RCA) process and, since the virus is not provided of a viral polymerase, it uses the host DNA-dependent RNA polymerase system, like viroids (Figure 18). The RCA consists of the unidirectional amplification of a circular nucleic acid molecule, and it is quite common in microorganisms (for example in bacteria's plasmid) to produce multiple

copies of the target molecule. The first RCA step in HDV life cycle starts from the circular RNA genome (3'→5') and generates a linear concatemer (Figure 18, panel A) formed by numerous complementary monomer copies with a positive orientation (5'→3') [170]. This process is expected to be carried out by the cellular RNA polymerase I. These monomer units will be then freed by the self-cleavage activity of the own viral ribozyme activity (the ribozyme will be deeply commented in the next paragraphs). This process consists of a transesterification reaction 33 nt downstream of the polyadenylation site of the *HDAG* with the liberation of linear antigenome molecules. The released antigenomes are circularized by a yet unknown mechanism and are used to repeat the RCA process [171,172] (Figure 18, panel B).

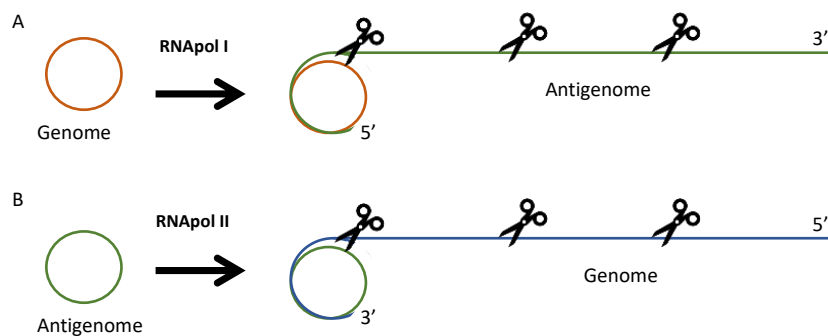


Figure 18. Rolling circle amplification (RCA) of HDV genome. A). The RCA starts with the amplification, led by the cellular RNA polymerase I (RNA pol I), of the genome sequence to produce a long linear molecule composed by multiple copies of the genome complementary sequence (concatemer with direction 5' -> 3'). From this concatemer, the antigenomic monomers are released thanks to the activity of the viral ribozyme (here represented with a scissor). B. The antigenomes previously produced are used as templates for another RCA cycle, this time driven by the cellular RNA polymerase II (RNA pol II) to generate another concatemer with multiple copies of the genome sequence (direction 3' -> 5'). Again, the different monomers are released through the viral ribozyme.

This second RCA process will give rise to a concatemer with several copies of the genome sequence. Again, the single unit of the genome will be self-cleaved by the viral ribozyme thus producing multiple monomers of circular genomic RNA [173,174]. The mechanism behind the circularization of the genomic RNA is still unclear. It might occur through a host RNA ligase [175], or through a possible self-ligation mechanism [176].

The circularized genomic molecules will then interact with the HDAG proteins (both short and large) into the nucleus thus forming the RNP complexes. The RNPs are so exported to the cytoplasm and thanks to the interaction of the L-HDAG with the HBsAg, new viral particles are assembled in the endoplasmic reticulum. New virions are so released into the bloodstream to infect the neighboring hepatocytes. Of note, more information about the function of the HDAG and ribozyme will be detailed in the next paragraphs.

2.3. HDV self-cleaving property: ribozyme

As mentioned before in the preceding section, the ribozyme is essential to release the genome and antigenome monomers from the concatemers produced during the RCA process. Crystallized structure was obtained for the first time in 1998 by Dr. Doudna's group. The sequence found in genomic (nucleotides 688-771) and antigenomic (nucleotides 900-948) RNA, was not identical, but had the same structure [177], this suggests that the preservation of the secondary structure of this region is mandatory to ensure its cleavage activity (Figure 19). It is composed by five helical regions (P1, P2,3 P1.1 and P4) that together form a nested pseudoknot, three single stranded regions (J1/2, J1.1/4 and J4/2) and two loop structures (L3 and L4) (Figure 19) [178]. Of note, in figure 19, the L3 and L4 are not highlighted.

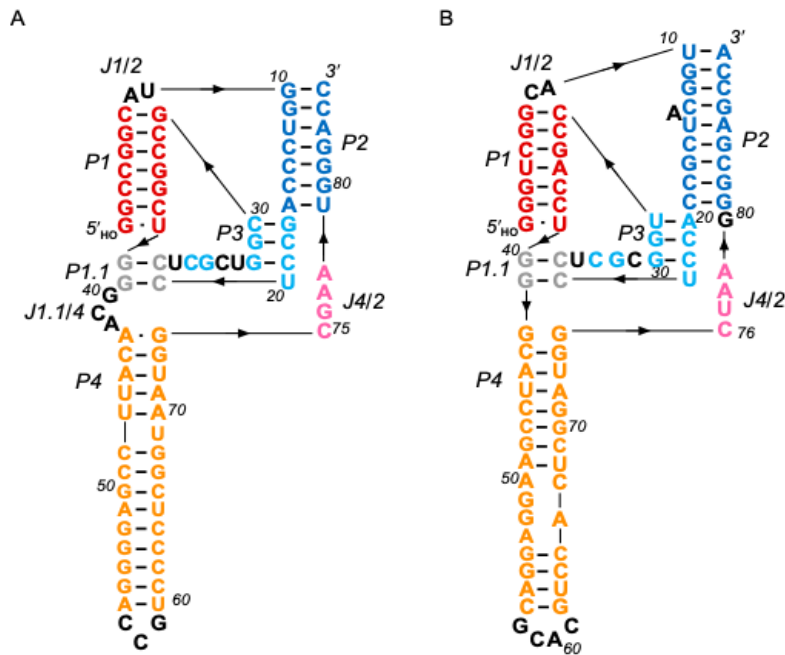


Figure 19. HDV ribozyme secondary structure in genomic (A) and antigenomic (B) sense. Each domain is highlighted by different colors. The helices are represented respectively in red (P1), dark blue (P2), light blue (P3), grey (P1.1) and orange (P4). The single stranded regions J1/2 and J1.1/4 are in black and the J4/2, which contains the catalytic cytosine, is in pink [177].

Although the entire secondary structure is essential for the activity, by a functional point of view, each region covers a role (Table 1). The P2 and P4 regions are important to stabilize the structure, whereas the P3, J4/2 and L3 regions are especially involved in the ribozyme's catalytic activity [177,179–182]. Moreover, the P1.1 pseudoknot has a significant role in promoting the catalytic activity since it promotes the "connection" between the catalytic domain and the substrate of cleavage (in P1), as reported by testing several mutants introduced in this domain [183,184]. The J4/2 portion also contains a cysteine at position 75 on the HDV genome, which is thought to be a key element in ribozyme self-cleavage. Experiments of mutagenesis and cross-linking suggested that this nucleotide is important in catalysis, since it could be involved in a dynamic network of hydrogen-bond interaction that helps the "activation" of the hydroxyl group, leading actor of the reaction of transesterification [178].

Table 1. HDV ribozyme regions functions. The table shows the function of each structural region of the HDV ribozyme, and their positions based on genomic orientation. Table from Been M, et al. 1997 [182].

Domain	Function/s	Position
P1	Presents the self-splicing site (nt 1). Requires base pairing to ensure proper self-splicing.	1-7/31-37
J1/2	Its function is still not totally known.	8-9
P2	Stabilizes the structure once the split has happened.	10-16/79-86
P3	Participates in the catalysis.	17-20/28-30
P1.1	Like P1, requires base pairing to ensure a correct self-splicing. G38 interacts with Mg ²⁺ during self-splicing.	21-22/38-39
L3	It is essential for the correct split.	23-27
J1.1/4	It presents some preserved residues which interact with Mg ²⁺ during self-splicing.	40-42
P4	Stabilizes the structure of the ribozyme. It does not participate in catalysis.	43-56/60-74
L4	Mainly structural region.	57-59
J4/2	It includes nt C75, which is very important for the catalytic activity of ribozyme as it interacts with Mg ²⁺ ions and promotes an acid-base catalysis during the self-splicing.	75-78

As cited before in the previous paragraph, the ribozyme release the monomers from the concatemer by a transesterification in which the hydroxyl group at 2' in the ribose of the nucleotide at position -1 (before the first nucleotide of the ribozyme) produces a nucleophilic “attack” to the contiguous phosphate group (at 3'), thus forming a 2'-3' cyclic phosphate (at nt -1) and a free hydroxyl group at 5' (at nt 1 in ribozyme) [185]. This cleaving mechanism appears to be the same as those described for plant viroids, virusoids and satellite RNAs [186–188]. Notably, to go ahead with the reaction divalent metal ions are required, such as calcium, magnesium, and manganese. The reaction could also occur in an alternative manner in the presence of monovalent ions, but the catalytic activity is greatly reduced [189].

2.4. The multifunctional HDAg protein

Another characteristic of HDV that distinguishes it from classical viroids is its ability to encode a specific viral protein. As commented before, it exists in two isoforms, the short S-HDAg, that consists of 195 amino acids and has a molecular weight of 24 kDa, and the long L-HDAg, which shares the same N-terminal sequence but has 19 amino acids more than the shorter isoform (214 amino acids with a molecular

weight of 27 kDa) [190]. Notably, both proteins are produced from the same ORF following the deamination of the adenosine presented in the S-HDAg codon stop.

This process, known as editing, is carried out by the cellular adenosine deaminase acting on RNA (ADAR) 1. ADAR1 is the most common editing form in *metazoa*. This cellular enzyme acts on double stranded RNA molecules (dsRNA). In humans the ADAR family presents two active forms (ADAR1 and ADAR2) and one inactive form (ADAR3) of the enzyme. ADAR1 is the one responsible for the HDV genome editing [191], and it has two major splicing isoforms, a small one of 110kDa (p110), which is constitutively and predominantly expressed in the cytoplasm, and a longer isoform of 150 kDa (p150) which is localizable in both the nucleus and the cytoplasm, and whose expression is triggered by the activation the interferon (IFN) pathway [192,193]. About the target, ADAR1 generally edits both short (around 20-30 bp) and long dsRNA presenting mismatched bases, bulges, and loops. In addition, ADAR1 has been shown to primarily edit extremes, preferring adenosines present at the 3' end of the RNA [194,195]. This enzyme deaminates the C6 of the adenosine ring thus forming an Inosine (I), which is similar, in term of sequence complementarity, to a guanosine.

During HDV replication, ADAR1 acts on the HDV antigenome where it deaminates the adenosine at position 1012, leaving an inosine (Figure 20). During the amplification of the antigenic material into genomic RNA, this modified nucleoside (I) is considered as a guanosine, resulting in the introduction of a cytosine (C) instead of an uracil (U) in the genomic sequence, and the codon in the genome is changed from AUC to ACC. When the genome is transcribed to produce the HDAg mRNA or new molecules of antigenomes, the modified codon generates, due to sequence complementarity, a UGG codon, thus modifying the Amber stop codon (UAG) in position 196. As a result, the modified mRNA expresses a tryptophan (*196W), thus resulting in the longer HDAg protein (L-HDAg). Of note, this process is not absolute, with each mean that during viral life cycle, both HDAg will be produced for respectively edited (L-HDAG) and unedited (S-HDAG) mRNAs [196–198].

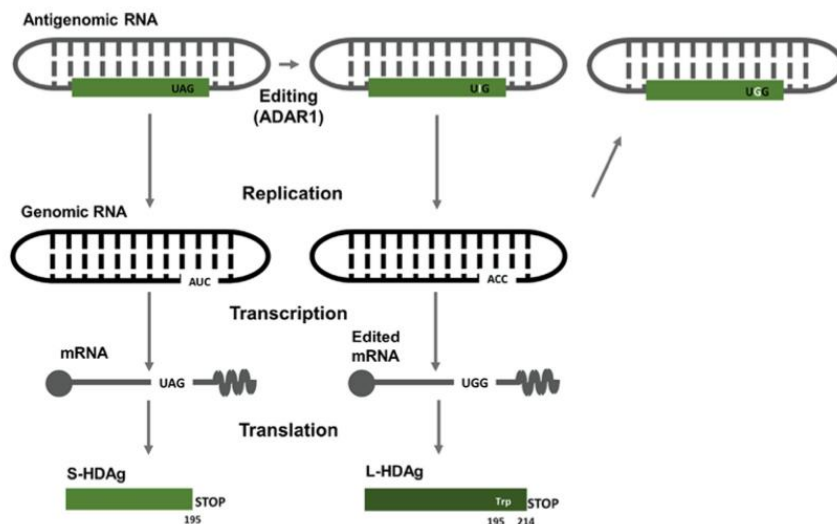


Figure 20. ADAR1 editing process on the HDV antigenome. ADAR1 catalyzes the editing of the antigenome RNA at nucleotide 1012 (aa 196; amber/W). The adenosine (A) at position 1012 is converted to an inosine (I), which is considered a guanine (G). During the next RCA step, the edited antigenome produces an edited genome that presents an ACC codon instead of AUC. When new mRNAs encoding the HDAg are produced from the genome monomers, the mRNA obtained from the edited genome will carry the UGG codon. This editing codon introduces an amino acid change from a stop codon (in the unedited mRNA) to a tryptophan coding codon (from the edited mRNA) at position 196, resulting in a 19 aa longer protein, the L-HDAg [169].

Although the two isoforms are both structural elements of the virions, they also play essential roles in viral replication [199].

The S-HDAg, produced in the early stages of the viral cycle, is essential for viral replication since it is necessary for the recruitment of the transcription machinery and promotes the elongation of viral RNA [200–202]. Studies have shown that S-HDAg interacts with several subunits of especially RNA polymerase II [203]. Both HDAGs have a sequence similarity with a Negative Elongation Factor (NEF), which inhibits RNA pol II elongation process. This similarity allows the HDAGs to interact with the RNA pol occupying the NEF site, thus promoting RNA elongation [204]. Moreover, the S-HDAg directly interacts with the RNA pol II clamp, which is a region of the RNA pol II that holds DNA and RNA in place, thus affecting RNA pol fidelity and its shift of template, from DNA to RNA [202]. On the contrary, L-HDAg begins to accumulate lately related the small isoform. Its main function is the assembly of virions [205]. It has been suggested that the balance between the ability of S-HDAg to promote viral replication and the ability of L-HDAg to enhance particle release is related to the amount of L-HDAg that accumulates during the time in the nucleus:

the more protein, the more ribonucleoprotein complex is ready to be exported to the cytoplasm [170].

When looking the protein sequence, several functional domains have been identified in the HDAg (Figure 21), most of them shared by both isoforms.

The RNA Binding Domain (RBD) eases the interaction with the viral RNA when forming the RNP complex. This domain contains two arginine-rich motifs between aa 97-107 and aa 136-146 (ARM) and has helix-loop-helix (HLH) between the ARM [206]. In the RNP the HDAg monomers are subjected to antiparallel coiled-coil mediated dimerization and dimers then form homo- and hetero-multimeric octamers disposed in groups of 4-5 around which the viral RNA is wrapped [170]. The dimerization occurs at the coiled-coil domain (CCD) between aa 12 and 60 [206]. The nuclear localization signal or NLS (aa 66 and 75), moreover, is important for the entry of the RNP into the nucleus [167,207,208] , whereas the Pol II Binding Domain (aa 150-195) allows the binding to the host RNA pol II during viral replication. In addition, the ribozyme binding site is located between amino acids 24 and 50 [204,209].

Related to the S-HDAg, the large isoform in the C-terminal portion of the L-HDAg (aa 195-214) resides the viral assembly sequence (VAS), which presents: the nuclear export signal (NES) between aa 198-210 and the isoprenylation signal between aa 211-214 (Iso), important for the following farnesylation of the L-HDAg. This VAS domain is essential for the formation of viral particles, due to the interaction between the ribonucleoprotein and HBsAg [210,211].

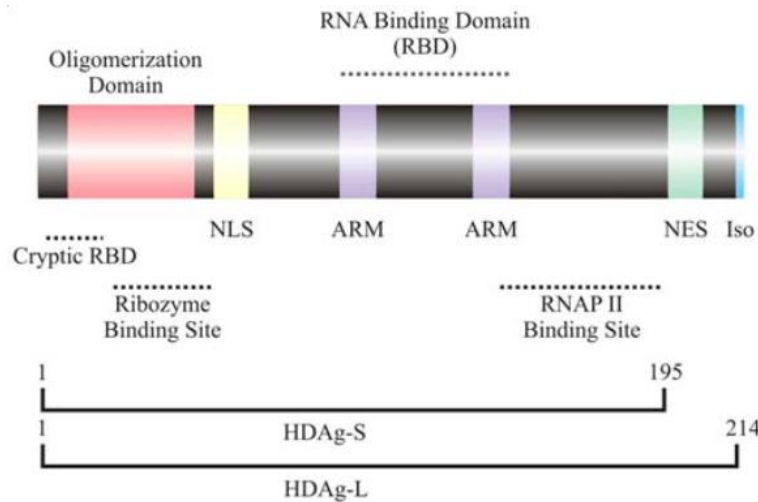


Figure 21. Schematic representation of the HDag functional domains. The different domains included in S-HDag and L-HDag are shown. NLS: Nuclear localization signal, ARM: arginine-rich motif, NES: Nuclear export signal, RBD: RNA-binding domain. Iso: Isoprenylation signal. The oligomerization domain (pink) and the Ribozyme and RNA polymerase II binding domains (dashed lines) are represented [203].

Of note, both the S- and L-HDag undergo to post-translational modifications such as phosphorylation, acetylation, or methylations, that promote the efficiency of the viral replication [212,213] (Figure 22).

Among all these modifications, a particular interest should be focused on the isoprenylation at the C-terminus of the L-HDag in the cysteine 211. This process consists in the addition of a farnesyl phenyl lipid moiety to the cysteine that is important for inhibiting viral replication and contributes to virion assembly by assisting the binding between L-HDag and HBV surface proteins [203]. This step is targeted by a therapeutic strategy currently under investigation, which consists of inhibiting farnesyltransferase with an inhibitor (Lonafarnib), thus preventing virion assembly [214].

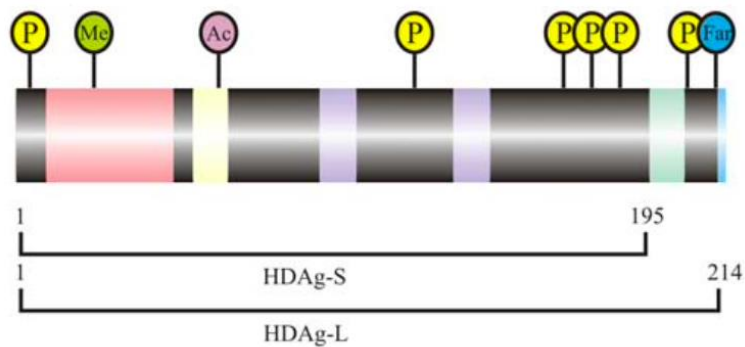


Figure 22. Schematic representation of the HDAg post-translational modifications. The post-translational modifications that occur during HDV replication are represented. P: phosphorylation, Me: methylation, Ac: acetylation, Far: farnesylation. The sumoylation of multiple lysine residues of S-HDAg has been reported [203].

In addition, both S-HDAg and L-HDAg have a high immunogenic capacity, presenting epitopic regions for B-cells, CD4⁺ and CD8⁺ T lymphocytes [215]. Studies have described that the association of HDAg aa changes with polymorphisms in the human leukocyte antigen may be associated with immune escape [216].

2.5. Current therapy against HDV

Since the 1980s, the main therapeutic protocol for patients with CHD was interferon Alpha (IFN α) or pegylated (peg-IFN α). This strategy, however, reached a virological response (undetectable HDV RNA levels) in just 17-47% of the patients (without HBsAg seroconversion) [217], and more than 60-97% of the treated patients experienced viral relapses [218]. It has been calculated that just one-fifth and one-third of the IFN α or peg-IFN α - treated patients achieve HDV clearance [219]. Moreover, this treatment has important side effects that can strongly affect patients' quality of life [57].

Of note, during the last years, the scientific research has actively improved the knowledge about HDV virology, contributing to the development of new therapeutic strategies. The main example is the Bulevirtide, which has been recently approved by the European Medicine Agency to treat CHD patients with compensated cirrhosis. As previously cited in section 1.2.3 of this chapter, this drug is a 47-nt acetylated fragment derived from the N-terminal domain of HBV preS1 HBsAg, which acts by

competing for the attachment of HBsAg to the NTCP receptor, thus inhibiting HBV and HDV entry into the hepatocyte (Figure 16) [220,221]. It is currently in phase III clinical trials [222].

Other strategies are based on the inhibition of the farnesyltransferase (Lonafarnib, LNF), which is a relevant step in viral replication (section 2.2 of this chapter), or in blocking HBsAg release through the nucleic acid polymers (NAPs) such as REP-2139-Mg or REP-2165-Mg (Figure 23) [223]. Studies conducted in CHB patients showed that these drugs, in combination with immunotherapies, had a strong antiviral activity [221,224]. REP-2139 is currently in phase II clinical trial. Other valuable therapeutic options are the new immunomodulators, especially the lambda interferon (IFN λ) which is currently in phase II of clinical trials [223].

Of note, all these new procedures do not directly interfere with HDV itself. A gene therapy strategy based on gene silencing could be extremely helpful to control this infection, since it could inhibit viral replication and interfere with the liver disease by directly acting on HDV. Small interfering RNAs (siRNAs) or Antisense nucleotides (ASOs). The JNJ-3989 siRNA that has been tested in CHB patients and a decrease of the HBsAg has been observed and it is currently under evaluation in phase II clinical trial in patients co-infected with HDV [128,223]. There are several ASOs in clinical trials, such as the Bepirovirsen (GSK3228836). A recent study reported a HBsAg clearance at 29 days after administration of Bepirovirsen in the 23.5% of patients with CHB, both naive and patients on with NA therapy [225].

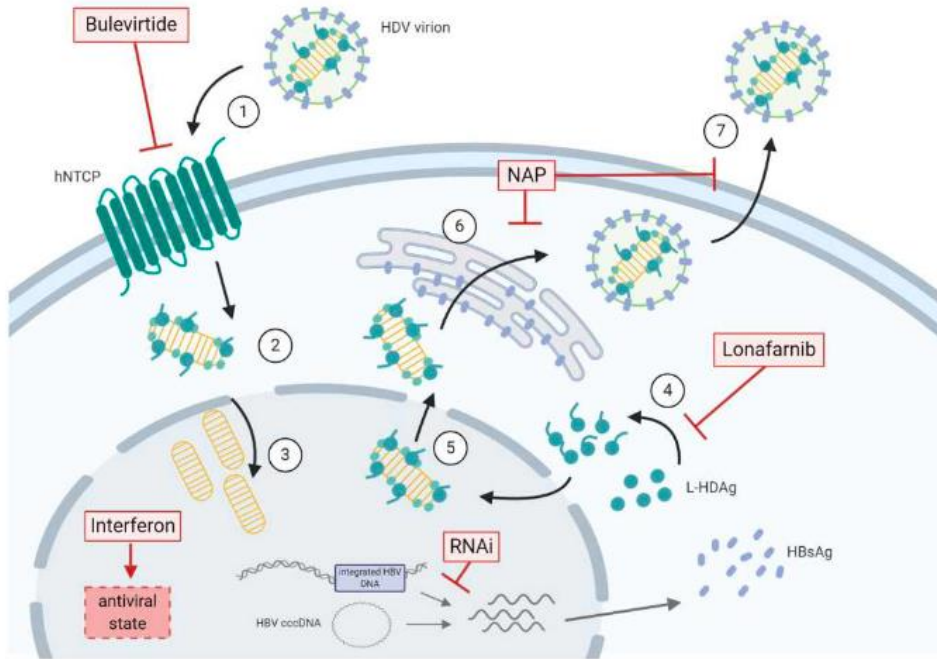


Figure 23. Current HDV therapeutic strategies. The picture shows the current therapeutic strategies (in red boxes) and the step of the life cycle in which each drug is active. RNAi: interfering RNA [223].

3. HDV-mediated HBV inhibition

Although HDV needs of the HBsAg to produce infectious viral particles, in presence of its satellite virus the replication of HBV is inhibited. Indeed, a significant reduction of HBV DNA levels has been reported in HBV/HDV patients [226–228], as well as a decrease of HBV replication intermediates in liver biopsies [153]. This viral interference was also confirmed *in vitro* in Huh7 cells transfected with DNA vectors expressing HBV and HDV [229], as well as in *in vivo* superinfection models in HBV infected chimpanzees, woodchuck hepatitis and in a humanized HBV-infected mouse model [230–233].

The mechanism behind this inhibition is still unknown. The fact that the inhibition can be completely reproduced *in vitro* might suggest that it could be independent of the adaptive immune system [229,233,234].

Some studies had suggested that the effect of the L-HDAg on cellular RNA pol II could limit the transcription of the cccDNA, without strongly affecting the expression of HBsAg which can also be produced from integrated DNA [235,236]. In addition, a reduction of the amount of the HBV RNAs were found in differentiated HepaRG cells *in vitro*, suggesting a possible inhibition of transcription or stability of HBV RNAs [234]. The inhibition of the transcription of cccDNA has been also suggested by another study that reported that both S- and L-HDAg are able to inhibit the activity of HBV Enhancers I and II [237].

Other possible source of inhibition can be related to the activation of the intracellular innate immunity. HBV is considered a “stealth” virus since it limitedly activates the intracellular Pattern Recognition Receptors (PRR) responsible of the activation of the intracellular IFN pathway. In contrast, HDV replication significantly induces innate immune response [238], and even take advantage of one of the interferon-stimulated gene (ISG), ADAR1, to produce the two delta antigen isoforms. It was reported, indeed, that when the HDV RNA reaches the cytoplasm, the ribonucleoprotein is recognized by one of the PRR, the MDA5 [239], which activates a mitochondrial antiviral signaling (MAVS) with the downstream activation of

transcription factors, such as IFN regulatory factor (IRF) 3/7 and nuclear factor- κ B (NF- κ B) [240], thus promoting the expression of type I IFNs (IFN α and IFN β) [241]. The expression and release of IFN is essential for the cellular response against viral infections [242,243]. In addition, the activation of the IFN response could promote the inhibition of the HBV expression through a direct mechanism the interferon-inducible cytoplasmic dynamin-like GTPase MxA. This latest could directly interact with HBcAg thus promoting a perinuclear accumulation of core antigen, limiting the formation and release of new viral particles without affecting viral mRNA levels [244]. Moreover, an *in vitro* study reported that the L-HDAg could increase of 3-fold the activation of the *MxA* promoter [237].

Of note, it is still unknown if the activation of the IFN pathway due to HDV could affect not only HBV expression, if not also its variability.

4. Viral quasispecies

A viral quasispecies (QS) is a dynamic mixture of closely related but not identical genomic variants called haplotypes. The term quasispecies was coined by Manfred Eigen and Peter Schuster in the 1970s to explain the organizational and adaptive capacity of replicative entities. The viral QS follows an ordered and stable distribution of mutants dominated by a principal genome, known as master sequence (those with the highest relative frequency) [245] (Figure 24). Differently, the term consensus refers to a sequence that shows the most common nucleotide (or amino acid) per each position in the QS.

The most important feature of a viral QS is its dynamicity, which is a product of several positive or negative selection processes where the best variants are selected. During each life cycle, viruses can acquire mutations that will generate new variants that might differently respond to the environment. Among all the generated variants, those that are less replicative might gradually disappear from the viral population. However, the presence of a selection pressure (such as an active immune response or antivirals) will maintain and select (positive pressure or genetic bottleneck) the variants that better respond to the modified environment (in some extents independently from the variant's fitness). These more resistant variants will generate a new viral population. When the pressure fails, the QS might potentially change again depending on new selection step and/or the viral fitness. Although the relative frequency of each haplotype can provide a measure of the fitness or of the selection of the unique variant, this value cannot optimally represent the complexity and variability of the QS [246]. In this doctoral thesis, the viral quasispecies has been studied by using several complexity and functional indices, that will be deeply described later, in the following chapters.

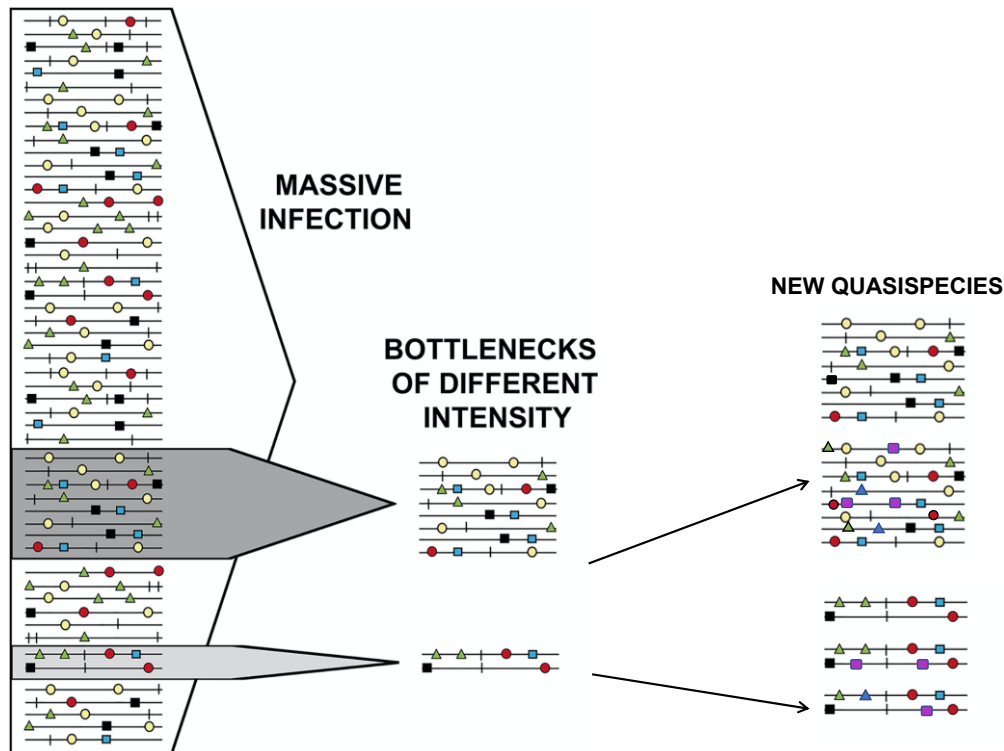


Figure 24. Schematic representation of a viral quasispecies and the bottleneck process. An infected individual will present a mutant spectrum corresponding to the viral population. This viral population will be composed by different closely related variants. Changes in the environment can provoke a bottleneck process, a selection process that will allow the survival of just those mutants able to adapt to this changing event. The bottleneck can vary in intensity depending on how restrictive the selection is. From the selected mutants, a new quasispecies will be formed due to the acquisition of new mutations. The different colored geometric shapes represent the acquired mutations.

The inspection of the dominant sequences (master) of a viral population can be a determining factor in the study of the course of the disease, the response to the host immune system or even to episodes of clinical relapses. On the other hand, the high variability of some viruses, allows them to re-adapt to a changing environment. For this reason, a deep study of the complete viral population, including minority variants that forms a QS, might be required [247,248]. Of note, HBV and HDV are both two extremely variable viruses and circulate as QS.

4.1. HBV variability

Although it is a DNA virus, its replication consists of a retrotranscription step. Of note, the viral retro-transcriptase lacks proof-reading activity, so it can introduce 1 mutation each 10^5 to 10^7 bases [249], thing that contributes to an evolution rate of 1×10^{-5} to 1×10^{-4} substitution/site/year [250–252], closed to those of RNA viruses [253]. Considering the high rate of viral replication (around 10^{13} viral particles per day in patients with acute infection and 10^{11} virions/day in patients with chronic hepatitis) [254,255], several variants can be produced per day of infection, generating a complex quasispecies.

In addition to that, intracellular antiviral response could potentially contribute to HBV QS variability. Some members of the AID/APOBEC deaminase family can hyper edit HBV as shown for the human immunodeficiency virus [256,257]. *In vivo* studies have reported that APOBEC had a deamination frequency of $\geq 1 \times 10^{-2}$ substitution/nucleotide in serum samples [258–260]. It has been reported that APOBEC deaminates cytidines in the intermediate of viral single-stranded DNA (during the retrotranscription step of the minus strand- C to U mutation), and adenosines are transcribed from uridines, leading to G-to-A hypermutations in second-strand DNA [261]. However, a study performed on HBV DNA from a database of patients with and without HCC observed that patients with HCC showed a higher ratio of APOBEC mutations in the negative strand of rcDNA than those without HCC, however, mutations in the positive strand of rcDNA were higher in those patients without HCC, thus suggesting that deamination might also occur during the formation or transcription of cccDNA in the nucleus. [262].

A proof of the high HBV variability is the presence of different viral genotypes and subgenotypes (Figure 25). HBV genomes have been classified in 10 genotypes (from A to I) with a sequence divergence of $\geq 8\%$. Within these genotypes, at least 35 subgenotypes have been identified, with a 4-8% divergence between them [263,264]. The various genotypes show a different geographical distribution, thus suggesting a different geographical evolution of the viral QS (Figure 25).

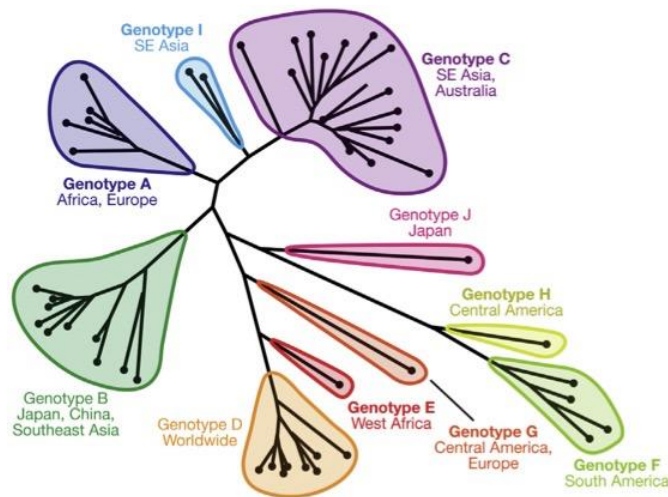


Figure 25. Phylogenetic three of HBV genotypes. The tree shows the genetic divergence between the different HBV genotypes (A-I) and prevalent sub-genotypes. The geographical area where each genotype is more prevalent is as reported. Figure from McNaught [265].

As commented before with each specific viral protein (throughout section 1.2 of this chapter), HBV variability can strongly impact the infection and the liver disease progression, such as the acquisition of mutations that confer resistance to antiviral treatment. The genotype itself can differently impact the antiviral response or disease progression. A recent study tested *in vitro* the expression of the different viral genotypes thus showing that A2 genotype mainly expressed precore mRNA than pgRNA and is less responsive to Lamivudine inhibition related to other genotypes such as genotype B2 or D1 [266]. In addition, genotypes C and D and subgenotype A1 have been identified to be associated with a higher risk of cirrhosis and HCC than other subgenotypes such as A2, B1 and B6 [267]. This variability is further enhanced due to the ability of different genotypes to recombine as mainly product of the co-infection of the same cells [264,268].

Considering this extreme high level of variability and how it influences the clinics, the study of viral QS could strongly help to identify prognostic or therapeutic factors, thus adapting the viral population evolution to patient's follow-up as a kind of personalized medicine [55].

4.2. HDV variability

Although the HDV genome replicates thanks to the cellular RNA polymerase (see section 2.2 of this chapter), which is provided of a 3'→5' proofreading exonuclease activity (with a very low error rate of (~1.00E-06 substitutions/nt) [269]. Nevertheless, HDV is the most variable virus among hepatitis viruses, with an evolution rate between 9.5E-03 to 1.2E-03 substitutions/site/year [270].

The most accepted theory considered that the cellular RNA pol can introduce errors (insertions, small nucleotide deletions or intramolecular recombination) due to the switch of template from DNA to RNA [271,272]. This hypothesis can be further reinforced by considering that the S-HDAg binds the RNA pol II, accelerating the direct translocation of this polymerase at the expense of its fidelity [202].

A recent study published by our group showed that in HDV QS the type of changes that occurred were mainly transitions (G↔A and C↔T) than transversions, thus suggesting a possible role of mutagenic enzyme (APOBEC, ADAR1) in HDV variability [273].

Again, a proof of this variability is the high divergence of the 8 genotypes, which is up to 16% within the same genotype and between 20% and 40% between different genotypes [274]. To cite some instances, between genotype 1a and 1b there is about 12% of dissimilarity, whereas genotype 3 is the more divergent comparing with the other ones (nearly 38%) (Figure 26) [275]. Like HBV, excepting genotype 1, which is worldwide distributed, the other genotypes are usually found in specific geographic areas are shown in Figure 26, panel A.

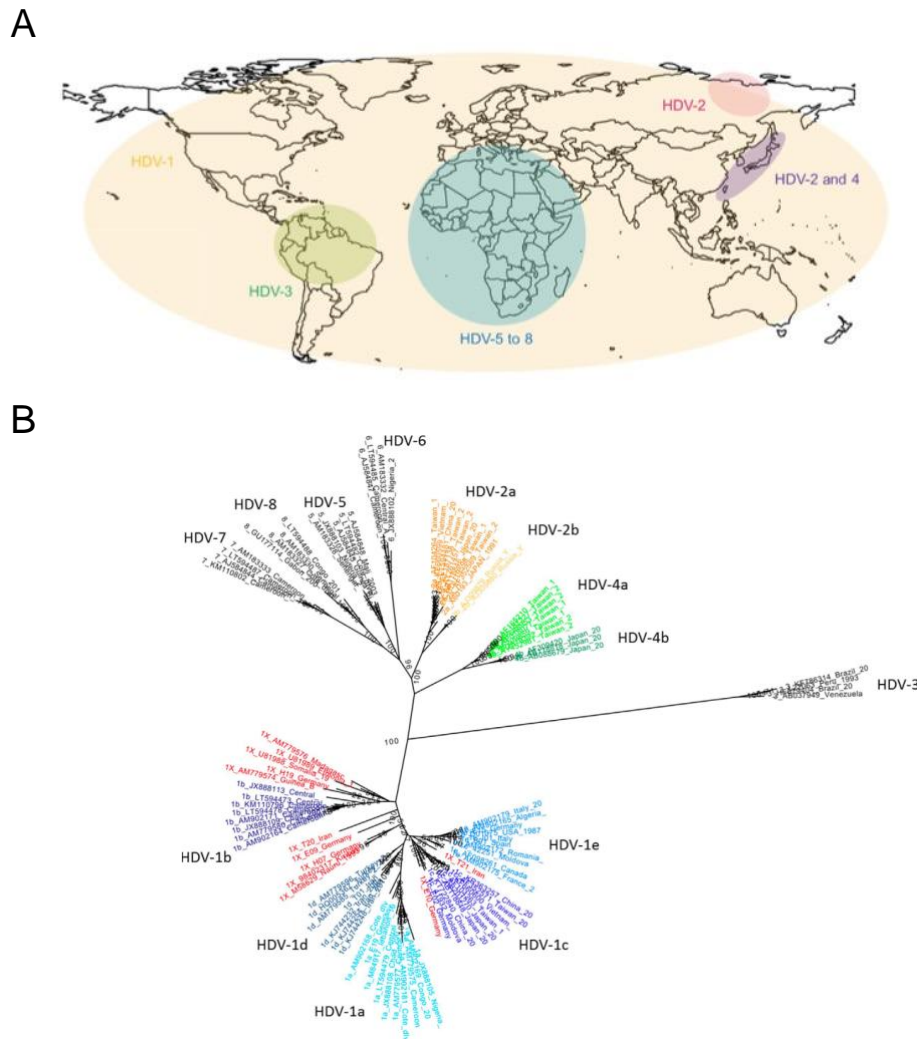


Figure 26. Global distribution and phylogenetic tree of HDV genotypes. Panel A. Genotype 1 (HDV-1) is worldwide distributed, whereas genotype 2 and 4 (HDV-2 and HDV-4) are mainly found in Taiwan and Japan. Genotype 2 (HDV-2) is also found in some regions of Russia. Genotype 3 (HDV-3) is mainly found in South America (mainly in the Amazon region). Genotype 5 to 8 (HDV-5 to HDV-8) are characteristics of the African continent [276]. Panel B. Phylogenetic tree representing the genetic divergence between the different HDV genotypes. Each of the genotypes (HDV 1-8) and the sub-genotypes are represented. This panel has been obtained from <https://abdominalkey.com/19-hepatitis-d/>.

As reported for HBV, the genotypes can affect the liver disease progression. Genotypes 1, 2 and 4, for example, have been related with milder liver disease than the genotype 3, which has been associated with a more severe course of acute infections and a higher risk of acute liver failure [277–279]. Even the sub-genotypes can influence the clinics, as reported in a study that shows that patients affected by genotype 1 can show a different progression of the disease depending on their geographical origin, suggesting that the African sub-genotype can be less pathogenic than Caucasian sub-genotype 1 [280].

The variability of HDV may also be manifested by the introduction of amino acid changes in the HDAg, which may interfere with the viral-specific adaptive immune response, such as previously reported in section 2.4 of this chapter.

5. Studying viral quasispecies: next generation sequencing (NGS)

During the past years, the first-generation sequencing techniques such as SANGER enabled studying viral genome variability, helping to genotype viruses and to identify mutations associated with antiviral resistance or immune escape. Although this technique facilitated the study of the consensus sequence (as reported before, the most common nucleotide in each position), it showed a big limitation since it cannot detect variants at a frequency <15-20% [281], and consequently it cannot be helpful to study the viral QS and its plasticity [55].

Differently, the next generation sequencing (NGS) produces millions of sequences permitting a deep study of the viral population [248]. This system is extremely helpful to study the quasispecies complexity, variability, and plasticity, and it allows to identify variants and their relative frequency in the population. Among the techniques belonging to this system, the Illumina platform has a high throughput, with less susceptibility to homopolymer errors and a limited error rate (introduction of single nucleotide substitution rate of ~0.1%) [55]. As commented in the previous paragraphs, HBV and HDV QS complexity and variability could play a role in the progression of the infection and/or the liver disease. In this thesis, the NGS and specifically the Illumina MiSeq platform have been used to analyze the viral QS of HDV, to inspect its evolution during the time, and to identify possible targets of new therapeutic strategy based on gene silencing. Thanks to its depth in sequencing, it has been also used to study how HDV and the activation of type I interferon can affect HBV variability. The application of this technology to achieve both aims shows its high applicability in both translational and basic research.

The Illumina technique is based on the formation of several clusters of identical molecules, each containing thousands of sequences. Moreover, the introduced nucleotide is recognized at the time of its insertion thanks to a procedure called Sequencing by Synthesis (SBS). Of note, the main steps are detailed below.

5.1. Cluster generation

The clusters are groups of clonally amplified identical sequences obtained from the previously prepared libraries (collections of amplified genetic fragments, hereafter called amplicons). As previously commented, each cluster contain thousands copies of the same sequence, thus generating a clearly bright spot that enables the correct identification of the nucleotide introduced during sequencing.

Clusters are generated in the flowcell (Figure 27, panel A), a glass device with channels randomly coated with an oligonucleotide lawn whose sequences are complementary to the two adapters (respectively forward and reverse) previously added to the amplicon during the preparation of the libraries. These adapters are introduced at the 5' and 3'-end of the amplicons (Figure 27, panel B) and comprise the P5 or P7 sequence, which are complementary to the oligos on the flowcell (respectively green and orange in Figure 18), and Rd1 and Rd2 sequences that are the targets of the Illumina primers used during the sequencing process.

In addition, the library's amplicons also include an M13 tail, which is a 20 nt-long tail added at the extremes of the amplicon during the nested PCR step. This tail is then used as target for the last nested-PCR step, when a sample-specific multiplex identifier (MID) has been added. This last one will help during the analysis step to identify the specific analyzed sample (Figure 27, panel C).

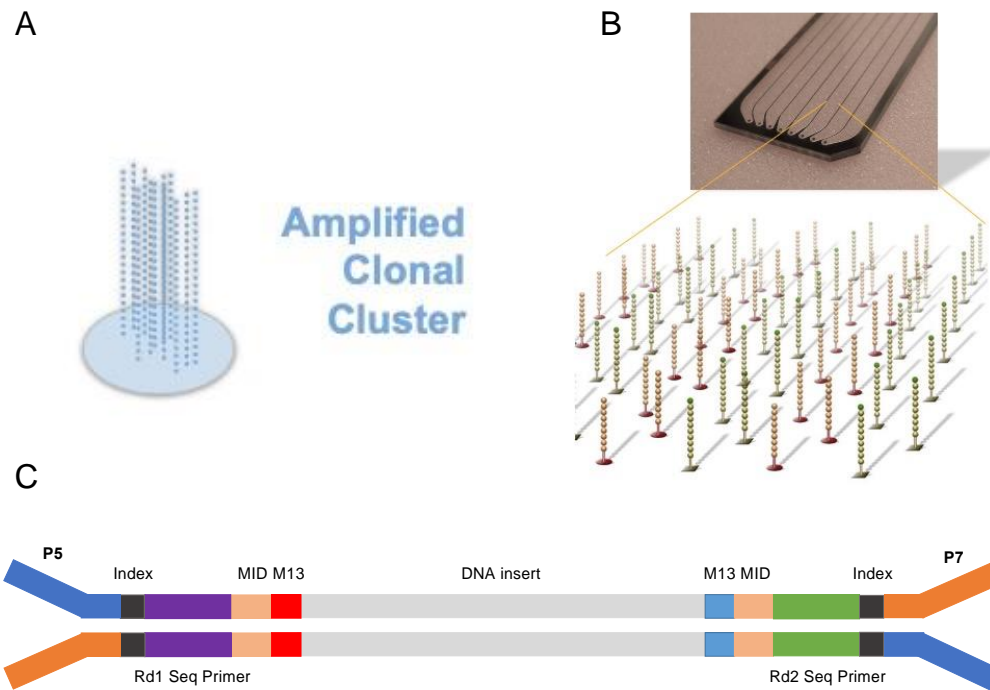


Figure 27. Cluster, flowcell and adapters. A) A cluster is a group of clonally amplified sequences. Each cluster represents thousands of copies of the same DNA strand in a 1-2-micron spot. B) The flowcell is a glass device with channels provided of a lawn of oligonucleotides (in green and red). C) Representation of the library amplicon. This figure shows the DNA insert (our amplified DNA sample) together with its sample specific adapters M13 and MID. The MID enables the recognition of each specific sample. The Rd1 Seq primer and Rd2 Seq portions are the binding site for the primer together with the index that is separately sequenced, and it helps to recognize the different pool of amplicons in a library, thus allowing sample multiplexing. Finally, the Miseq specific adapters P5 and P7 are complementary to the oligonucleotide sequences attached on the flowcell. Images obtained and modified from Illumina website [282].

To check the sequencing quality, a commercial DNA sequence (PhiX) is added and analyzed during sequencing. The PhiX is a DNA fragment at a concentration of 10 nM provided of an average size of 500 bp and a balanced base composition (about 45% GC and 55% AT) and derived from a well-characterized bacteriophage small genome. Moreover, it is also used, in sequencing, to artificially increase sequence diversity, which is important for an optimal run performance, cluster generation, sequencing, alignment, and high-quality data generation.

Once attached to the flowcell, each fragment (or amplicon) is isothermally amplified thus forming the cluster through a process known as bridge amplification. The amplicons in the library are denaturized and the single-stranded DNA (both forward and reverse) is attached on the flowcell thanks to the hybridization of the P5 or P7 sequence and the oligos on the flowcell surface (Figure 28, panel A). The

complementary strand is so produced starting from the oligo covalently bound to the flowcell resulting in double-stranded DNA molecules (Figure 28, panel B). A new cycle of denaturation occurs, and the original template strand is run away, thus leaving the new synthesized complementary strand, which is covalently attached on the flowcell surface. Thanks to the adapter, this strand forms a new bridge (Figure 19 panel C) with the following synthesis of the complementary strand up to produce another double-stranded bridge (Figure 28, panel D). After another cycle of denaturation, the results will be two covalently bound copies of the original strand (Figure 28, panel E). The other extreme of the *de novo* strands can now hybridize the other nearby flowcell's oligos thus forming several bridges (Figure 19 panel F) that will give rise to thousands of clonally amplified sequences forming each cluster (Figure 28, panel F).

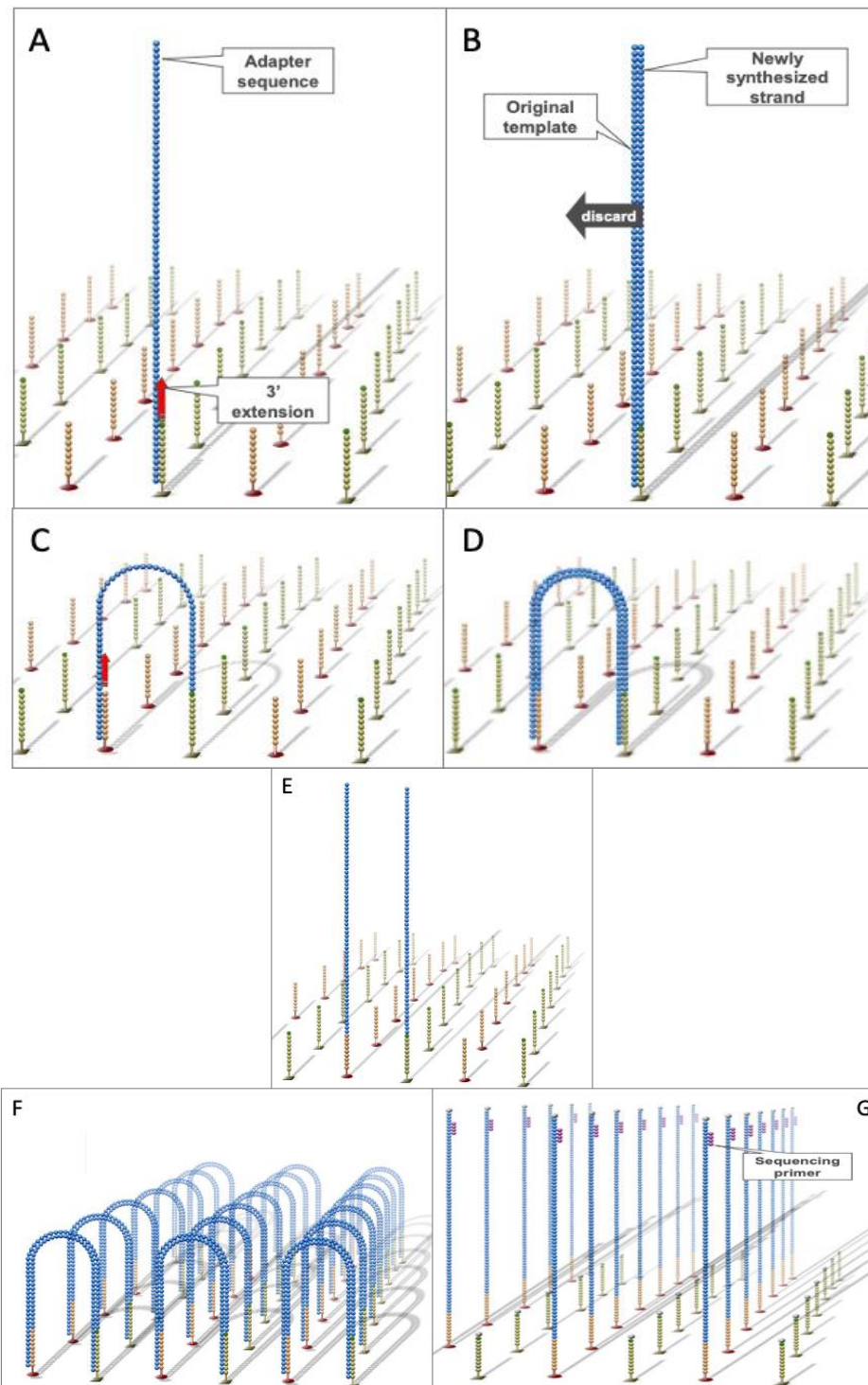


Figure 28. Overview of Illumina cluster formation. A: hybridization of the sequence with the flowcell thanks to the P5 and P7 adapters and extension of the complementary strand. B: a new cycle of denaturation occurs, and the original strand is discarded leaving just the covalently attached DNA strand. C: The free extreme end can anneal, thanks to the other adapter, a near oligo on the flowcells thus forming a bridge. D: synthesis of the complementary strand. E: denaturation of the double strands thus leaving the single stranded DNA sequence covalently attached to the flowcell F: formation of new bridges and new cycle of denaturations (G) thus resulting in several copies of the same amplicon attached to the flowcell surface, better known as clusters, that will be later sequenced thanks to the annealing with the sequencing primers. Extracted and modified from Illumina [282].

5.2. Paired end sequencing

Another essential step of the Illumina platform that strongly improves the yield of the technique is the paired-end sequencing, which consists in the sequencing of both forward and reverse strand, thing that is extremely important specially during the bioinformatics analysis, since it enables the identification of errors or indels. To guarantee an optimal sequencing efficiency, the two complementary strands are separately sequenced.

Starting from the cluster, the reverse strand is cleaved and run away, leaving the clusters with the only forward molecules. To prevent new bridge formation, the 3'-end of the forward strand is blocked. The sequencing process (read1) occurs thanks to the Illumina primers that anneals the Rd1 region in the adapter. The nucleotide added to the growing sequence will be detected during sequencing (SBS). This step will be explained with more details in the next section.

At the end of this first step of sequencing, the sequenced strand is stripped off and the 3'-extreme of the templates unblocked. This enables the formation, again, of bridges as previously explained, thus leaving two adjacent covalently attached forward and reverse strands. The original forward strand is so cleaved, and the reverse is sequenced by using the sequencing primers annealing the Rd2 sequence (read 2).

A set of sequences (reads) will be finally obtained from both forward and reverse strands (read 1 and read 2). These reads will be bioinformatically filtered (Chapter 4, section 1.4.1) and collapsed to obtain the haplotypes (genomic variant) sequences.

5.3. Sequencing by synthesis (SBS)

Another element that contributes to the performance of the Illumina platform is that the added nucleotide is recognized in real-time is a process known as sequencing by synthesis (SBS). Each of the four nucleotides is bound to a different fluorescent dye

(Figure 29) as a reversely terminator label. When the dyed nucleotide is incorporated into the nascent sequencing strand, the attached fluorochrome is excited and release from the nucleotide emitting a light of a unique wavelength that will be captured by the camera in the device and recorded. Thanks to that, at each cycle each cluster will be appeared like a colored spot. This process is repeated until the end of the sequencing. This procedure allows to read all the clusters from different amplicons and samples at the same time, enabling a base-by-base sequencing and limiting errors due to homopolymeric regions. The fragments analyzed by Illumina can be long up to 2x300 nt, although shorter lengths are recommended.

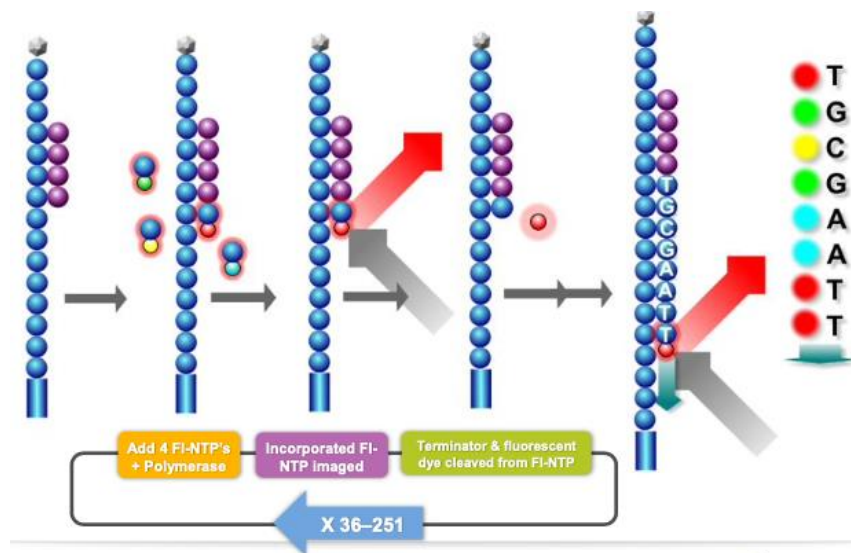


Figure 29. Detailed scheme of Sequencing by synthesis (SBS). Each nucleotide is excited by a fluorescence lamp and emit a light specific for each nucleotide and recorded.

To summarize, the clustering step permits the obtention of thousands and thousands of identical sequences, reaching an optimal coverage even for those genomic variants that display very low relative frequency. The paired-end sequencing, on the other end, contributes to the coverage and provides a double-check for the detection of the potential introduced errors. Finally, the sequencing by synthesis allows the contemporary detection of the introduced nucleotide in each nucleotide, thus sequencing thousands and thousands of strands at the same time and limiting the introduction of errors due to the existence of homopolymeric regions.

Thanks to all the above-mentioned processes, the Illumina platform is an extremely valuable instrument for studying quasispecies of highly variable viruses such as HBV and HDV.

This thesis is especially focused on using this technology to inspect the variability of HDV and the evolution of its quasispecies and to investigate if the introduction of mutations in the genome of HBV might be a contributing factor of the inhibition of the replication of HBV observed in presence of HDV.

Hypothesis and objectives

Study 1. Inspecting HDV ribozyme conservation and comparison of quasispecies variability and evolution between ribozyme and 5'-HDAG

Hypothesis

HDV infection affects more than 5% of the HBV-infected patients, and strongly worsens liver disease progression. The most widely used therapeutic strategy against HDV is based on pegylated IFN- α , which does not guarantee a sustained suppression of the viral replication. New therapeutic strategies are now under study, including silencing strategies based on silencing RNA or antisense oligonucleotide targeting the HDV helper virus, the HBV.

Like plant viroid, its genomic RNA molecule is characterized by a catalytically active RNA portion called ribozyme, which plays a key role in the viral life cycle since it is responsible for the liberation of genomic and antigenomic monomers from the concatemers during the replication. Of note, HDV is also provided of a genomic region encoding for the unique viral protein, the delta antigen (HDAG), which is involved in both virion formation and release, and in viral replication.

Significantly, HDV is the most variable among the hepatitis viruses, reaching a substitution rate of 1.2×10^{-3} to 9.5×10^{-3} nt substitutions/site/year. As proof of this variability, 8 viral genotypes have been described with an inter- and intra-genotype divergence of 16% and 20-40%, respectively. Thanks to this extremely high variability, HDV circulates as a complex mixture of closely related variants known as quasispecies (QS). The next-generation sequencing (NGS) is a highly sensitive technique that allows a deep analysis of the viral QS complexity, variability, evolution, and conservation.

Based on these assumptions, in the first part of this thesis, the following hypothesis were made:

- The high genomic variability of HDV could strongly hinder the design of silencing molecules directly targeting HDV mRNA rather than HBV.
- Considering the functional role of the ribozyme in viral replication, this portion of the HDV genome might be highly conserved and could be a valuable target of a gene therapy strategy.
- Related to the previous point, the presence of mutations in this region should potentially impact HDV replication, thus limiting it or promoting viral persistence.
- Based on the previous hypothesis, the ribozyme variability should be different from the QS variability observed in the *HDAG* sequence.

Objectives

Considering the above-mentioned hypothesis, the following objectives were set:

1. To study by NGS the conservation of the HDV ribozyme sequence to identify hyper-conserved regions that may serve as targets for gene therapy.
2. To detect mutations along the ribozyme, and to study their potential effect on viral replication.
3. To compare ribozyme QS to that of the *HDAG* (in the 5'-extreme) to inspect differences in terms of QS complexity, variability, and evolution between the two regions.

Study 2. *In vivo* study of the involvement of HDV in HBV variability: the role of type I interferon pathway

Hypothesis

HDV is a satellite virus of HBV. Of note, in presence of the delta virus, the replication of the HBV is strongly inhibited. The mechanism behind this inhibition is still completely unknown, however, studies suggest that the activation of the IFN pathway might play a role. The HBV, indeed, is considered as a “stealth” virus, since it can limitedly activate the intracellular interferon response, differently from the HDV, that even takes advantage of this pathway to produce the two isoforms of the delta antigen (by editing through the ADAR1 the Amber/W stop codon). The recognition of HDV by the intracellular PRRs induces the expression of different ISGs including those encoding for the mutagenic enzymes APOBEC/AID and ADAR1. Although HBV is a DNA virus, it shows a degree of variability like RNA viruses. It replicates through a retrotranscription step, which is considered the main source of mutations. However, mutations associated with APOBEC have been also reported. An *in vivo* model of the HDV superinfection using HBV transgenic mice that constitutively express high levels of HBV, knock-out and wild-type for the type I IFN pathway, might be extremely helpful to study the interaction between both viruses and the role of the HDV-induced IFN in viruses’ variability.

Starting with these promises, the following assumptions were made:

- The activation of the type I IFN pathway due to the presence of HDV may have a direct role in inhibiting HBV replication.
- The activation of the type I IFN pathway promotes the expression of different ISGs including those that encode for the APOBEC/AID and ADAR1 deaminases, which could be hyper-mutagenic in the HBV genome thus introducing lethal mutations that affect viral replication.
- Moreover, the strong activation of this immune pathway might also contribute to HDV variability, since this virus is characterized by a double-stranded RNA molecule that could be edited by ADAR1 in more sites rather than the unique Amber/W codon.

Objectives

In the second part of this thesis, the following objectives were planned:

1. To evaluate the inhibition of HBV expression in presence of HDV, and its relationship with the interferon type I pathway, using HBV transgenic mice.
2. To study if the activation of the type I IFN due to the presence of HDV might promote HBV variability by analyzing the intrahepatic HBV RNA (before the retrotranscription step) through NGS in two regions (the SRT and HBX) that are highly important in the viral replication. To reach this objective, the viral RNA QS is analyzed at two different timepoints (7 and 21 days post-AAV injections).
3. To inspect the role of type I IFN in HDV RNA variability in the region of the 5'-HDAG to check if the activation of ADAR1 might contribute to HDV variability.

Chapter 3

Materials

1. Primers and probes

The name, and position of the primers used in this thesis are listed in the following tables. Of note, the assay where the primers were used is also reported.

Table 2. List of the used primers for qPCR. The table shows the list of all the primers used to produce the results explained in this thesis, including the target virus, positions in viral genome and sequence. LC610 and BBQ: Fluorochrome and Quencher of TaqMan probe for qPCR. LC640, fluorescein and phosphate: Fluorochrome and Quenchers of FRET probes for qPCR.

Name	Virus	Positions	Sequence 5' → 3'
Full X Fw	HBV	1263-1283	GATCCATACTGCGGAACTCCT
Full X Rv	HBV	1408-1388	GHAGGATCCAGTTGGCAGYAC
Probe Full X	HBV	1289-1313	LC610-CTTGTTTTGCTCGCAGCMGGTCTGG--BBQ
pgRNA Fw	HBV	2407-2429	GTCCCCTAGAAGAAGAACTCC
pgRNA Rv	HBV	2472-2495	CATTGAGATTCCCGAGATTGAGAT
Probe pgRNA	HBV	2449-2469	LC610-TCTCAATCGCCGCTCGCAGA-BBQ
Primer DP15 Fw	HDV	815-836	CGTCCTTCTTTCCTCTTCGGGT
Primer DP2 Rv	HDV	900-919	AGTGGCTCTCCCTTAGCCAT
Probe HDV-FL	HDV	848-870	TCCTCCTTCGGATGCCAGGTCTG - fluorescein
Probe HDV-LC	HDV	873-895	LC640 - CCGCGAGGAGGTGGAGATGCCAT - phosphate

Table 3. List of the used primers for NGS. The table shows the list of all the primers used to produce the results explained in this thesis, including the target virus, positions in viral genome, sequence, and the procedure. Underlined M13 tail. MID: Multiplex identifier.

Name	Virus	Positions	Sequence 5' → 3'	Application
Oligo(dt) ₁₇	HBV	-	TTTTTTTTTTTTTTTTTTT	HBV reverse transcription
M13 SRT FW	HBV	596-614	<u>GTTGTA</u> AAACGACGGCCAGTACCTGTATT CCCATCCCAT	SRT amplification
M13 SRT RV	HBV	970-990	<u>CACAGGAA</u> ACAGCTATGACCACATACTTT CCAATCAATAGG	SRT amplification
EXT X FW	HBV	579-599	TGTATTCCCATCCCATCATC	HBX amplification
EXT X RV	HBV	1912-1936	AGWAGCTCCAAATTCCTTTATAAGG	HBX amplification

M13 5HBX FW	HBV	1234-1255	GTTGTA AAAACGACGGCCAGTATGCGTGGA ACCTTTGTGGCT	5'HBX amplification
M13 5HBX RV	HBV	1631-1611	CACAGGAAACAGCTATGACCATGGGCGTT CACGGTGGTCT	5'HBX amplification
M13 3HBX FW	HBV	1581-1600	GTTGTA AAAACGACGGCCAGTGCACTTCGC TTCACCTCTG	3'HBX amplification
M13 3HBX RV	HBV	1936-1912	CACAGGAAACAGCTATGACCAGWAGCTCC AAATTCTTTATAAGG	3'HBX amplification
Ribo RT	HDV	1435-1454	TGGCTGGGAAACATCAAAGG	Ribozyme amplification
Ext Ribo FW	HDV	1435-1454	TGGCTGGGAAACATCAAAGG	Ribozyme amplification
Ext Ribo RV	HDV	308-326	CCTCCAGAGGACCCCTTCA	Ribozyme amplification
M13 Ribo FW	HDV	883-900	<u>CACAGGAAACAGCTATGACCTCGGCATGG</u> CATCTCCAC	Ribozyme amplification
M13 Ribo RV	HDV	663-683	<u>GTTGTA AAAACGACGGCCAGTCGCGTTCCA</u> TCCTTTCTTACC	Ribozyme amplification
HDAg RT	HDV	728-747	CGGTCCCCTCGGAATGTTG	5'HDAg amplification
Ext HDAg FW	HDV	865-884	AGGTCCGACCGGAGGAGGT	5'HDAg amplification
Ext HDAg RV	HDV	306-328	GCTGAAGGGTCTCTGGAGGTG	5'HDAg amplification
M13 HDAg FW	HDV	886-909	<u>GTTGTA AAAACGACGGCCAGTGAGATGCCA</u> TGCCGACCCGAAGAG	5'HDAg amplification
M13 HDAg RV	HDV	1272-1295	<u>CACAGGAAACAGCTATGACCCGACGAAGG</u> AAGGCCCTCGAGAAC	5'HDAg amplification
MID FW	-	-	MID-GTTGTA AAAACGACGGCCAGT	HBV and HDV amplification
MID RV	-	-	MID-CACAGGAAACAGCTATGACC	HBV and HDV amplification

Table 4. List of the used primers for Ribozyme Mutagenesis. The table shows the list of all the primers used to produce the results explained in this thesis, including the target virus, positions in viral genome, sequence and the assay or procedure. Underlined the place where the mutation was inserted.

Name	Positions	Sequence 5' → 3'	Application
T23C FW	1596-1619	CGGCGCCAGCG <u>GG</u> GAGGCTGGGA	Mutagenesis
T23C RV	1596-1619	TCCCAGCCTCC <u>CG</u> CTGGCGCCG	Mutagenesis
C64del FW	1547-1577	TCCCATTGCCATTACCGAG <u>GG</u> ACGGTCC	Mutagenesis
C64del RV	1547-1577	GGACCGTCC <u>C</u> TCGGTAATGGCGAATGGGA	Mutagenesis
T69C FW	1544-1571	GGGTCCCATTGCCATT <u>G</u> CCGAGGGGA	Mutagenesis
T69C RV	1544-1571	TCCCCTCGG <u>C</u> AATGGCGAATGGGACCC	Mutagenesis
Seq Ribo FW	1418-1436	TCGGCATGGCATCTCCAC	SANGER
Seq Ribo RV	1803-1823	ACTTATCGTCCCCATCTAGC	SANGER

2. Commercial kits

Different commercial kits were used in this project. A list of all of them, with their application, is provided below.

2.1. Serological Markers

- Elecsys® HBsAg II kit and Elecsys® HBeAg kit (Roche Diagnostics, Rotkreuz, Suiza): This kit was used for the determination of HBV serological markers HBsAg and HBeAg, respectively, on the COBAS 8000 automated system. This determination is performed using a commercial electrochemiluminescence

immunoassay from 150 μL of serum extracted from EDTA tube. From stable substrates, products capable of emitting photons of electrochemical origin are generated, the antibody used coats magnetized microparticles, which, when the antigen-antibody complex is formed, are attached to an electrode by magnetism. This antibody is conjugated to a marker capable of emitting photons when a small potential difference is applied to the electrode. This light energy is detected in a photomultiplier. Lipemic, icteric or haemolyzed samples do not affect this quantification.

- HBV RNA investigational Assay (Roche Diagnostics, Rotkreuz, Switzerland): This kit allows the determination of circulating HBV RNA in 200 μL of serum by qPCR on COBAS 6800 automated system. This kit has excellent precision and linear detection from 10 copies/mL to 10^9 copies/mL. This kit has a good accuracy, it has standard deviation below 0.15 log₁₀ copies/mL.
- Cobas® HBV Test (Roche Diagnostics, Mannheim, Germany): this kit quantifies HBV DNA by quantitative PCR (qPCR) on the COBAS 6800 system. This system uses a universal sample and reagent preparation procedure to isolate, purify and extract total sample nucleic acids from 650 μL of plasma extracted in EDTA tube. Once extracted, qPCR simultaneously amplifies and quantifies the amplification product absolutely. This technique has a linearity range from 10 to $1.00\text{E}+09$ IU/mL and a detection limit of 2.7 IU/mL. Lipemic, icteric or haemolyzed samples do not affect this quantification.

2.2. DNA and RNA extraction

- High Pure Viral Nucleic Acid Kit (Roche Diagnostics, Mannheim, Germany): Enables the purification of viral nucleic acids from 100 μL mammalian serum, plasma, or whole blood. This kit allows viral lysis thanks to the incubation of the sample with a special lysis buffer/binding buffer containing proteinase K, which is a broad-spectrum serine protease that allows the cleavage of proteins for the release of nucleic acids. These nucleic acids bind to the

surface of the glass fibers in the presence of a chaotropic salt, and are washed to remove PCR inhibitors, salts, proteins, or cellular impurities and finally eluted with elution buffer.

- MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics, Mannheim, Germany): automatized purification of total nucleic acid (viral DNA and RNA) from mammalian serum, plasma, and whole blood. From a minimum volume of 250 μ L of plasma extracted in EDTA tube, lysis of the viral particles is carried out by incubation at high temperature with a proteinase K and a chaotropic binding/lysis buffer. This releases the nucleic acids and protects the released RNA from ribonucleases (RNAses) present in the plasma. Magnetic glass particles (MGP) are then added, and the nucleic acids present in the sample are bound to their surface. Unbound substances, such as salts, proteins, and other impurities, are removed when washing the MGPs. Finally, the adsorbed nucleic acids are eluted at high temperature with a low-salt buffer.
- QIAamp Viral RNA mini kit protocol (QIAGEN, Hilden, Germany): purification of viral RNA from cell-free body fluid. The presence of optimized buffers and enzymes lyses the samples, stabilizes the nucleic acids, avoiding phenol-chloroform for extraction, and enhancing the selective adsorption of RNA to the QIAamp silica membrane, while contaminants pass through. To ensure RNA integrity, samples are lysed under highly denaturing conditions to inactivate RNAses. Alcohol is added and the lysates are loaded onto the QIAamp spin column. Wash buffers are used to remove impurities and PCR inhibitors, such as divalent cations and proteins, which are completely removed in two efficient washing steps and then pure, ready-to-use RNA is eluted in low-salinity buffer or water.

2.3. RNA and DNA quantification

- GoTaq® qPCR Master Mix (Promega, Wisconsin, USA): This qPCR kit includes a fluorescent DNA-binding dye, BRYT Green® Dye, which is characterized by increased fluorescence upon binding to double-stranded DNA (dsDNA).
- Qubit RNA HS Assay Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, USA): For the rapid detection and quantification of the RNA, based upon the use of a fluorogenic dye that selective bind RNA molecules, even at low concentrations.
- LightCycler RNA Master HybProbe (Roche Diagnostics, Mannheim, Germany): Easy-to-use Reaction Mix for One-Step RT-PCR using the LightCycler® Carousel-Based System. It contains a Tth DNA Polymerase, which is a thermostable enzyme with RNA-dependent reverse transcriptase activity and DNA-dependent polymerase activity, together with aptamers that are oligonucleotides that bind the RNA sample preventing the attachment of primers to the target at temperatures below the optimal reaction temperature of the Tth enzyme. This principle is used to quantify RNA samples using FRET technology.
- LightCycler® FastStart DNA Master HybProbe (Roche Diagnostics, Mannheim, Germany): Easy-to-use reaction mix for the qPCR of DNA molecules using the LightCycler® Carousel-Based System. This kit allows specific and accurate detection of both DNA and cDNA. This reaction employs the enzyme FastStart Taq DNA Polymerase, which is active at high temperatures, where the primers do not establish non-specific binding. This principle quantifies DNA and cDNA samples through FRET technology.
- Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, Camarillo, CA, USA): enables the quantification of double stranded DNA through a fluorophore that reacts with the target even at low concentrations (0,25 pg/ml).

2.4. DNA and RNA amplification and transcription

- MEGAscript T7 Transcription Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, USA): kit provides of a T7 RNA polymerase that allows the in vitro transcription of DNA sequence under T7 promoter control.
- Accuscript HiFi reverse transcriptase PCR kit: The Accuscript HiFi enzyme enables the retro-transcription of long RNA (0.1–9.6 kb in length) with limited amount of starting RNA level (from 10–1000 ng of total RNA). More importantly, this enzyme shows a high specificity and fidelity, which is an essential factor to avoid the introduction of mutations that could alter NGS results.
- AccuScript High Fidelity RT-PCR System (Agilent Technologies, Waldbronn, Germany): This kit features a very accurate retrotranscriptase. This RT-PCR System allows the amplification of cDNA from 0.1 to 9.6 kb in length from 10 to 1000 ng of total starting RNA.
- PfuUltra II Fusion HD DNA polymerase (Agilent Technologies, Waldbronn, Germany): This standard enzyme for high-fidelity PCR, has a lower error rate than other polymerases (1 error/ 2.5 million bp). Can amplify up to 19 Kb of genomic DNA targets and amplifies up to 70-80% faster than other polymerases.
- FastStart Taq DNA Polymerase dNTPack (Roche Diagnostics, Mannheim, Germany): This enzyme is a combination of FastStart Taq DNA Polymerase and ready-to-use PCR grade nucleotides. This enzyme is thermostable and rapidly activates at 95°C but is inactive below 75°C to prevent the extension of non-specific fragments. It is versatile and can be used in a wide range of applications and instruments.

- DNase I (Thermo Fisher Scientific, Camarillo, CA, USA): This deoxyribonuclease I digests single and double-stranded DNA into oligodeoxyribonucleotides presenting 5' phosphate. This DNase is effective in amounts of DNA between 10,000 to 25,000 units/mg.
- TURBO DNase (Thermo Fisher Scientific, Camarillo, CA, USA): the method of removing DNA is the same as previously mentioned for DNase I, however, Turbo DNase is genetically modified with respect to DNase I, improving DNA affinity and presenting a 50-fold higher catalytic activity and 350% higher efficiency than DNase I.

2.5. Nucleic acid purification

- MEGAclean Transcription Clean-up Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, USA): kit for the purification of products obtained by in vitro transcription using the MEGAscript T7 Transcription Kit.
- QIAquick Gel Extraction Kit protocol (QIAGEN®, Hilden, Germany): for the purification of an amplicon with a specific length after separation through electrophoresis in 1.5% agarose gel. This kit removes any remaining impurities from the samples (nucleotides, salts, enzymes, agarose, between others) and ensures a recovery of 80% of the DNA. The gel band is dissolved in a buffer with an optimum pH indicator for DNA binding. This mixture is applied to a QIAquick column, and the nucleic acids are adsorbed on the silica membrane. The columns are washed with 80% ethanol to eliminate impurities and the DNA is finally eluted in 60 μ L of EB buffer (10 mM Tris-Cl, pH 8.5).
- D1000 ScreenTape kit (Agilent Technologies, Waldbronn, Germany): Automated method for the analysis of the quality of RNA and DNA samples. It is an automated electrophoresis solution that allows to analyze the size, quantity, and integrity of our samples in a fast (1-2 minutes/sample) and

efficient way. It is recommended to guarantee the quality and purity of samples for NGS processing.

- Exosap IT (Thermo Fisher Scientific-Life Technologies, Austin, TX, USA): Exosap IT is an enzymatic clean-up product for the amplified PCR product. It hydrolyses excess nucleotides and primers in a single step, thus minimizing sample loss in the clean-up process.
- BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, USA). A set of reagents suitable for fluorescence-based cyclic sequencing with single or double-stranded DNA (Sanger). This kit offers high throughput for a wide variety of DNA sequences, maximizing read lengths and allowing the identification of variants up to 5% in the sample after sequencing.
- BigDye® Xterminator™ Purification Kit (Thermo Fisher Scientific-Life Technologies, Austin, TX, USA): Allows to purify the sequences obtained from amplification with BigDye terminators. It purifies quickly and fluently the samples by eliminating unincorporated BigDye terminators and salts.
- Kapa Pure beads (Roche Diagnostics, Rotkreuz, Switzerland): positively charged paramagnetic beads in an optimised buffer containing Polyethylene glycol (PEG)/NaCl that bind the DNA target through the electrostatic interaction with the negative charge of the phosphate group in deoxyribonucleic acid. The proportion of the beads used with this system is a determinant element for the correct purification of the desired amplicon. Depending on the length of the PCR product, a different proportion of beads must be used. In this study the magnetic spheres were used at 0.8X related to the samples in both studies. In manual purification, magnetic beads with amplified DNA are incubated for 10 minutes at 22°C for DNA attachment to beads and subsequently there are incubated on a magnet, the DNA is captured together with the magnetic beads on the magnet and the shortest fragments are removed by washing twice with 80% ethanol. For the elution

of the cleaned DNA samples, they are incubated with elution buffer (EB) at 37° for 5 minutes for the dissociation of the DNA from the beads, these beads are retained thanks to the magnet and the supernatant containing our purified genetic material is recovered.

2.6. NGS library preparation

- Kapa Hyper Prep Kit (Roche Sequencing, Pleasanton, USA): This kit provides protocols for several steps of library preparation. It combines several enzymatic steps and minimizes bead cleanups to reduce sample handling time. This kit presents the enzymes and buffers needed for: end repair and A-tailing, adapter ligation and library amplification (explained later).
- SeqCap Adapter Kit (Roche Sequencing, Pleasanton, USA): which contains the set of adapters that must be used in the library ligation step.
- KAPA HiFi HotStart ReadyMix PCR Kit (Roche Sequencing, Pleasanton, USA): This kit is used for the amplification step of NGS libraries, the kit contains the KAPA HiFi HotStart DNA polymerase, which has been developed for fast and versatile high-fidelity PCR.
- MiSeq Reagent Kit V2 and V3 (Illumina, San Diego, USA): kit with the pre-filled, ready-to-use reagent cartridges for MiSeq sequencing with a read length of respectively 2x250 and 2x300bp. This kit allows reaching 15 (v2) and 25 (v3) millions of reads per run.
- KAPA Library Quantification Kit (Roche Sequencing, Pleasanton, USA): This kit contains all the reagents needed for the accurate, reliable, and reproducible qPCR-based quantification of NGS libraries prepared for sequencing on Illumina platforms.

2.7. Mutagenesis, plasmid purification and transfection

- QuickChange Lightning site-directed mutagenesis kit (Agilent Technologies, Waldbronn, Germany): It offers a fast (3 hours) single and multiple sites directed mutagenesis protocol, including the possibility of introducing insertions or deletions. This kit uses a linear amplification method where primers are incorporated but not copied. Suitable for lengths from 4 to 14 kb with an efficiency of more than 80% for single directed mutagenesis and more than 55% to mutagenesis efficiency for multiple directed mutagenesis. The kit allows the mutated plasmid purification and cloning into XL-gold bacteria.
- Endotoxin free NucleoBond Xtra Midi plus (Macherey-Nagel, Duren, Deutschland): this kit is designed for the purification of plasmids, starting from a bacterial inoculum with the cloned plasmid. Passing through steps of plasmid extraction from the bacteria, filtering and washing, we obtain the plasmid DNA ready for further use. It features anion-exchange technology for faster and more efficient filtration, as well as columns in the kit for better filtration. In addition, this kit is endotoxin-free, thanks to a washing step that removes endotoxins from our material. The final step with a finalizer included in the kit, allows a better and faster filtration of plasmid DNA.
- Magnetofectamine O2 (OZBiosciences, Marseille, France): kit designed for transfection of primary or difficult to transfect cells. This kit allows a fast and efficient transfection, minimizes cell toxicity and enhances gene expression thanks to the transfection system using a magnetic field. The transfection method of this kit is performed in several steps, first binding the transfected DNA with a reagent presenting magnetic complexes and subsequently incubation of this beads-DNA complex, later they are added to the culture plate above a magnetic plate for 20-30 minutes allowing the magnetic field complexes entering the cells.

3. Cell lines

- HUH7: Hepatocarcinoma cells isolated in 1982 from a 57-year-old male. They are well-differentiated hepatocytes with a high replicative capacity. They are adherent cells and grow in 2D monolayers. They were originally used for HCV and hepatoma research. Currently it is an *in vitro* model used for HBV infections and HDV transfections. The culture medium used is Dulbecco's Modified Eagle Medium (DMEM) supplemented with Glutamax (L-alanyl L-glutamine in 0.85% of NaCL) and antibiotic (penicillin and streptomycin), with 10% fetal bovine serum (FBS).

Study 1

Inspecting HDV ribozyme conservation and comparison of quasispecies variability and evolution between ribozyme and 5'-HDAG

1. Methods

1.1. Patients and samples

One of the aims of this thesis was to study HDV quasispecies and its evolution in two genomic regions that cover essential roles in viral replication: the ribozyme and the delta antigen. For this purpose, HDV RNA was extracted from plasma samples of patients suffering chronic hepatitis delta (CHD) and the viral genome was analysed by next generation sequencing as detailed in the following paragraphs.

Patients were selected from those attending the outpatient clinics at Vall d'Hebron University Hospital (Barcelona, Spain). All were diagnosed with CHD and presented at least 3log IU/mL of HDV RNA. Moreover, none of them were treated with IFN at the time of samples collection. The treatment with NUC against HBV was not considered as exclusion factor, since these drugs, not acting on HDV, shouldn't influence the HDV QS. To limit factors that could have an impact of viral QS evolution, patients with HIV infection or autoimmune diseases were excluded.

As previously commented (Chapter 1, section 4.2), HDV is a highly variable virus. To evaluate how viral QS can change during the time, two plasma samples were collected per each patient: one sample at the beginning of the study and the other after at least 5 months.

Viral and biochemical markers were evaluated in both samples. HBV virological markers (HBsAg and HBeAg) were tested in serum by chemiluminescence assays with the COBAS 8000 instrument (Chapter 3, section 2.1). The HBV DNA was quantified qPCR through the COBAS 6800 instrument (Chapter 3, section 2.1). Since samples were surplus of the samples collected during the normal clinical follow-up of the patients, the data about ALT and AST levels were extrapolated from the follow-up lab report.

The titter of HDV RNA was quantified through an in-house protocol and by using an international viral RNA standard (1st World Health Organization International

Standard for Hepatitis D Virus RNA for Nucleic Acid Amplification Techniques-based assays. PEI code number: 7657/12) [283]. The protocol used for the quantification is detailed in the following paragraphs.

Data regarding HDV genotypes was also collected. Viral genotypes were previously studied in Dr. Sara Sopena's doctoral thesis. Briefly, a region coding for the *HDAG* (between nt 910-1270 of HDV genome) were analysed by NGS and haplotypes were aligned in reference to each genotype sequence (from genotype 1 to 8). The genotypes were determined using the Kimura-80 model and a dendrogram was constructed using the unweighted pair group method with arithmetic mean (UPGMA) [284].

1.2. HDV RNA extraction

Depending on the volume of sample available, the HDV RNA was extracted by either automatic or manual kits. For volumes larger than 250 μ L, the viral RNA was isolated by using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Chapter 3, section 2.2) (with the Automated sample preparation system MagNA Pure 24 [285]. In the other side, samples with less than 250 μ L were manually treated with the QIAamp Viral RNA mini kit (Chapter 3, section 2.2) following the manufacturer's instructions.

Of note, the WHO international standard used for the viral RNA quantification is a lyophilized preparation of HDV genotype 1 strain obtained from a positive human plasma further diluted with human negative plasma [144]. It was estimated that the standard has a potency of 5.75E+05 International units per mL (IU/mL) when reconstituted in 0.5 mL of nuclease-free water, and both its potency and stability were tested by an international collaborative study involving 15 laboratories performing a wide range of HDV real-time Nucleic Acids Amplification technology (NAT) assays [286].

1.2.1. HDV RNA: WHO Standard Curve and Viral Load Estimation

Since a standardized and automatized technique for the quantification of HDV RNA is still not available, viral genome was quantified by an in-house one-step quantitative RT-PCR technique using the WHO international standard.

To try to overcome the limitation related to the lack of standardization for HDV RNA quantification, the WHO international standard was serially diluted to obtain a curve for the viral titer quantification. Once extracted, the standard HDV RNA was eluted in 60 μ L of AVE elution buffer (provided by the kit) which contains 0.04% sodium azide diluted into RNase free water to prevent microbial growth and undesired RNA degradation, and it was 10-fold serially diluted to obtain a curve within a range of $5.75E+05 - 5.75E+01$ IU/mL [283].

Since the HDV RNA acquires a rod-like structure due to the high internal rate of base pairing (about 70%) [159,287], and forms a ribonucleoprotein complex with the HDAg (Chapter 1, section 2.1), a denaturation step is required to open the RNP structure and to allow primers annealing. This step consisted of a heat-shock at 95°C for 5 minutes to break the linkage between the complementary strands, followed by 5 minutes at -80°C to prevent possible renaturation. The HDV RNA quantification was performed by using the Florescence Resonance Energy Transfer (FRET) system with the primers and probes reported by Schaper *et al.* [228] (Probes HDV-FL and HDV-LC and primers DP15 and DP2 in (Chapter 3, section 1, table 2). The FRET system consists of two oligonucleotide probes provided of two different dyes which wavelength overlap. When the probes hybridize the target, the dye at 3' (donor) in the first probe is excited thus liberating an emission spectrum that overlaps with the excitation spectrum of the dye (acceptor) at the 5' of the second probe. This quantification of viral RNA in a one-step RT-qPCR was performed by using the LightCycler RNA Master HybProbe kit on the LightCycler 2.0 system (Chapter 3, section 2.3) The above-mentioned steps (denaturation and qPCR with FRET probes) were applied to both the standard and sample's RNA molecules. The protocol applied is exposed in table 5.

Table 5. Protocol used for HDV RNA quantification.

	Target (°C)	Hold (Time)	Acquisition Mode	Analysis Mode
Reverse transcription	61	20 min	None	None
Desnaturalization	95	10 min	None	Quantification
Amplification (45 cycles)	95	2 sec	None	Quantification
	55	15 sec	Single	
	72	15 sec	None	
Cooling	40	30 sec	None	None

To finally quantify HDV RNA, the number of cycles at the intersection between the amplification curve and the threshold (named as crossing point -Cp- or cycle threshold -Ct) per each dilution were correlated to the specific standard's concentration to obtain a regression line with the correspondent R-squared (R^2). This value represents the goodness-of-fit of the regression curve, and just those with a $R^2 \geq 0.99$ were considered valid for quantification.

1.3. Preparation of the library for the analysis by NGS of the ribozyme and 5'-HDAG

1.3.1. HDV ribozyme amplification

As previously commented in Introduction and Objective chapters, the ribozyme is an essential element of HDV replication as it is responsible of the release of genomes and antigenome monomers during viral rolling circle amplification (Chapter 1, section 2.2) This part of my thesis was focused on studying ribozyme variability and conservation (Chapter 2, "Hypothesis and Objectives").

The amplicon used in this study and including the ribozyme encompassed the positions 683 to 899, (ribozyme 688 and 771) on the HDV genome. The HDV RNA was obtained from the serum samples as reported above in section 1.2.1 of the present chapter. After extraction, it was subjected to a heat shock step as previously reported (section 1.2.1), and then retro-transcribed by using a primer reverse

complementary to the positions 1434-1453 of HDV genome (Ribo RT. Chapter 3, section 1, table 3) with the AccuScript high-fidelity reverse transcriptase (RT) (Chapter 3, section 2.4). Three nested-PCRs were performed, as shown in figure 30, with a sensitivity limit greater than 1.00E+02 IU/mL. Although the limit of PCR sensitivity was 1log below the viremia criteria used at inclusion, we selected samples with at least 3log IU/mL to guarantee enough number of reads after sequencing. The first PCR step amplified a larger region (between nt 326 to 1453 HDV genome) (Ext Ribo FW and Ext Ribo RV, Materials, 1) that included the ribozyme portion. In the second amplification, the ribozyme was amplified with the addition of M13 sequence (from nt 683 to 899 of HDV genome); M13 Ribo FW and M13 Ribo RV (Materials, 1). This sequence, as previously explained in the introduction chapter (Introduction, 5) enables the addition of the multiplex identifier (MID), the 10 nucleotide-long tail specific for each sample [270]. All the PCR steps were performed using the high fidelity Pfu Ultra II DNA polymerase (Chapter 3, section 2.4), protocols of amplification used in each step are reported in Table 6.

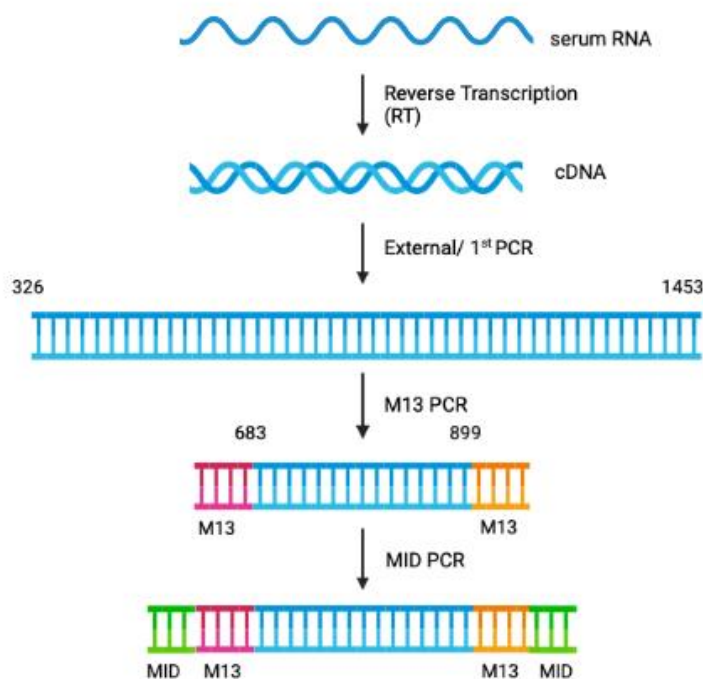


Figure 30. Graphical representation of ribozyme amplification. The HDV RNA was retrotranscribed and converted into cDNA that later underwent to 3 nested amplifications. The first or external PCR encompasses from nt 326 to 1453. The 2nd PCR allows the addition of the M13 tail and amplified from nucleotide 683 to 899, whereas the 3rd PCR added to the 2nd sample a specific multiplex identifier (MID). All the positions marked in this table correspond to HDV genomic sequence.

Table 6. Amplification protocol for HDV ribozyme amplification. MID: multiplex identifier, specific for each sample. RT: reverse transcription.

Amplification step	Primer	Protocol
RT	Ribo RT	RT 42°C 60 min; inactivation 70°C 10 min; cooling 20°C ∞
1 st PCR	Ext Ribo FW	95°C 1 min; (94°C 20 s, 54°C 20 s, 72°C 45 s) × 40 cycles; 72°C 3 min
	Ext Ribo RV	
M13 PCR	M13 Ribo FW	95°C 2 min; (94°C 20 s, 60°C 20 s, 72°C 30 s) × 35 cycles; 72°C 3 min
	M13 Ribo RV	
MID PCR	MID FW	95°C 2 min; (94°C 20 s, 60°C 20 s, 72°C 45 s) × 25 cycles; 72°C 3 min
	MID RV	

To inspect the correct amplification of the desired amplicon and ensure the absence of contamination, the quality of the PCR products were checked by gel electrophoresis on 1.5% agarose gel diluted in 1X tris-acetate-EDTA (TAE) buffer (Corning Mediatech Inc., Manassas, VA, USA), which enables the separation of the DNA molecules based on their molecular weight. Those bands corresponding to the amplicon were cut, weighed, and purified using the QIAquick Gel Extraction Kit protocol (Chapter 3, section 2.5) following the manufacturer's instructions. As commented in Materials, this kit enables the extraction of the DNA from agarose gel and its purification by applying a column composed by silica membrane that adsorbs the nucleic acids.

The quality of the purification was verified through the 2200 TapeStation System and the D1000 ScreenTape kit (Chapter 3, section 2.5). This system, together with the Agilent 2100 Bioanalyzer software, allows the visualization of the purified amplicon by electrophoresis. Moreover, thanks to the standard curve provided by the D1000 ScreenTape kit, it furnishes details about the presence and length of different DNA fragments in the sample. This step is essential because the presence of shorter contaminant fragments might reduce the sequencing yield.

1.3.2. 5'-HDAG amplification and library preparation

As previously explained in the introduction (Chapter 1, section 2.4), the *HDAG* encodes for the unique HDV-specific protein, the HDAG (small and large), which is an essential structural component of the viral virion and plays a key role in viral replication. Another open question that I tried to address in my thesis was to evaluate the existence of a difference in QS variability between the ribozyme and the 5'-*HDAG*. For this purpose, the 5'-*HDAG* was also analyzed by NGS from the same samples reported for the ribozyme.

Here, a region between nucleotides 910 to 1270 (genomic positions), corresponding to the 5'-end of the delta antigen (in genomic sense), was amplified as follows. Of note, this amplified region also contains the editing site (nt 1012 of the HDV genome, and amino acid 196 within the HDAG protein in antigenomic sense) that gives rise to the long isoform of the delta antigen (L-HDAG) (Chapter 1, section 2.4).

The preceding steps of viral RNA isolation, heat shock and reverse transcription were the same that those used for the ribozyme amplification, but the primer used for reverse transcription of 5'-HDAG encompassed from nt 728 to 747 of HDV genome (RT HDAG, Chapter 3, section 1, table 3).

At amplification, 3-nested-PCRs were performed (Figure 31). In the first PCR a bigger region (between nt 884 to 328 of HDV genome) that encompassed the HDAG region was amplified (primers: Ext HDAG FW and Ext HDAG RV, Chapter 3, section 1, table 3). In the second amplification step, the 5'-*HDAG* was amplified with the addition of the M13 tail-sequence (from nt 910 to 1295 of HDV genome; primers M13 HDAG FW and M13 HDAG RV, Chapter 3, section 1, table 3), whereas during the third amplification step the sample-specific multiplex identifier (MID) was added. All PCR steps were performed using the Pfu Ultra II high-fidelity DNA polymerase (Chapter 3, section 2.4).

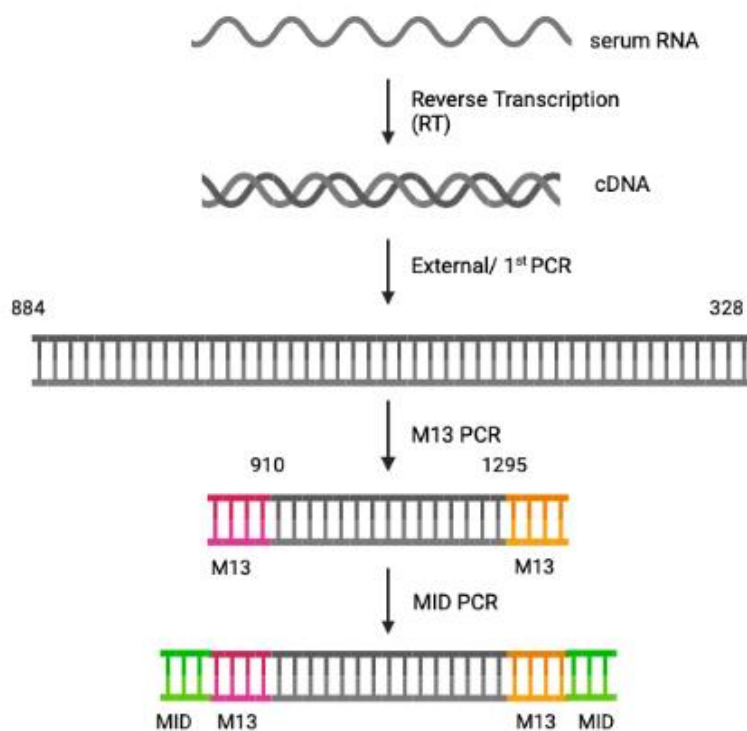


Figure 31. Graphical representation of ribozyme amplification. The HDV RNA is retrotranscribed and converted into cDNA that later underwent to 3 nested amplifications. The first or external PCR encompasses from nt 884 to 328. The 2nd PCR allows the addition of the M13 tail and amplified from nucleotide 910 to 1295, whereas the 3rd PCR added to the 2nd sample a specific multiplex identifier (MID). All the positions marked in this scheme belong to HDV genome sequence.

The detailed protocol used for 5'-HDAG amplification is reported in table 7.

Table 7. Amplification protocol for HDV 5'HDAG amplification. MID: multiplex identifier, specific for each sample. RT: retro transcription.

Amplification step	Primer	Protocol
RT	HDAg RT	RT 42°C 60 min; inactivation 70°C 10 min; cooling 20°C ∞
1 st PCR	Ext HDAg FW	95°C 1 min; (94°C 20 s, 54°C 20 s, 72°C 45s) × 40 cycles; 72°C 3 min
	Ext HDAg RV	
2 nd PCR	M13 HDAg FW	95°C 2 min; (94°C 20 s, 63°C 20 s, 72°C 30s) × 35 cycles; 72°C 3 min
	M13 HDAg RV	
MID PCR	MID fw	95°C 2 min; (94°C 20 s, 60°C 20 s, 72°C 45s) × 25 cycles; 72°C 3 min
	MID rv	

To inspect the correct amplification of the desired amplicon and ensure the absence of contamination, the quality of the PCR products were checked by gel electrophoresis on 1.5% agarose and purified using the QIAquick Gel Extraction Kit protocol (Chapter 3, section 2.5), and the quality of the purification was done as reported before for the ribozyme.

1.3.3. Quantification, normalization of the amplicons and library production

The purified amplicons were quantified with the automated system Freedom EVO® (Tecan, Mannedorf, Switzerland) coupled to the Infinite 200 Pro (Tecan, Mannedorf, Switzerland) fluorimeter using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Chapter 3, section 2.3). This assay allows the ultra-sensitive quantification of double-stranded DNA. Knowing the concentrations of each amplicon per each sample is essential to normalize all of them to the uniform concentration of 0.5 ng/ul (with EB buffer: 10mM Tris-HCl, pH 8.0-8.5). This step guarantees an equal representation of each amplicon in the library when pooled.

1.3.4. Preparation of the library

The term library refers to the pool of the amplicons that will be later sequenced. As explained in the introduction (Chapter 1, section 5.1) to be sequenced the library must be attached to the flow cell. This occurs through the interaction of the adapters, that must be added at the extremes of each amplicon, with the oligos on the flow cell's surface. The steps required to attach the adapters to the amplicons are detailed below.

First, the cohesive ends of the amplicons must be converted into blunt ends to later permit the ligation of the adapters. For this purpose, the 5' end was phosphorylated and an adenine nucleotide (A) added to the 3' end. This step was performed using the Kapa Hyper Prep Kit using the End Repair & A-tailing buffer and the End Repair & A-tailing Enzyme Mix (Chapter 3, section 2.6) at 20°C for 30 min and 65°C for another 30 min.

The ligation of the adapters to the amplicon's ends was carried out with the SeqCap Adapter Kit. The thymine (T) present at the end of the adapter anneals with the adenine (A) added in the previous repair step. To promote the ligation of the adapter, the amplicons' pool was incubated with the adapter (at a concentration of 15 μ M), in presence of a ligase at 20°C for 15 minutes.

The products of the ligation were, again, purified by KAPA Pure Beads at a proportion of 0.8X. The clean-up procedure is reported in Chapter 3, section 2.5.

1.3.5. Preliminary amplification of pooled amplicons libraries

A further amplification step is usually recommended for libraries composed by amplicons at low concentration, or difficult to amplify, which is the case of those obtained from HBV and HDV. For this amplification, the KAPA HiFi HotStart ReadyMix PCR Kit (Chapter 3, section 2.6) containing the KAPA HiFi HotStart Ready Mix (2x) and the KAPA Library Amplification Primer Mix (10x) was used following the manufacturer's instruction. The KAPA HiFi HotStart enzyme has a low error rate (1 error per 3.6E+06 incorporated nucleotides) thanks to the strong 3'→5' exonuclease (proofreading) activity, which reduces the introduction of artefacts.

The products of this preparatory amplification were purified, again, by using the Kapa Pure Beads at a proportion of 0.8X, as reported in Chapter 3, section 2.5.

The obtained purified pools were quantified by real time qPCR on the LightCycler 480 instrument by using the KAPA Library Quantification Kit (Chapter 3, section 2.6). This kit contains the KAPA SYBR FAST qPCR master mix, which includes a DNA polymerase capable to amplify different-in-length DNA fragments with the same efficiency. The kit also incorporates a platform-specific library quantification primer premix (banding with P5 and P7 adapters), and a pre-diluted set of DNA standard (six dilutions). These purified pools were diluted up to 1:10000 in EB buffer and quantified by using the KAPA Library Quantification Kit in LightCycler 480.

1.3.6. Preparation of the master pool and sequencing

The term *masterpool* indicates the combination of different libraries in the same tube. To ensure that each library is equally presented and to limit unbalanced amplification during the sequencing procedure, all the libraries were normalised to 4 nM with EB buffer and pooled together. The *masterpool*, diluted at 1:400 and 1:4000, was quantified through the KAPA Library Quantification Kit (Chapter 3, section 2.6). Once quantified, the *masterpool* was adjusted to 4 nM with EB buffer. Both *masterpool* and PhiX were denatured with 0.2N NaOH and later neutralised by adding 200 mM Tris-HCl solution pH 7-8. More information about the PhiX and its usefulness in sequencing are reported in the Introduction chapter (Chapter 1, section 5).

After denaturalising the master pool and PhiX, both were diluted to 12-15 pM with HT1 hybridisation buffer from the MiSeq Reagent Kit v2 for Ribozyme and MiSeq Reagent Kit v3 for 5'HDAG (Chapter 3, section 2.6). Both master pool and PhiX were mixed at a ratio of 4:1 respectively. The mix was subjected to a denaturation step at 95°C for 2 minutes and subsequently loaded into the sequencer.

1.4. Sequencing quality filters and bioinformatic analysis

1.4.1. Quality filters

The reads (sequences) obtained from the sequencing were subjected to several bioinformatic filters by using the open-source R software [288]. The filters are listed below:

1. The reads were analysed position per position and those sequences with indeterminations or those reads shorter than the expected amplicon were eliminated. Moreover, general parameters of the MiSeq Illumina sequencing quality were also evaluated.
2. Since MiSeq Illumina enables a paired-end sequencing (Introduction, 5) both forward (R1) and reverse strands (R2) were collapsed by using

FLASH package in R software [289], imposing a minimum of 20 bp overlap between R1 and R2 with a maximum of 10% mismatches. The reads that do not meet the criteria were excluded.

3. In this step, the Phred punctuation was used. This parameter measures the quality of the nucleotides identification during the sequencing by determining the probability of error. A value of Phred (or Q) equals to 30 corresponds to an estimated accuracy of 99.9% [290]. Based on this reference value, the reads with 5% or more of bases with a quality lower than 30 (Q30) were excluded.
4. Each read was assigned to the specific amplicon and sample by detecting the specific sequence MID (Chapter 1, section 5.1). Briefly, the Illumina adapters were used to distinguish between the different *masterpool* libraries and the M13 to discriminate between the genome regions and between the two strands (up to 3 mismatches allowed). During the MID assignation, just a base mismatch was accepted since these sequences are just 10 nucleotide-in-length. At aligning, MID, M13 and adapters were trimmed, obtaining a *fasta* file for each MID-adapter-strand combination. These reads were then collapsed into haplotypes with the corresponding frequencies.
5. In each *fasta* file, the haplotypes were aligned with the reference sequence, called *master* sequence (the most abundant haplotype in the file) and a quality filter (performance >90%) was applied. Through this filter, the haplotypes that didn't cover the entire amplicon, that had more than two indertermiations, three or more gaps and haplotypes with more than 99 different bases related to the *master* sequence were discarded. The tolerated indertermiations and gaps are repaired according to the reference sequence.
6. Haplotypes with abundance equal or greater than 0.1% at both strands were selected, and those unique for one strand discarded. The coverage

of the haplotypes was calculated as the sum of the reads that passed the quality filter at both strands.

7. For the final selection of haplotypes, those with abundances below 0.25% were discarded. With the rest of the haplotypes, the following computational analyses were implemented.

Both the quality filter and the following bioinformatic studies were carried out by Dr. Josep Gregori.

By applying all the different quality filters, we finally obtained a multiple alignment of the haplotypes forming all the viral QS, accompanied by the relative frequencies.

1.4.2. Study of the QS conservation

HDV genome is characterized by a high variability, as extensively explained in the Introduction (Chapter 1, section 4.2) and the identification of hyper-conserved regions could help us to identify new targets for gene therapy. In this first study of my thesis, I analyzed by NGS the QS of HDV to inspect the conservation and variability of the ribozyme region, and moreover, compare it with the variability of the 5'-HDAG. For this purpose, several bioinformatic analysis were applied.

QS conservation was studied by quantifying the information content (IC) at each position of the haplotypes' multiple alignments. This analysis is based on Shannon's uncertainty. It is defined as the number of binary decisions (number of yes or no questions) required to find the correct element in a set of N elements. In other words, it represents the probability to find a nucleotide in those specific positions of the multiple alignments. The conservation is measured in bits, where the IC ranges are from 0 bits, indicating maximal variability or uncertainty, to 2 bits indicating maximal information or conservation [291] (Figure 32).

$$CI_j = \log_2(N) - \sum_{i=1}^j p_{ij} \log_2(p_{ij})$$

Figure 32. Mathematical formula to quantify the information content (IC). j: position in the alignment, p_j : frequency of the haplotype in the viral quasispecies at position j. N: number of possibilities (4 nt or 22 aa). The information content (IC) varies from 0 (indicating maximum uncertainty) to 2 (maximal conservation, as results of the $\log_2(4)$ where 4 indicates the number of nucleotides).

The information content along the entire amplicon's length is calculated by applying a sliding window analysis, in which the IC is calculated in windows of 25 nucleotide that steps 1 nucleotide forward (Figure 33) At each step the average of bits is calculated thus resulting in the sliding window's curve where each point shows the conservation of each specific position in the multiple alignment. This analysis could be conducted by considering or not the haplotypes' relative frequency.

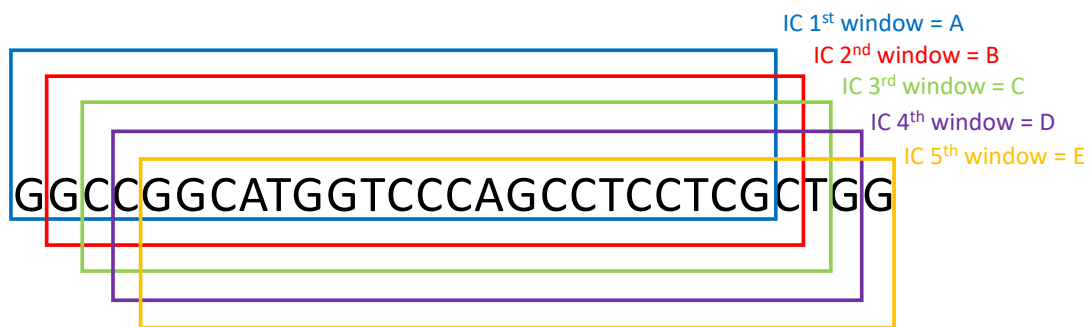


Figure 33. Schematic representation of the information content analysis by sliding window of the viral quasispecies. A corresponds to the information content calculated for the first 25 nt-long window, B for the second 25 nt-long window and successively for C, D and E.

The conservation can also be displayed by using a logo representation where the sequence is showed by the letters representing the four nucleotides and whose height gives a measure of the specific nucleotide conservation (from 0 to 2 bits) [292].

1.4.3. Analysis of HDV quasispecies: incidence-based diversity indices

To have a preliminary idea of the composition of the viral population, regardless the abundance of each variant, we first evaluated incidence-based indices such as the number of reads per sample, the number of reads of the master sequence (Mstr), the percentage of the master sequence in the quasispecies (Mptc), and the number of haplotypes. As commented in the introduction (Chapter 1, section 4) the master sequence is the most prevalent sequence in the viral quasispecies. The more this sequence is prevalent the less complex is the quasispecies. Other indices are the number of polymorphic sites (sites involved in mutations) and the number of the mutations in the quasispecies per each sample, which provide a measure of the variability of the viral population.

Although this first evaluations can give a preliminary idea of the quasispecies's diversity, these indices cannot reflex the entire complexity of the viral quasispecies, since they do not take in consideration parameters such as the relative frequency and the relation between the different haplotypes, which can profoundly impact the

complexity and variability of the quasispecies. Moreover, the viral quasispecies is the result of a dynamic system characterized by continuous cycles of acquisition of mutations and selections, which implies that different indices are needed to have a complex picture of the viral population and its changes [246,293].

1.4.4. Analysis of HDV quasispecies: abundance-based indices

The Abundance-based indices evaluate quasispecies complexity considering the haplotypes that form the viral population and their relative frequency. They provide a measure of either diversity (number and frequency of different haplotypes) or evenness (uniformity of the haplotype distribution) of the viral quasispecies [246]. Here, we focused on two indices: the Shannon entropy (S_n) and the Gini-Simpson diversity index (H_{Gs}).

The Shannon Entropy (S_n) defines the complexity of the viral quasispecies by considering the relative frequency of the haplotypes and, in some measure, their fitness. In other words, it defines the degree of disorder (increasing randomness), that is the uncertainty in assigning a random genome to the corresponding haplotype in the studied population: the greater the disorder the greater difficulty in associating a random genome to a specific haplotype. It is calculated by considering the relative frequency of each unique sequences (Figure 34). Values range from 0 to 1 bit, where 0 is a quasispecies composed by a single haplotype (no disorder, identical viral population) and 1 a high degree of uncertainty. In term of haplotypes the presence of rare haplotypes will be contributed to a higher value of the index.

$$H_s(p) = -\sum_{i=1}^H p_i \log(p_i)$$

Figure 34. Formula calculating the Shannon entropy (S_n): p_i : relative frequency of the haplotype i in the QS; \log is the logarithm; the symbol Σ indicates the summation of the value obtained per each haplotype and H is the number of haplotypes in the QS population.

The Gini-Simpson index (H_{GS}) indicates the probability of two genomes selected randomly from the viral population correspond to different haplotypes. This index ranges from 0 (a unique haplotype) to 1 (infinite number of haplotypes at the same abundance) [246,294]. The formula considers the relative frequency of the haplotypes (as seen for the Shannon entropy), but without calculating the logarithm (Figure 35), which implicates that this index is less sensitive to the presence of rare haplotypes (at lower frequency), since it gives more weight to those with higher frequency.

$$H_{GS} = 1 - \sum_{i=1}^H p_i^2$$

Figure 35. Formula calculating the Gini- Simpson index (H_{GS}). p_i : frequency of haplotype i in the viral population. the symbol Σ indicated the summation of the value obtained per each haplotype.

1.4.5. Analysis of HDV quasispecies: functional indices

Differently from the abundance-based, the functional indices provide information about the QS variability, the haplotypes' diversity, and the presence of mutations. These are highly important since they consider the mutations that distinguish a haplotype from another in the viral population. In this part of my thesis, two functional indices were analyzed: the mutation frequency and the nucleotide diversity.

The mutation frequency (M_f) shows the fraction of mutated sites in the multiple alignments with respect to the dominant haplotype (master sequence) normalized to the number of haplotypes that form the QS (Figure 36). Based on this definition, the presence of highly mutated haplotypes (even at low frequency) will implicate an increased mutation frequency, as more mutated sites are observed correlated to the most abundant haplotype [295].

$$Mf = \frac{1}{H} \sum_{i=1}^H d_{1i}$$

Figure 36. Formula calculating the mutation frequency (Mf). d: Proportion of mutations in the haplotype i related to the master sequence. H: the number of haplotypes.

The other functional-based index used in my thesis is the nucleotide diversity (π or P_i). This index evaluates the average number of substitutions between randomly paired genomes of the viral QS. This provides a measure of the overall heterogeneity of the population, considering the differences between any two genomes in the population [246,296]. It is calculated by considering the distance (the number of substitutions) between two haplotypes of the same population, taking into account their relative frequency (Figure 37).

$$\pi = \sum_{i=1}^H \sum_{j=1}^H d_{ij} p_j$$

Figure 37. Formula calculating the nucleotide diversity. p_i : the relative frequency of the haplotype i; p_j : the relative frequency of the haplotype j; d_{ij} : the number of substitutions between i and j; the symbol Σ indicate the summation of the value obtained per each couple of haplotypes.

1.4.6. Variability in HDV quasispecies and single nucleotide variations.

To inspect the presence of single nucleotide variations (SNVs), the distance between paired quasispecies was calculated and represented by a heatmap. It was calculated as the logarithm of the division between the distance between the quasispecies X and the quasispecies Y (D_{XY}) and the sum of the nucleotide diversity average of each quasispecies (D_X and D_Y) (Figure 38).

$$\log_{10}(2D_{XY}/(D_X + D_Y))$$

Figure 38. Formula calculating the quasispecies distance. D_{XY} is the distance between the quasispecies X and Y, whereas D_X and D_Y show the nucleotide diversity average of each one of the quasispecies.

In the obtained heatmap, the tone of the green indicates the degree of distance, from the lighter green, which corresponds to a shorter distance, to the darker green, which indicates greater sample-sample distance. Distance was calculated by considering the QS consensus of each sample.

Once analyzed the global distance between the different quasispecies, we also studied the single nucleotide variants (SNVs) in each haplotype. The SNV were detected by aligning the haplotypes with its specific quasispecies consensus. The evolution rate during the follow-up for both *ribozyme* and *HDAG* was also studied by normalizing the number of polymorphic sites for the amplicon length and the follow-up time (years). This enabled to compare *ribozyme* and *HDAG* variability during the time considering that both amplicons were different in length.

1.5. Analysis of the effects of the observed mutations on viral expression *in vitro*

As previously commented into the introduction (Chapter 1, section 2.3), the ribozyme plays a key role in HDV replication, since it is responsible of the liberation from the concatemer of the single genome and antigenome monomers during genome amplification. Considering this function, mutations in the HDV ribozyme sequence could potentially affect the viral replication.

To answer this question, the effect of most relevant mutations was studied *in vitro* to inspect how they can impact on viral replication. For this purpose, the mutations were introduced in an HDV-coding plasmid and the ability of the mutated plasmids to sustain viral expression was tested by cellular transfection. The methodology used in these steps is detailed in the paragraphs below.

1.5.1. Introduction of the mutations in pCMV-HDV-1.2X plasmids

The pCMV-HDV-1.2x plasmid is a plasmid containing 1.2 copies of genomic cDNA of HDV under the control of a strong promoter, the cytomegalovirus promoter (CMV) (Figure 39) [297]. This promoter is commonly used for high-level production of recombinant proteins in mammalian cells [298]. Moreover, this plasmid carries an antibiotic resistance (against ampicillin) for its following selection in bacteria. The mutations were introduced using site-directed mutagenesis (QuickChange Lightning site-directed mutagenesis kit, Chapter 3, section 2.7) following manufacture's protocol. The primers for the mutagenesis were designed by using the Agilent online tool, and they were totally complementary, and both included the desired mutation in the middle of their sequence. We introduced two-point mutations (T23C and T69C) and a 1nt-long deletion (C64del). The sequence of the primers is reported in Materials (Chapter 3, section 1, table 4).

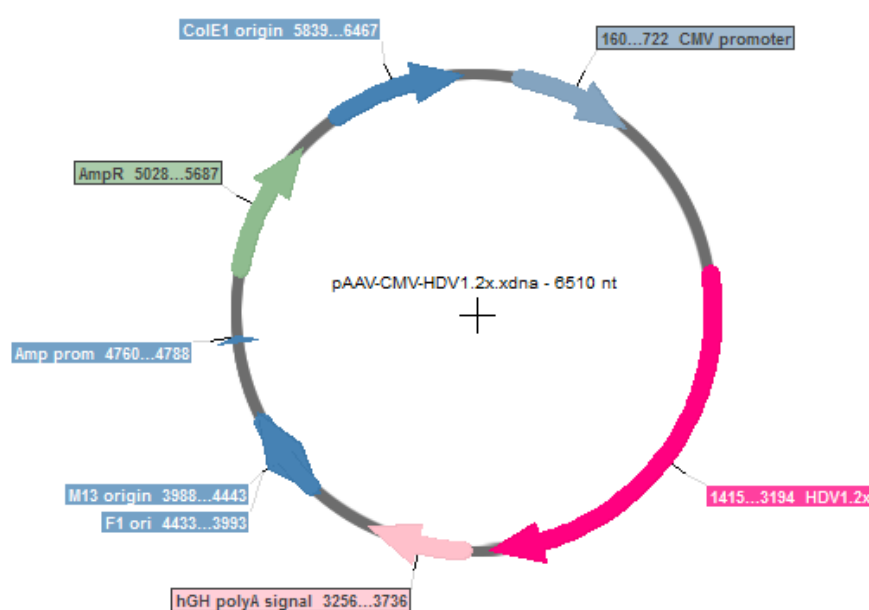


Figure 39. pCMV-HDV-1.2x plasmid main regions. The HDV genome (HDV1.2x- nucleotides 1415-3194- in pink) is under the control of a strong promoter, the CMV promoter (160-722 in dark grey) and the human Growth Hormone (hGH) polyA signal (nt 3256-3736 in light pink), for the termination and polyadenylation of the mRNA produced into the cells. The selection is provided by a gene that confers resistance to the ampicillin (promoter: nt 4760-4788 and gene nt 5028-5687, in respectively blue and green), whereas the replication of the plasmid into bacteria is allowed thanks to the origin of replication in position 5839-6467 (dark blue- ColE1 origin). The F1 and M13 origin (nt 4433-3993 and 4433-3993, in blue) are two portions (with reverse orientation) for the encapsidation of the linearized plasmid into phage particles in presence of phage viral proteins.

The QuickChange Lightning site-directed mutagenesis kit enables the introduction, in a plasmid, of site-directed mutations or of limited-in-length deletion or insertions. The enzyme provided in the kit is a proprietary blend that includes a derivative of the PfuUltra high-fidelity (HF) DNA polymerase (Chapter 3, section 2.4), for ensure the optimal amplification with high fidelity of the two strands of the plasmid. Once amplified, the modified plasmid is subjected to a digestion with DpnI enzyme which selectively targets methylated DNA (that are those plasmids directly extracted from bacteria, such as the original wild-type plasmid).

The mutated and digested plasmids were subsequently amplified by transforming in XL10-Gold ultracompetent bacteria cells by heat shock (30' at 4°C, followed by 30" at 42°C and 2' at 4°C). Transformed bacteria were pre-incubated with SOC medium (2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄ and 20 mM glucose) in the shaker at 37°C for 1 hour at 225 rpm. SOC medium does not contain antibiotics, so it allows the unselective growth of all the bacteria presented. To later support the growth of just those bacteria that were optimally transformed with the plasmid (selection), the bacteria in SOC were plated on ampicillin-treated (1mg/mL) agar plates and incubated overnight at 37°C.

Several colonies were picked and maintained in nuclease-free water at 4°C for the following PCR and sequencing steps, which will be explained with more details in the following paragraph. Just that colony that presented the desired mutation in a clean sequence underwent the extraction protocol.

First the colony was pre-inoculated into 500uL of LB broth with Ampicillin (1mg/mL) for at least 1 hour at 37°C and 225rpm. After this time, bacteria were moved in 100mL of LB broth with Ampicillin (1mg/mL) and incubated at 37°C and 225rpm over-night. The following day, the bacteria were centrifugated and the plasmid extracted from the pellet through the Endotoxin free NucleoBond Xtra Midi Plus kit (Chapter 3, section 2.7) following manufacturer's instructions. The purity of the plasmid was evaluated by charging it on the NanoDrop One system (Thermo Fisher Scientific, Wilmington, USA) which is an UV-spectrophotometer that enables the analyses of RNA, DNA, protein, pigments, etc. allowing the determination of the

extracted DNA's purity in terms of contamination by RNA or alcohols. The concentration (ng/ μ L) of the extracted plasmids was determined by fluorometric quantification using Qubit fluorometer (Chapter 3, section 2.3).

1.5.2. Sanger sequencing

To ensure the correct introduction of the mutations, the plasmids were sequenced by Sanger sequencing. For this purpose, the single picked colonies maintained at 4°C, as described in the previous paragraph, were sequenced. A region covering the full ribozyme was first amplified (primers Seq Ribo FW and Seq Ribo RV, Materials, 1) using the FastStart Taq DNA Polymerase dNTPack (Chapter 3, section 2.4). The PCR products were analyzed by gel electrophoresis and then purified by ExoSAP-IT™ PCR Product Cleanup Reagent (Chapter 3, section 2.5), which enables the clean-up of the PCR products by enzymatically eliminate both the excess of unpaired-primers and nucleotides. Purified PCR products were afterwards analyzed by Sanger Sequencing.

This method of DNA sequencing is also known as chain terminator method because it is based on a PCR called chain-termination PCR, that works like a classical PCR but differs in the addition of 3'-OH lacking nucleotides (dideoxy ribonucleotides, ddNTPs). Each of these nucleotides are provided with a different fluorochrome (4 fluorochromes per the 4 nucleotides). In the PCR mix, the ddNTPs are added at low ratio related to the normal dNTPs. During the elongation step, the nucleotides are added according to the complementarity of the sequence, until the introduction of the ddNTPs ceases the elongation step. The result of this step is a mix of DNA fragments of different length that present at the end a fluorescent ddNTP. These fragments are thereupon separated and ordered by capillary electrophoresis, nucleotides are excited by a laser and detected by the fluorochromes. A software then computes the overlapping or contiguous sequencing resulting in an electropherogram that shows the sequence of the region of interest.

In this part of my thesis, the BigDye® Terminator v3.1 Cycle Sequencing Kit (Materials, 2.5) followed by a purification step with the BigDye® Xterminator™

Purification Kit (Thermo Fisher Scientific, Foster City, USA) was performed following manufacturers' guidelines.

1.5.3. Culture of Huh7 cells and *in vitro* testing

The hepatocarcinoma Huh7 cells were cultured with Dulbecco Eagle's minimal essential medium (DMEM) supplemented by 10% of fetal bovine serum (FBS), penicillin (100 U/mL) (Materials, 3), streptomycin (100 µg/mL), and Glutamax (2 mM). Cells were plated at 160,000 cells/mL and they were transfected with 0.5 µg of wild-type (WT) or mutated DNA plasmids using the Magnetofectamine O2 kit (Chapter 3, section 2.7) according to the manufacturer's protocol. Since HDV needs of the HBV surface antigen to produce and release new viral particles, a plasmid containing 1.3-length HBV genome (pTriEx-HBV) was included in each condition. Both pTriEx-HBV and pCMV-HDV-1.2x WT or mutated were co-transfected, so the amounts of plasmids were in equal proportion (0.25 µg of each).

The supernatant was collected 72 h after transfection and HDV RNA was extracted using MagNA Pure LC Total Nucleic Acid Isolation Kit (Chapter 3, section 2.2) with the Automated sample preparation system MagNA Pure 24 (Roche Applied Science, Indianapolis, IN, USA). Once extracted, the viral RNA was treated with DNase I (Chapter 3, section 2.4) to eliminate potentially contaminant DNA plasmid. HDV RNA was quantified using the LightCycler RNA Master HybProbe kit (Chapter 3, section 2.3), on the LightCycler 2.0 system (Roche Diagnostics, Mannheim, Germany) with DP15 and DP2 primers and HDV-FL and HDV-LC probes (Chapter 3, section 1, table 2), as reported in this chapter section 1.2.1, table 5.

2. Results

2.1. Patients selected in the study of ribozyme conservation and variability

In the first part of my thesis, I analyzed the conservation and variability of ribozyme in chronic delta patients (CHD) with more than 3 log₁₀IU/mL of HDV RNA and attended in Vall d'Hebron University Hospital (Barcelona, Spain).

Twenty-five patients were included in this study. Considering that one of the goals of this part was to follow-up the ribozyme quasispecies during the time, we considered two samples per each patient with a total of 50 samples and a mean of follow-up of 2.25 years.

Although HDV RNA was extracted and amplified in all the samples, after applying the quality filter as reported in this chapter, section 1.4.1, only 19 patients (38 samples) were finally considered in the study, with a median (Q1-Q3) of reads of 4550.5 (1342.24–5892.74) per patient. All patients presented more than 3 log₁₀IU/mL of HDV RNA, with no changes between the two timepoints (median (Q1-Q3) log₁₀ HDV RNA of 5.76 (5.02–5.83) and 5.76 (3.68–5.76), respectively) (Table 8). Moreover, they were not significant differences in term of transaminases and platelets between both follow-up samples (Table 8). The HBV DNA was undetectable in both samples in all patients. HBV surface antigen was resembtant in both samples (median (Q1-Q3) of 3.96 (3.57-4.11) to 3.92 (3.39-4.93) log₁₀ IU/mL).

Table 8. Main clinical and viral characteristics of chronic hepatitis delta patients Included in the study. The table shows the clinical and viral characteristics of the patients considered in this study between the 2 follow-up samples (sample A and sample B). The p values were obtained by applying a Kruskal Wallis-test (p-value = 0.05). Abbreviations: AST -Aspartate aminotransferase (normal value 12-50 IU/mL); ALT- Alanine aminotransferase (8-50 IU/mL); platelets (140-400 UI/mL); Q1-Q3, Quartile range; HBsAg, Hepatitis B virus surface antigen; ns, not statistically relevant.

Markers	Sample A	Sample B	P
HDV RNA Median (Q1-Q3) Log ₁₀ (IU/mL)	5.76 (5.02 – 5.83)	5.76 (3.68 – 5.76)	ns
AST (UI/L) Median (Q1-Q3)	88 (45 – 133)	90 (36.25 – 126)	ns
ALT (UI/L) Median (Q1-Q3)	100 (58 – 158.5)	89.25 (44.5 – 133.75)	ns
PLATELETS (UI/L) Median (Q1-Q3)	89 (123 – 212)	79 (118 – 197)	ns
HBV-DNA (IU/mL)	Low/IND	Low/IND	
HBsAg Median (Q1-Q3) Log ₁₀ (IU/mL)	3.96 (3.57-4.11)	3.92 (3.39-4.03)	ns

As previously explained in this chapter's 1.1 section, HDV genotypes were analyzed by NGS through the sequencing of the 5'-end of HDAG (nt 910 -1270). Of note, all the patients were infected by genotype 1. Just a patient showed a non-1 genotype, however, after the sequencing, these samples did not pass the quality filter, so this genotype was discarded in the study.

2.2. Conservation of the quasispecies in the ribozyme region

QS conservation was analyzed by aligning all the haplotypes of each sample and calculating the information content through the application of a sliding window analysis. As shown in figure 40, the ribozyme (between positions 688 and 771 of HDV genome) was overall highly conserved and 85% of the nucleotide positions presented 2 bits of information content (100% of conservation) (Figure 40, panel A). This analysis was conducted by both considering or not (respectively red and blue line in figure 40, panel A) the haplotypes' relative frequency. Of note, no difference in term of conservation was observed when comparing both lines.

To deeply inspect the conservation along the ribozyme, the sequence was displayed as a sequence logo. As explain in Chapter 4, section 1.4.2, the logo shows the sequence and the degree of conservation of each nucleotide, providing a more detailed representation of the sequence conservation. As shown in Figure 40, panel B, of the 85 nucleotides included in the ribozyme sequence, just 3 nucleotide positions (around 3.5% of the total) presented a conservation less than 1.5 bits. Although these three positions seemed less conserved, the degree of bits was still quite high (more than 1.2 bits), thus reinforcing the high level of conservation of this portion of the HDV genome.

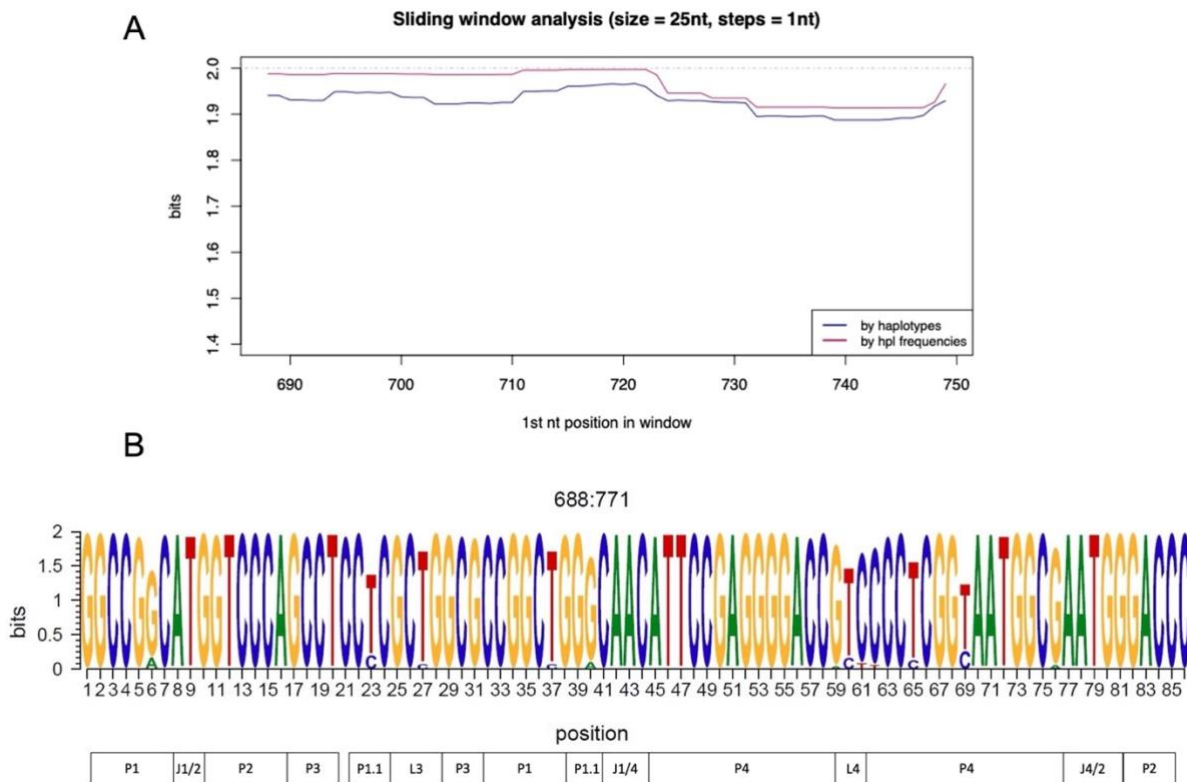


Figure 40. Conservation of ribozyme region by sliding window analysis and represented by logo. (A) The sliding window analysis is the result of the mean information content (bits) of the 25-nt windows with a displacement between them of 1-nt obtained by multiple alignments of all the quasispecies (QS) haplotypes. The analysis was implemented by considering (red line) or not (blue line) haplotypes frequency. The dashed line represents the maximum level of conservation (2 bits). (B) Logo representation of the nucleotide sequence corresponding to the entire ribozyme region from the genome positions 688 (corresponding to position 1 in the ribozyme) to 771 (corresponding to position 86 in the logo representation). The height of each letter represents the grade of conservation from a maximum of 2 bits to a minimum of 0. The sequence is shown in the genome sense. The different structural and functional domains of the ribozyme are reported at the bottom [368]

Despite the general conservation of the ribozyme quasispecies, we tried to identify the most conserved and most variable portions of the region, since one of the aims of this study was to identify possible target of a gene silencing therapeutic strategy against HDV infection.

The most conserved region was between the nucleotide positions 715 and 745 (nucleotides 28 to 58 if considering just the ribozyme region). The 93.6% of the sequence presented 2 bits, whereas only two nucleotide positions (the 725 and 728, corresponding to the positions 38 and 41 of the ribozyme) presented a conservation higher than 1.5 bits. Of note, this highly conserved region comprised part of double-stranded regions (P3, P1, P1.1 and P4), the complete single-stranded J1.1/4 region and part of the looped L4 region (Figure 41, panel A). The role of all these above quoted domains in the ribozyme structure and functionality are reported in the Introduction chapter (Introduction, 2.3).

The most variable region (Figure 41, panel B) encompassed the positions 739 and 769 (nt 51 to 82 of the ribozyme). Here, 19.3 % (6/31) of the nucleotides presented lower degree of conservation (between 1.2 and 1.5 bits). The region partially overlapped with the highly conserved region in a lengthy portion of six nucleotides (between nucleotides 739 and 745) however, all of them presented the maximum level of conservation. When looking for the functional domains that fell in this “more variable” region, we observed that these nucleotides comprise part of the double-stranded P2 and P4 regions, the entire L4 loop region and single-stranded J4/2 region [299].

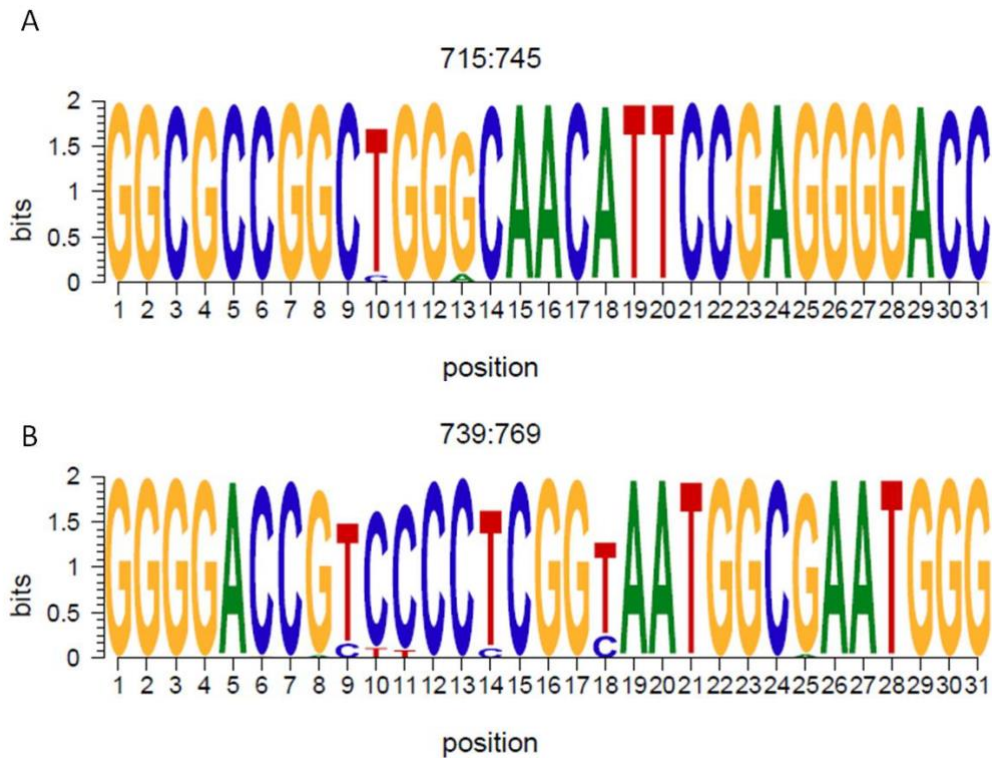


Figure 41. Logo representation of the most conserved and most variable portion of the ribozyme. The logo represents the most conserved (A) (nt 715–745) and most variable (B) (nt 739–769) regions of the ribozyme. The height of each nucleotide represents its information content in bits (2 bits indicates 100% of conservation). The sequence is shown in the genome sense.

2.3. Ribozyme Quasispecies (QS) evolution and variability during patient’s follow-up

The genetic distance (Chapter 4, section 1.4.3) between the samples was determined to evaluate the ribozyme QS evolution during the patient’s follow-up (Figure 42).

As expected, considering the high level of conservation, the distance between sample A and B was very short (<0.25) for most of the patients (white-light green in Figure 42). Although patients P2, P12, P13, P21, and P23 seemed to present a greater distance (>0.50) compared to the others, when their sequences were examined in depth, very few nucleotide substitutions were observed. Due to the high degree of sequence conservation, the presence of few mutations might magnify the distance between the quasispecies.

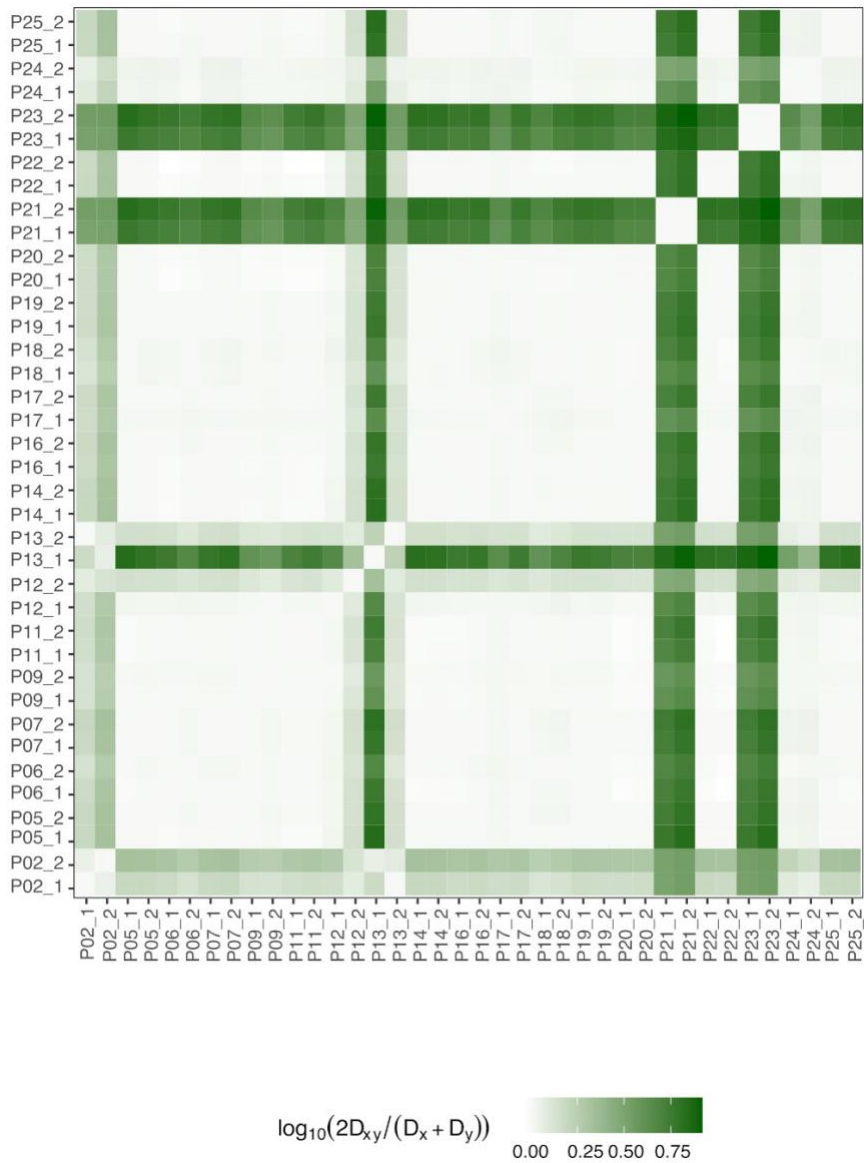


Figure 42. Sequences' distance between samples. The heatmap shows the distance of the ribozyme sequences between the different samples included in the study as the logarithm of the ratio between quasispecies nucleotide distance (D_{xy}) and the nucleotide diversity average of each quasispecies (D_x and D_y). The tone of the green indicates the degree of distance, from the lighter green, which corresponds to a shorter distance, to the darker green, which indicates greater sample-sample distances. Distance was calculated by considering the quasispecies consensus of each sample.

To obtain a complete picture of the viral population and its composition in the HDV ribozyme, we analyzed different quasispecies complexity and diversity indices, as summarized in Table 9. For more details related to the indices and their determination, please refer to the chapter 4, section 1.4.3. Again, and consistent with the high degree of conservation and the limited genetic distance between sample A and B, no differences were observed.

Table 9. Complexity and diversity indices calculated in samples A and B in the ribozyme region. The table shows the median and Q1-Q3 of the following indexes: number of reads per sample; number of master sequence reads (Mstr); percentage of the master sequence (Mpct); number of haplotypes; number of polymorphic sites; Shannon index; Gini-Simpson coefficient; functional attribute diversity (FAD); mutation frequency (Mf); nucleotide diversity (Pi) and Pi to Mf ratio. P values were obtained by applying a Kruskal Wallis-test (p-value = 0.05). ns = not statistically relevant.

Complexity index	Sample A Median (Q1-Q3)	Sample B Median (Q1-Q3)	<i>p</i>
N reads	3785 (1196.5 – 4981.5)	4909.5 (1550.5 – 6460)	ns
N reads of master (Mstr)	2462.5 (1109 – 3571.5)	3161.5 (1440 – 4601.5)	ns
% Master (Mpct)	9.15 (85.17 – 94.33)	10.52 (83.9 – 94.4)	ns
N haplotypes	2.5 (6.5 – 9)	2.5 (6.5 – 9)	ns
Polymorphic sites	3 (5 – 8)	3 (5 – 8)	ns
N mutations	3 (5 – 8)	3 (5 – 8)	ns
Shannon index	0.32 (0.31 – 0.63)	0.41 (0.29 – 0.71)	ns
Gini Simpson coefficient	0.16 (0.1 – 0.26)	0.18 (0.1 – 0.28)	ns
Functional attribute diversity (FAD)	0.89 (0.7 – 1.6)	0.77 (0.1 – 1.48)	ns
Mutation frequency (Mf)	0.001 (0.0006 – 0.0017)	0.0012 (0.0006 – 0.0018)	ns
Nucleotide diversity (Pi)	0.002 (0.0012 – 0.0033)	0.0023 (0.0012 – 0.0036)	ns
Pi to Mf ratio	0.10 (1.86 – 1.97)	0.11 (1.86 – 1.97)	ns

2.4. Analysis of Mutations

The presence of single nucleotide variations (SNVs) in the ribozyme were studied by aligning the haplotypes with their master sequence. Although the ribozyme was conserved overall, when studying each sample individually, a total of 48 mutations were observed in the full region of the ribozyme. When looking at mutation type, we observed that the most abundant changes were transitions: C→T (16 mutated positions), G→A (10 mutated positions) and T→C (9 mutated positions) (Figure 43).

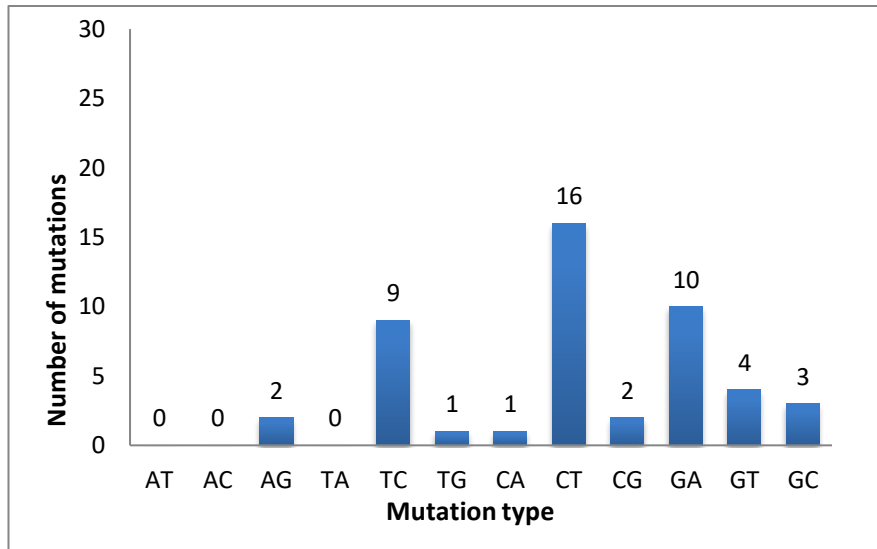


Figure 43. Type of the mutations observed in ribozyme sequence. The bar plot shows the number of mutations per each transition or transversion observed in the ribozyme quasispecies. The number is also reported.

Of the 48 observed mutations, 12 of them appeared in at least six patients in both follow-up samples. Each mutation is reported in the table 10 accompanied by its relative frequency in respectively samples A and B.

Table 10. Most abundant mutations observed in the ribozyme. The table shows the 12 most abundant mutations observed in the ribozyme. The relative frequency of each mutation in respectively sample A and sample B is reported as mean \pm standard deviation. The positions of the mutations are reported by considering the HDV genome or the first nucleotide of the Ribozyme.

Mutation (Position in HDV genome)	Mutation (Position in Ribozyme)	Sample A (%freq)	Sample B (%freq)
G693A	G6A	0.84% \pm 0.43%	0.83% \pm 0.31%
T710C	T23C	3.04% \pm 2.27%	5.55% \pm 8.91%
T714C	T27C	0.12% \pm 0.3%	0.88% \pm 3.5%
G727A	G40A	0.25% \pm 0.22%	0.41% \pm 0.24%
G746A	G59A	0.07% \pm 0.03%	0.08% \pm 0.02%
T747C	T60C	6.15% \pm 30.64%	6.42% \pm 35.01%
C748T	C61T	0.15% \pm 0.24%	0.14% \pm 0.27%
C749T	C62T	0.12% \pm 0.17%	0.08% \pm 0.12%
C751d	C64d	47.17% \pm 1.66%	48.88% \pm 2.47%
T752C	T65C	6.02% \pm 39.66%	6.29% \pm 44.04%
T756C	T69C	12.02% \pm 21.99%	12.75% \pm 44.04%
G763C	G76A	0.53% \pm 2.49%	0.44% \pm 17.68%

Of the observed mutations, only three involved at least eight patients and were maintained or selected (mutation frequency that changes between the two timepoints) during the follow-up: T710C (position 23 if considering the first nucleotide in ribozyme from position 688), T756C (position 69 in ribozyme) and the deletion in position 751 (position 64 in ribozyme). The T23C mutation involved 17 of the 19 patients in sample A with a mean frequency of 2.99%, and 18/19 of the patients with a mean frequency of 4.96% in sample B (Figure 44, panel A). The T69C mutation was found in both samples in 100% of patients, with a relative frequency that was maintained during the follow-up (10.6% and 10.8% in respectively sample A and B) (Figure 44, panel C). The deletion in position 64 (C64d) was highly prevalent in quasispecies, with a frequency of 98.7% in sample A and 98.3% in sample B and involved half of the patients (8/19 patients in both samples) (Figure 44, panel B).

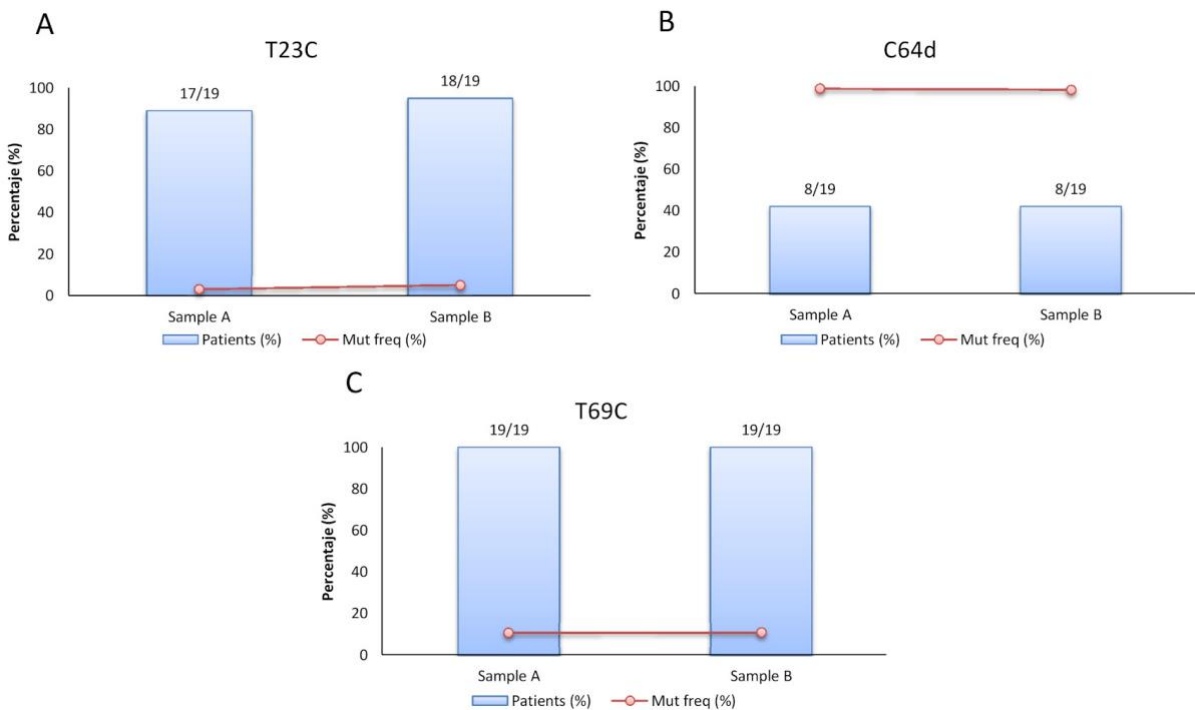


Figure 44. Incidence of the main mutations observed in the ribozyme. The graphs show the relative frequency (red line), the percentage (%) and number of patients (blue bar) presenting the mutation T23C (A), C64d (B) and T69C (C) between the 2 follow-up samples (sample A and B).

By looking the mutations' relative frequency in each patient, we didn't observe marked changes in most of the patients with some exceptions (Figure 45). The mutations T23C and T69C, indeed, were positively selected in patient 2 (from 5.29% to 17.07% and from 43.88% to 71.85% for respectively the positions 23 and 69), and in patient 12 (from 9.52% to 36.97 and from 2.79% to 24.11%) (Figure 45 panel A and C). Of note, the T69C was also negatively selected in patient 13 (from 91.15% to 39.09%) (Figure 6C). In the case of the C64d deletion, it was maintained between the different timepoints in patients at almost 100% (Figure 45, panel B).

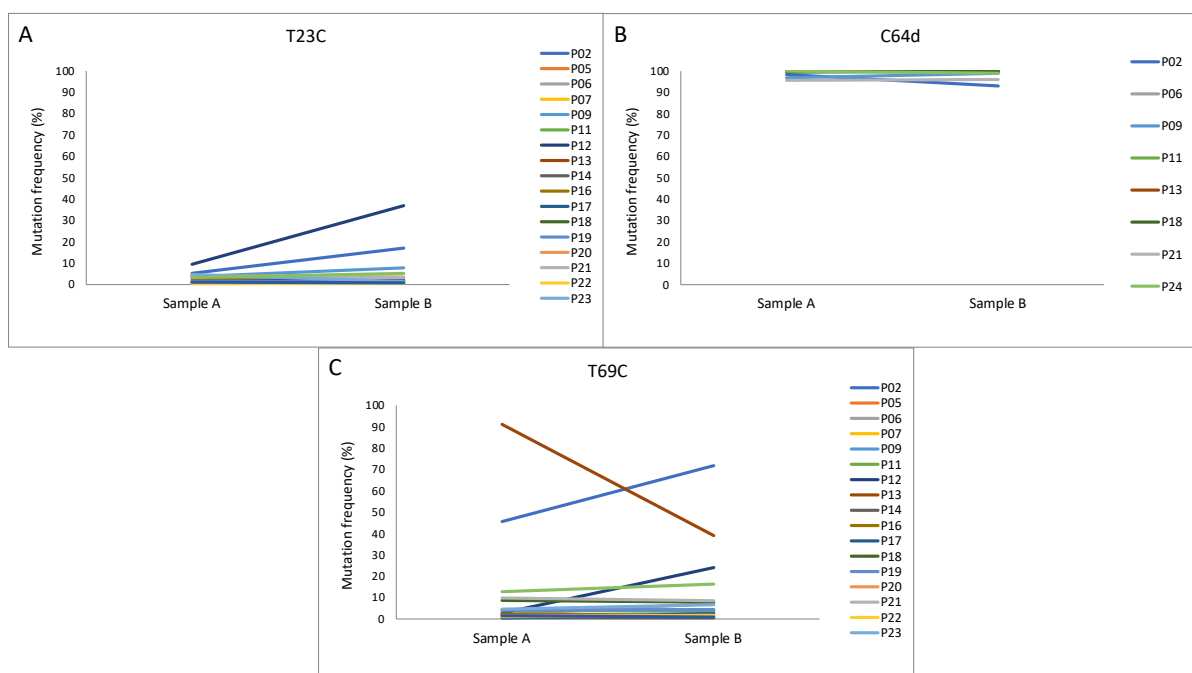


Figure 45. Evolution of the mutation frequency between the two follow-up samples (A and B). Each patient is represented by a different color. The x-axis represents the two-follow-up sample, the basal (Sample A) and the follow-up (Sample B) whereas the y-axis shows the relative frequency (%) of the mutation in the viral quasispecies. T23C mutation; Panel A, C64 deletion; Panel B and T69C mutation; Panel C.

2.5. Effects of the mutation on the viral expression *in vitro*

Considering the high degree of ribozyme conservation, the observed mutations may potentially affect HDV fitness, thus promoting or inhibiting viral expression. To test the effect of the three main detected mutations (T23C, C64d, T69C) on HDV expression, they were introduced in a plasmid containing 1.2-in-length HDV full

genome and mutated plasmid were used to transfect HuH7 cells. HDV RNA titer was quantified in transfected cells supernatant at 72 h after transfection.

Through this system we efficiently expressed HDV *in vitro*, reaching a HDV RNA titer in supernatant of $2.86 \pm 0.61 \log_{10}$ IU/mL. The T69C mutant showed a similar replication rate than WT (2.78 ± 1.04 UI/mL, $p = 0.662$). Similarly, the mutation T23C did not affect viral replication (mean \pm std \log_{10} IU/mL = 2.7 ± 0.38 IU/mL, $p = 1$). Differently, the interference with viral expression was stronger in presence of the deletion in position 64 (C64d), that caused a reduction in viral expression of around 1.23 log related to the WT virus (mean \pm sd \log_{10} HDV RNA = 1.63 ± 0.71 UI/mL, $p = 0.08$) (Figure 46). P-value were calculated by applying a Mann-Whitney test.

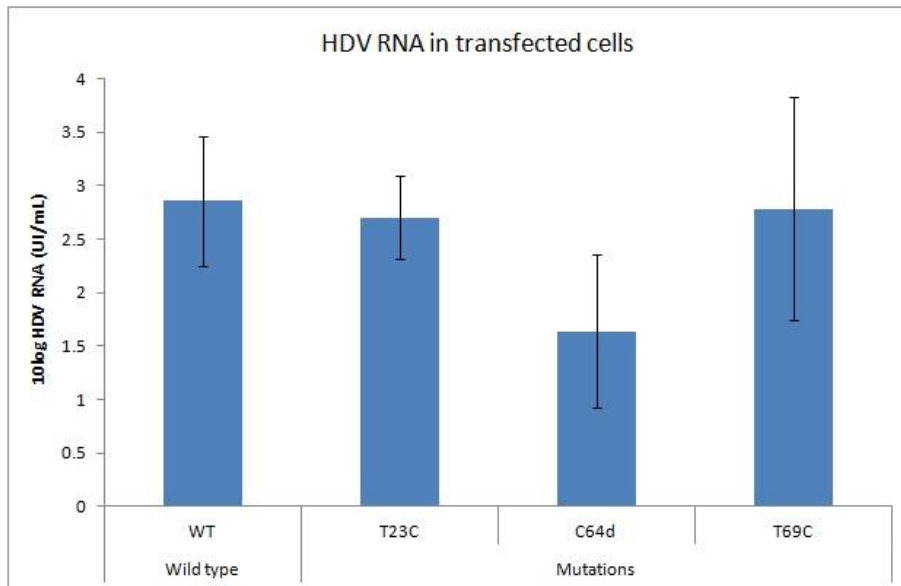


Figure 46. Effect of the most relevant observed mutations in HDV replication *in vitro*. The graph shows the HDV titer in cellular supernatant after transfection with HBV and WT or mutated HDV after 72h of transfection.

2.6. Ribozyme quasispecies variability and evolution compared to the 5'-HDAG

To have a more complete sketch of the quasispecies of HDV and its evolution during the time, we also analyzed HDAG quasispecies (in the 5'-extreme). As explained in this chapter's 1.3.2 section, this portion of the viral genome comprises the C-terminal domain of the HDAG, including the Amber/W site, where the editing process of the messenger RNA occurs. To evaluate the differences in term of quasispecies between the ribozyme and the 5'-HDAG, in this second part of my thesis, we amplified the 5'-HDAG and analyzed this region by NGS by using the same samples previously analyzed in the ribozyme.

2.6.1. Clinical characteristics of the patients included in the study

The 5'-HDAG was amplified in all the 19 patients of the previous described study, however after passing the quality filters just 9/19 patients were suitable for the following analysis, and finally just these patients were involved in the study. As done for the ribozyme, we analyzed per each patient two longitudinal samples with an average \pm standard deviation follow-up interval of 1.3 ± 0.5 years. Patients presented similar level of HDV RNA and transaminases (ALT and AST) between the two follow-up samples (Table 11). After the application of the different quality filters a median (Q1-Q3) of 14720.5 (1679.5-16399.75) reads per sample were obtained.

Table 11. Virological and biochemical markers. The table shows the evolution of both virological (HDV RNA) and biochemical (ALT and AST) markers between the two follow-up samples. Each marker is shown as median (Q1- Q3). The two samples were statistically compared by applying a Wilcoxon test.

	Follow-up median (Q1-Q3)		p-value
	A	B	
HDV RNA (log ₁₀ IU/mL)	5.4 (4.6-5.7)	4.6 (3.8-5.7)	ns
ALT (UI/L)	58.5 (44.2-71)	63.0 (46-84)	ns
AST (UI/L)	44(33.5-64)	57 (34-90)	ns

2.6.2. Viral quasispecies complexity between ribozyme and 5'-HDAG

By comparing the viral quasispecies between the ribozyme and the 5'-HDAG, a strong difference was observed in term of both abundance and complexity indices. For an explanation of the different indices considered in this study, please refer to sections 1.4.4, 1.4.5 and 1.4.6 of this chapter. The ribozyme showed a higher frequency of the master sequence related to the 5'-HDAG (Figure 47, panel A), respectively 90.3 % (85.2-94.2) and 19.1% (12.9-27.5) (median(Q1-Q3)). This result was reverted when comparing QS complexity, the 5'-HDAG showed higher levels of both Sn and GiniS (median (Q1-Q3) respectively 3.4 (2.9-3.7) and 0.9 (0.9-1.0)) than ribozyme (Sn and GiniS of 0.5 (0.3-0.6) and 0.2 (0.1-0.3) (Figure 47, panels B-C).

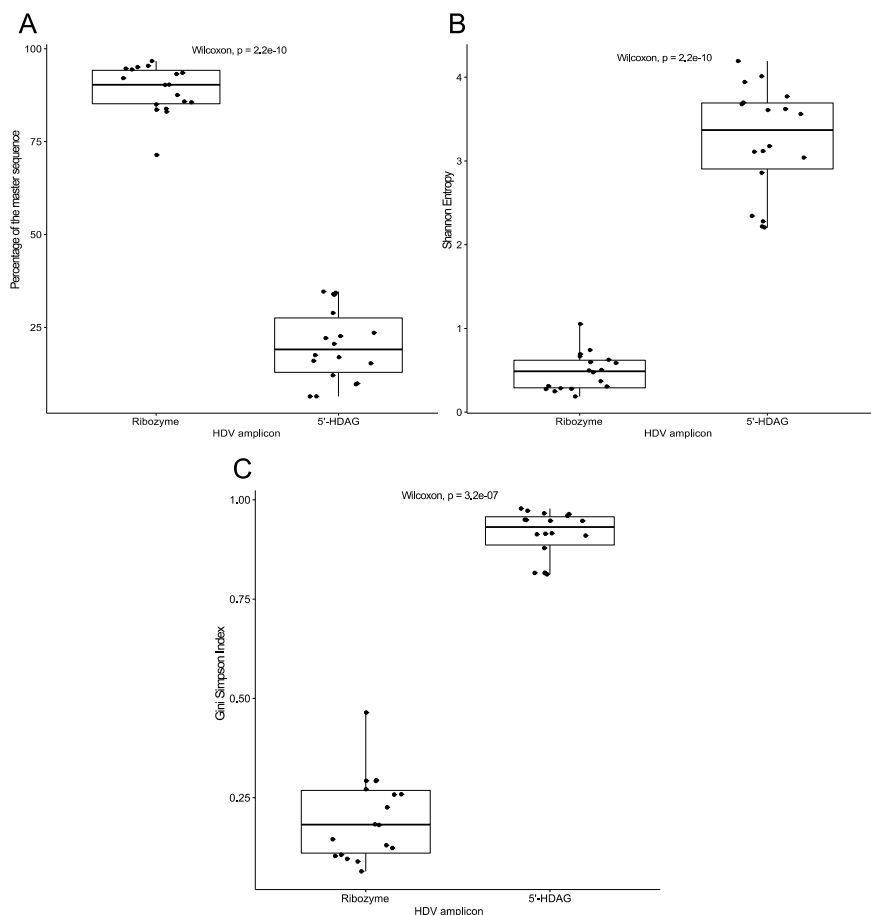


Figure 47. QS complexity indices between the two different HDV genomic regions. The figure shows the comparison between ribozyme and 5'-HDAG in term of frequency of the master sequence (A), complexity indices (Gini Simpson index and Shannon Entropy, respectively panel B and C). The index for each sample (independently of the timepoint) is represented by a dot. The p-value and the statistical test used in the comparison is also reported.

In term of variability, the 5'-HDAG region presented a higher mutation frequency (0.004 (0.003-0.007) versus 0.001 (0.001-0.002) in ribozyme, Figure 48, panel A) and a higher number of polymorphic sites (15.5 (13.3-17.8) versus 7 (5;8), Figure 48 panel B). Of note, when considering ribozyme and 5'-HDAG quasispecies throughout the time, we didn't observe any change in term complexity and variability between both follow-up samples.

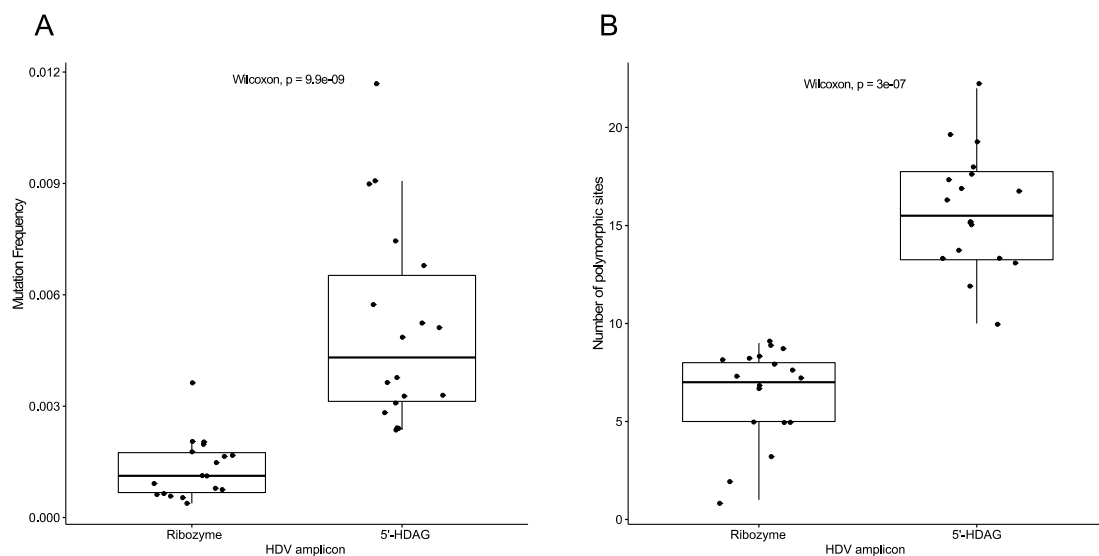


Figure 48. QS variability indices between the two different HDV genomic regions. The figure shows the comparison between ribozyme and 5'-HDAG in term of diversity indices Mutation frequency (panel A) and number of polymorphic sites (panel B) The index for each sample (independently of the timepoint) is represented by a dot. The p-value and the statistical test used in the comparison is also reported.

2.6.3. Quasispecies evolution rate between *ribozyme* and *5'-HDAG*

Since we did not observe great changes when comparing the quasispecies between the two timepoints for both the viral genomic regions, the evolution rate between samples A and B was calculated. Because of the different length between the amplicons, the evolution rate of the QS was calculated by considering the polymorphic sites per year of follow-up per position. When comparing the evolution rate between the two HDV genomic regions, we didn't observe a strong difference between the regions, although the HDAG appeared to show a tendency of higher evolution rate (median (Q1-Q3) of respectively 0.021 (0.014-0.026) and 0.025 (0.02-0.038), Figure 49, panel A).

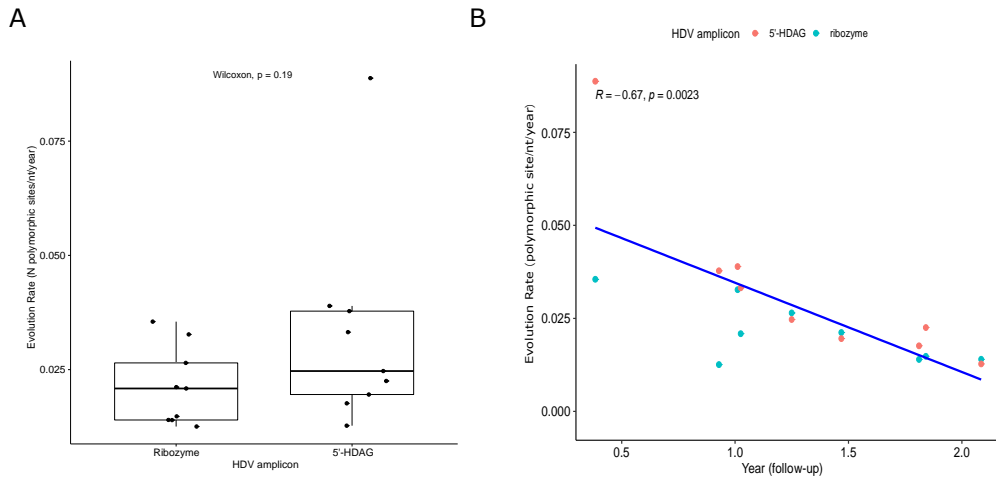


Figure 49. Evolution rate of the QS related to the region and the elapsed time. A) Boxplot shows the comparison of the Evolution rate (number of polymorphic sites normalized per the amplicon length and the year of follow-up) between the two HDV genomic regions. The statistical test and the p-value is reported. B) Spearman correlation between the evolution rate and the time of follow-up. The different colors show the different HDV amplicons (pink for 5'-HDAG and light blue for ribozyme).

Notably, the highest evolution rate was mainly observed for shorter follow-up time, and a negative strong correlation was detected when considering together the evolution rate and the time between the two samples (Figure 49, panel B). This correlation was even more strong when just considering the 5'-HDAG region alone (rho = -0.93, p-value = 0.00075- Figure 50, panel B) but not with the ribozyme (Figure 50, panel A).

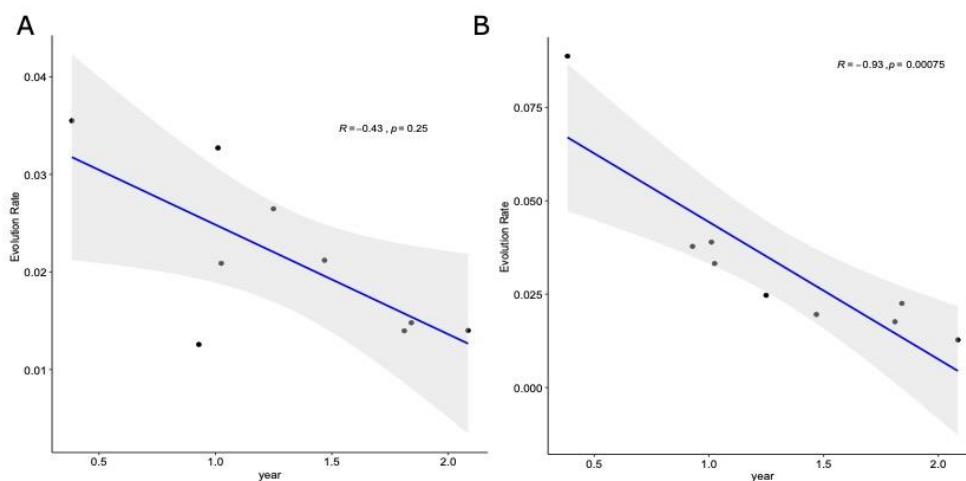


Figure 50. Correlation between evolution rate and follow-up time in respectively ribozyme and 5'-HDAG. The scatter plots show the correlation between ribozyme (A) or 5'-HDAG (B) evolution rate and the time of follow-up. Each dot represents a patient. The Spearman rho and the p-value are also reported.

Study 2

In vivo study of the involvement of HDV in HBV variability: the role of type I interferon pathway

1. Methods

1.1. Superinfection mouse model

To determine the variability of HBV in presence of HDV, we used a model of HDV-related superinfection in collaboration with the group of Dr. Gloria Gonzalez Aseguinolaza from the Centre for Applied Medical Research (CIMA) in Pamplona. The mice employed in this study were kindly donated by Dr. F.V. Chisari (La Jolla, California). These mice were modified to constitutively express HBV thanks to the insertion of a terminally redundant over-length 1.3 viral DNA (HBVtg) [300]. The insert starts just before the X promoter and Enh I region and ends after the single polyadenylation site of HBV and promotes the expression of viral particles at high level (more than 10^6 copies/mL in serum/plasma) comparable to what observed in patients with chronic hepatitis B (Figure 51).

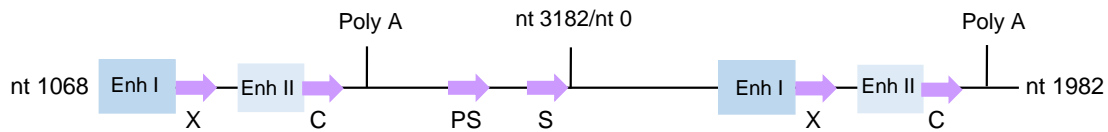


Figure 51. Inserted HBV genome. The figure shows the composition of the over-length 1.3 HBV DNA inserted in HBV transgenic mice, detailing all the regions and ORF included. There are represented the different ORFs (X, C, PS and S) and both enhancers (Enh I and Enh II). There are also represented the PolyA sites and the end/beginning of the HBV genome (nt 3182/ nt 0). Figure modified from Guidotti *et al.* 1995 [300].

To develop the HDV superinfection model, HBVtg mice were administered with 5×10^{10} vg adeno-associated vectors (AAV) containing an over-length 1.2 copy of the HDV genotype 1 antigenome (rAAV-HDV). The 1.2x HDV antigenome was obtained from pDL456 plasmid and cloned into a pAAV-MCS vector (Agilent Technologies, Santa Clara, United States), as previously presented by Dr. Gracian Camps in his thesis. To guarantee the tissue-specific expression of HDV, viral genome was inserted behind a liver-specific chimeric promoter composed by the albumin enhancer (EAlb) fused to the hepato-specific promoter of the α -1 antitrypsin (AAT), followed by the β -globin intron (1), which promotes mRNA intracellular stability and protein expression [301] (Figure 52). Moreover, the plasmid, so called pAAV-

EAlb/AAT-HDV, was also provided of a polyadenylation sequence to promote the RNA nuclear export and translation. The expression cassette was flanked by two inverted terminal repeats (ITR) which are essential for the encapsidation of the DNA into the recombinant AAV (rAAV). These latter were produced by co-transfecting HEK293T cells with pAAV-EAlb/AAT-HDV and pDP8.ape helper plasmid (Plasmid Factory, #PF478) that provides the genes essentials for AAV replication and encapsidation. The rAAV containing the 1.2 HDV genome was subsequently concentrated and purified by Iodixanol gradient ultracentrifugation. Once purified and quantified, rAAV-HDV particles were intraocularly injected to HBVtg mice (HBVtg AAV-HDV). To avoid biased results, in the HBV monoinfection model the HBVtg mice were injected with 5×10^{10} vg of rAAV expressing luciferase (HBVtg AAV-Luc).



Figure 52. Insert of the pAAV-EAlb/AAT-HDV. To promote the encapsidation of the 1.2 HDV genome into the rAAV particles, the viral genome (in blue) was flanked by two IRT (Inverted terminal repeats) sequence. The high level of expression of the viral genome was ensured by the presence of a strong and tissue-specific promoter (EAlb/AAT in green). The β -globin intron and the polyadenylation sequence (in respectively orange and red) supported the stability and translocation of the produced viral RNA. Figure obtained from Dr. Gracián Camps thesis.

1.1.1. HBVtg mice knock-out for the IFN α / β R

As reported in the Introduction chapter (Chapter 1, section 3), and differently from HBV, HDV infection strongly activates the intracellular interferon pathway, which might impact HBV variability due to the activation of mutagenic enzymes such as ADAR1. To evaluate this effect, a HBVtg mouse model knock-out for the expression of IFN α / β receptor (IFN α / β R) was produced by Dr. Gracian Camps in Dr G Gonzalez's group by using CRISPR-Cas9 system.

Briefly, HBVtg mice were crossed with IFN α / β R KO mice to obtain a progeny heterozygous for the IFN α / β R KO. Heterozygous female and male were later crossed together to obtain a homozygous progeny (HBVtg IFN α / β R KO). Recombinant AAV

(both HDV and Luc) were injected in both WT or HBVtg IFN α / β R KO to obtain monoinfected or superinfected models in presence of both, normal and inhibited type I interferon pathway. A brief report of the four experimental conditions used in this study is reported in figure 53. Notably, these models were maintained and manipulated at the animal facility of the CIMA-University of Navarra.

Superinfected and monoinfected mice were sacrificed at 7- or 21-days post-injection (dpi).

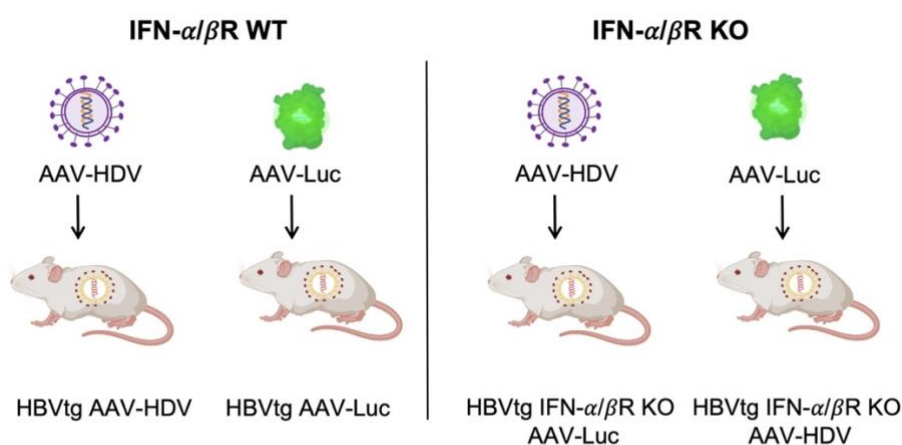


Figure 53. Experimental conditions. The figure shows a resume of the experimental conditions used in this study. Both IFN α / β R WT and KO mice were injected with 5×10^{10} rAAVs encoding luciferase or HDV to lead to the monoinfected and superinfected models respectively.

1.2. Analysis of HBV expression in serum

Serum samples were collected from HBVtg mice before (day -2 of AAV injection) and weekly after rAAV injection.

HBV titter was measured by Dr. Camps as reported in his thesis. Briefly, viral DNA was extracted from 100 μ l of serum by using the High Pure Viral Nucleic Acid Kit (Chapter 3, section 2.2). Eluted HBV DNA was later quantified by qPCR with a couple of primers specific for the HBcAg using the GoTaq $\text{\textcircled{R}}$ qPCR Master Mix (Chapter 3, section 2.3). A standard DNA of known concentrations was used for the absolute quantification of HBV DNA and results were reported in vg/ml.

Circulating HBV RNA was quantified by real-time PCR on the COBAS 6800 instrument using an HBV RNA Investigational Assay (Chapter 3, section 2.1). This technique has a linearity of 10 to 10E+09 copies/mL and a limit of quantification of 10 copies/mL [135,302].

As previously commented in the Introduction (Chapter 1, section 1.4), HBV RNA is a direct marker of cccDNA expression [135,302]. In this case, although it represents the expression from the integrated DNA in HBVtg, this marker might help us to hypothetically determine the step of the HDV-related inhibition of HBV expression (if at the time of retrotranscription, or directly at HBV DNA expression).

1.3. Analysis of the intrahepatic HBV expression: pgRNA

Viral RNA was obtained from liver tissue samples at both timepoints (7 and 21 dpi) and then retro-transcribed to obtain the cDNA for the following quantification and sequencing of the intrahepatic HBV RNA. All the materials used in the next paragraphs are reported in Chapter 3, sections 1 and 2.

The commercial High Pure Viral Nucleic Acid Kit (Chapter 3, section 2.2) was used to extract the viral RNA from mice liver. To avoid DNA contamination, the extracted RNA was first treated with Turbo DNase (Chapter 3, section 2.4) at 37°C for 30 min followed by a heat-inactivation at 75 °C for 10 min and Mg⁺² remotion by the addition of EDTA (25 mmol/L EDTA solution).

The transcription of the RNA into cDNA was performed in two steps. In the first step, the digested RNA was denatured by treating it at 65°C for 5 minutes in presence of the Oligo(dT)₁₇ primer (Chapter 3, section 1, table 3), which is complementary to the poly-A tale of each messenger RNA. To promote the annealing of the primer to the denatured RNA, the reaction was cooled up to 20°C for another 5 minutes. In the second step, the AccuScript High-Fidelity RT-PCR System enzyme (Chapter 3, section 2.4), was added to the reaction together with the RNase Out inhibitor (Thermo Fisher Scientific-Life Technologies, Austin, United States), which is a recombinant protein that inhibits the ribonuclease thus preventing RNA

degradation. The reverse transcription step was carried out by incubating at 42°C for one hour, followed by an inactivation step at 70°C for 15 minutes.

Once retrotranscribed, intrahepatic pgRNA was quantified by qPCR by using a standard curve consisting of serial dilutions of an in-house produced standard pgRNA obtained by the *in vitro* transcription of a plasmid containing the full-length HBV genome (Figure 54). The standard was obtained by cloning the HBV 1.3 genome from pTriEx1.1-HBV1.3x plasmid into a pEF6/V5-His-TOPO™ vector (Invitrogen, Carlsbad, California) (Figure 54, panel A) which is provided of a T7 promoter for the following *in vitro* transcription. The HBV DNA was inserted between the T7 promoter and a NotI restriction site (pEF6.HBV1.3). To ensure that the standard RNA followed the same route of the RNA extracted from the samples, a poly-dA tale was inserted by site-directed mutagenesis (QuickChange Lightning site-directed mutagenesis kit, Chapter 3, section 2.7) between the end of the insert and the NotI site (pEF6.HBV1.3-polyA). To obtain the pgRNA standard, pEF6.HBV1.3-polyA was linearized by an overnight digestion with NotI (37°C followed by a heat-inactivation at 65°C). Linearized DNA was *in vitro* transcribed through the MEGAscript T7 Transcription Kit (Chapter 3, section 2.4) following the manufacturers' protocol specification (Figure 54, panel B). The obtained RNA was purified using the MEGAclean Transcription Clean-up Kit (Chapter 3, section 2.5). To further prevent DNA contamination, a DNase I treatment was performed. Standard RNA purity was evaluated through the NanoDrop spectrophotometer (Thermo Fisher Scientific-Life Technologies, Austin, TX, United States), whereas its concentration was determined through the Qubit fluorimeter using the Qubit RNA HSAssay Kit (Chapter 3, section 2.3). RNA concentration was later used to calculate the concentration in number of copies per mL. The standard curve was made by applying 10-fold serial dilutions from 7.51 to 1.51 Log₁₀ copies/mL (Figure 54, panel C).

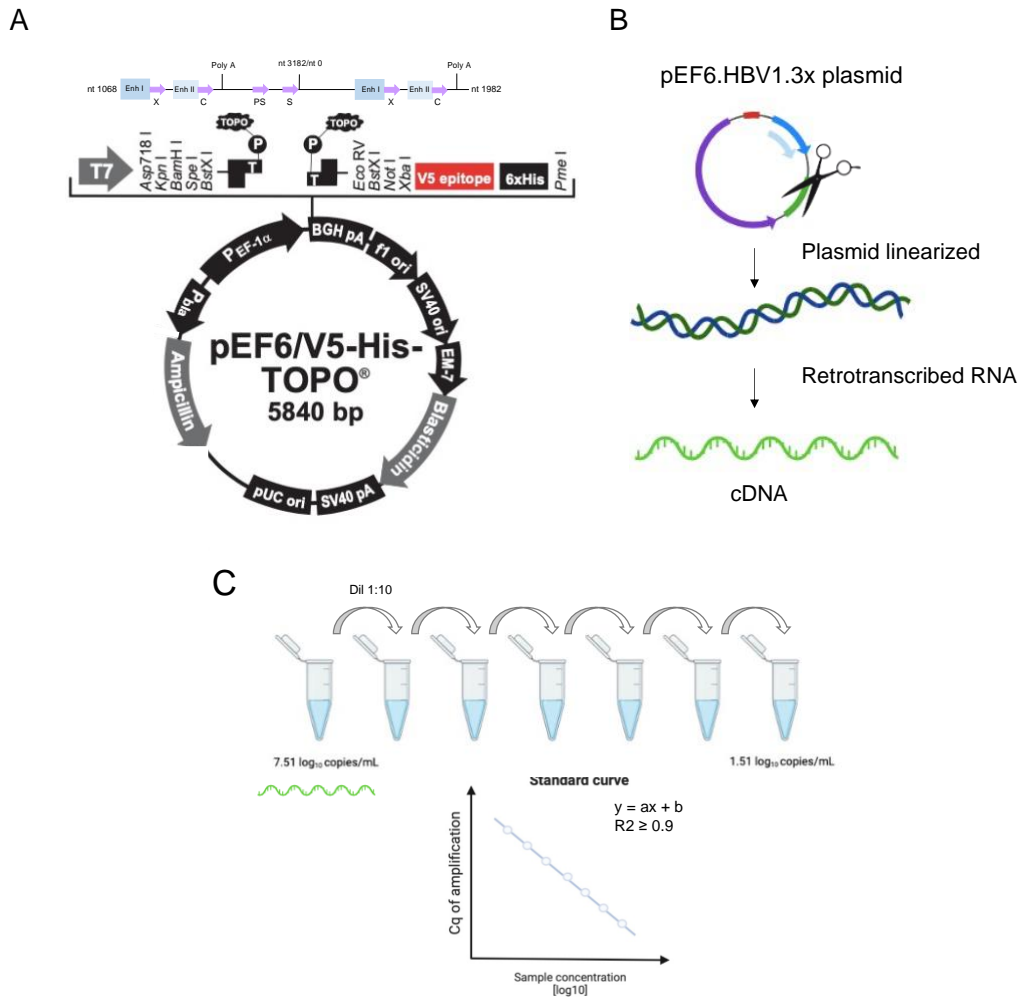


Figure 54. Graphical representation of the in-house obtention of the pgRNA standard curve. A) The insert from the pTriEx1.1-HBV1.3x plasmid was cloned into a pEF6/V5-His-TOPO™ vector between the T7 promoter and the NotI restriction site (pEF6.HBV1.3). B) To produce the pgRNA standard, the pEF6.HBV1.3 was linearized by treating it with NotI enzyme and then in vitro transcribed. The standard RNA also contains a polyadenylation sequence for the following reverse transcription. C) The obtained standard pgRNA was serially diluted up to 1.51 log₁₀ copies/mL and analyzed by the Lightcycler 2.0 system. The quantification cycle (Cq) obtained per each dilution was determined in duplicate and related to the standard's concentrations to obtain a regression curve. The standard curve is considered optimal when the R-squared (R²) was higher than 0.9.

To quantify intrahepatic pgRNA, the viral RNA molecules were purified from mice liver as reported above, and then treated with Turbo DNase to eliminate contaminant DNA.

The quantification cycle (Cq) for both the standard's dilutions and the mice were quantified by using the LightCycler RNA Master HybProbe on the the LightCycler 2.0 system in presence of pgRNA Fw and pgRNA Rv primers and pgRNA probe (Materials, 1). The used protocol is described in Table 12. The concentration of the

pgRNA in the sample was finally obtained by applying the linear regression curve through the software available in the LightCycler 2.0 system.

Table 12. Protocol used for the pgRNA quantification in LightCycler 2.0 system.

	Temperature (°C)	Hold (Time)	Cycles
Reverse transcription	63	3 min	
Desnaturalization	95	30 sec	
Amplification	95	10 sec	45
	65	30 sec	
	72	1 sec	
Cooling	40	10 sec	

Of note, to ensure that the contaminant DNA was efficiently eliminated during the DNase step, treated RNA was analysed on the LightCycler 2.0 system through the LightCycler® DNA Master HybProbe (Materials, 2.3) with the following protocol: Denaturalization at 95°C for 10 minutes, 45 cycles of amplification at 95°C for 5 seconds, 30 seconds at 65°C and 72°C for 15 seconds). This protocol enables to determine if contaminant DNA was still presented in the DNase-treated RNA sample, since no reverse transcription step was used.

1.4. Next-generation sequencing analysis of HBV and HDV quasispecies variability in the liver of superinfected IFN α / β -R WT or KO mice

To inspect the effect of the activation of HDV superinfection and type I interferon in HBV variability, the viral RNA was extracted from the mice's liver, treated with TURBO DNase and retrotranscribed with Oligo(dt)₁₇ using the Accuscript HiFi reverse transcriptase PCR kit following the protocol reported in the previous paragraph. The obtained cDNA was subsequently amplified to analyse by NGS the viral RNA quasispecies in all the four experimental conditions (HBV monoinfected and HDV superinfected mice, wild-type or knock-out for the type I interferon

signalling). The different regions of the HBV RNA that were studied will be described with more details in the next sub-paragraph (section 1.4.1.).

Since HDV RNA is known to be targeted of intracellular mutagenic enzyme such as ADAR1, the QS of HDV RNA were also studied in both WT and KO superinfected mice (as reported in chapter 3, section 1.3.2 for *5'-HDAG*).

1.4.1. Amplification of the regions of interest in HBV genome

The acquisition of mutations in key portions of viral genome could affect viral fitness thus promoting the inhibition of the viral replication. With this purpose, we analysed two main portions of the HBV pgRNA: the X gene, and a portion of the *HBS* that overlaps with the viral RT (SRT).

As previously commented in the introduction chapter (Chapter 1, section 1.2.4), the full X gene encodes for a key protein responsible for transactivating HBV expression. Moreover, some portions of this sequence that belongs to pgRNA acquire a double stranded configuration (the α -nt 1292-1321-, β - nt 1411-1433- and ϵ -nt 1847-1907- loops) that could be targets of mutagenic enzymes such as ADAR1. The SRT amplicon is also important as mutations at this level might affect the release of new viral particles and/or the retro-transcription of pgRNA to produce new molecules of rcDNA.

The DNase-treated HBV RNA, obtained as above, was retrotranscribed using the AccuScript High Fidelity RT-PCR System enzyme (Chapter 3, section 2.4) in presence of the Oligo(dT)₁₇ primer (Chapter 3, section 1, table 3). This step was common for all the studied regions of the intrahepatic viral RNA. This reaction was performed in three phases: a first denaturation phase (at 65°C for 5 minutes and 20°C for 5 minutes), followed by the retrotranscription (at 42°C for 1 hour), and the RT inactivation at 70°C for 15 minutes. The produced cDNA was later amplified by using different protocols depending on the region under investigation.

The full X gene was amplified by a 3-step-PCR process with two overlapping amplicons corresponding respectively to the 5' and 3' portion of the gene (Figure 55). In the first PCR or external PCR, a bigger amplicon (from nt 599 to 1936) that included all the X gene was produced. The 1340bp-long PCR product was later used for a nested amplification step using two couple of primers (as reported in table 13) generating two overlapping amplicons: the 5'-*HBX* (nt 1255 to 1611) that encompassed the *HBX* promoter and the HBx coding region from nt 1374, and the 3'-*HBX* (nt 1596 to 1936), which comprised the transactivating N-terminal domain of the HBx and overlapped with a portion of preCore. During this second step, an M13 tail was added to both ends of each amplicon, serving as a template for the last PCR, in which a sample-specific multiplex identifiers (MIDs) was added. The primers used for the preparation of the library are reported in chapter 3, section 1, table 3, whereas the protocol used in each step is reported in table 13.

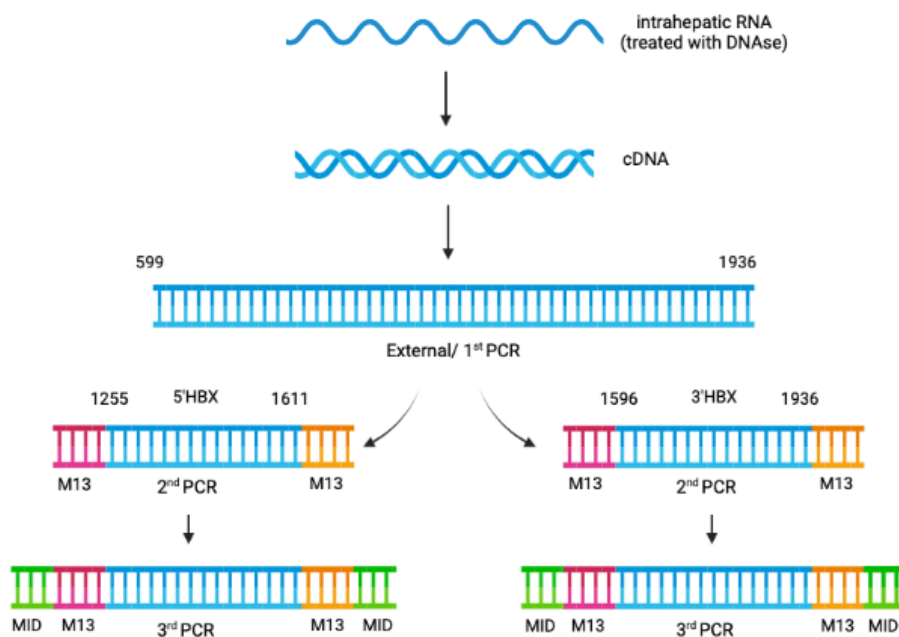


Figure 55. Graphical representation of the amplification of the HBX gene. The viral RNA extracted from the liver of the mono and superinfected mice (both WT and KO for the type I interferon pathway) was retrotranscribed by obtaining a cDNA that was further used as target for the different amplification processes. In case of HBX, the cDNA first underwent to an external amplification encompassing from the nt 599 to 1936. This big amplicon was further amplified during the 2nd PCR step producing two different amplicons, the 5'-HBX (from nucleotide 1255 to 1611) and 3'-HBX (from nucleotide 1596 to 1936). During this step, a M13 tail was added. The 3rd and last PCR used as template the M13 tails and allowed the addition at the extremes of a multiplex identifier (MID) specific for each sample and essential to identify the unique sample during the analysis step.

Table 2. Amplification protocol and primers for the HBX gene.

Amplification step	Primer	Protocol
1st PCR	Ext X FW	95°C 5 min; (95°C 20 s, 53°C 20 s, 72°C 15s) × 35 cycles; 72°C 3 min
	Ext X RV	
2 nd PCR (5' end)	M13 5HBX FW	95°C 2 min; (95°C 20 s, 60°C 20 s, 72°C 15s) × 35 cycles; 72°C 3 min
	M13 5HBX RV	
2 nd PCR (3' end)	M13 3HBX FW	95°C 2 min; (95°C 20 s, 62°C 20 s, 72°C 15s) × 35 cycles; 72°C 3 min
	M13 3HBX RV	
MID PCR	MID FW	95°C 2 min; (95°C 20 s, 60°C 20 s, 72°C 30s) × 35 cycles; 72°C 3 min
	MID RV	

Differently, the SRT amplicon (between the positions 616 and 990) was obtained from the cDNA by a two-step amplification: the primers M13 SRT FW and M13 SRT RV (Chapter 3, section 1, table 3) to produce the amplicon with the addition of the M13 tail and a MID PCR as done for the X amplicons. The protocol used is reported in the table 14.

Table 14. Primer and amplification protocol for SRT gene.

Amplification step	Primer	Protocol
M13 PCR	M13 SRT FW	95°C 1 min; (95°C 20 s, 65°C 20 s, 72°C 30s) × 35 cycles; 72°C 3 min
	M13 SRT RV	
MID PCR	MID FW	95°C 2 min; (95°C 20 s, 60°C 20 s, 72°C 30s) × 35 cycles; 72°C 3 min
	MID RV	

1.4.2. Amplicon purification

The purification of the amplicons is an essential stage to guarantee an optimal sequencing, especially considering the high sensibility of the Illumina system. The amplicons were purified by using the KAPA Pure beads (Chapter 3, section 2.5) in the automatized TECAN instrument (Tecan Trading AG, Switzerland). The proportion of the beads used with this system is a determinant element for the correct purification of the desired amplicon. Depending on the length of the PCR product, a different proportion of beads must be used. In this study the magnetic spheres were used at 0.8X related to the samples (0.8 μ L of KAPA Pure Beads per μ L of DNA). This kit enables the positive selection of the desired beads-attached amplicon during the ethanol-based washing thanks to the interaction of the beads with a magnet.

The quality of the purification was verified through the 2200 TapeStation System and the D1000 ScreenTape kit (Chapter 3, section 2.5). This system, together with the Agilent 2100 Bionalyzer software, allows the visualization of the purified amplicon by electrophoresis. Moreover, thanks to the standard curve provided by the D1000 ScreenTape kit, it furnishes details about the presence and length of different DNA fragments in the sample. This step is essential because the presence of shorter contaminant fragments might reduce the sequencing yield.

Sample quantification, normalization and library preparation were performed as described in the previous study (Chapter 4, from section 1.3.3 to 1.3.6). For both HBV and HDV sequencing MiSeq Reagent Kit v3 was used (Chapter 3, section 2.6).

1.5. Sequencing quality filters and bioinformatic analysis

After sequencing, the samples from this study were filtered according to the same criteria as the samples from the first study. For review, please refer to Chapter 4, section 1.4.1.

Both the quality filter and the following bioinformatic studies were carried out by Dr Josep Gregori.

The finally obtained haplotypes were used to study the viral quasispecies variability in HDV-superinfected mice in presence of wild-type or knock-out type I interferon response.

1.5.1. Rare Haplotype Load study

The Rare Haplotype Load (RHL) defines the quasispecies complexity so far as it is enriched of haplotypes at low frequency. It provides an estimation of the fraction of molecules at low frequency (in the case of this study at 1%) composing the viral QS. A study on HCV revealed that this index could be extremely relevant for the study of highly mutagenic viruses especially in presence of selection pressure, such as the presence antiviral drugs [303].

1.5.2. Analysis of single nucleotide variations

The term single nucleotide variation (SNV) refers to the substitution of one nucleotide for another. These mutations were identified by aligning the haplotypes sequence with the master sequence. SNV's frequency was calculated as the sum of the relative frequency of the haplotypes carrying the mutations per each quasispecies, whereas the frequency per group was obtained computing the average of the SNV's frequency in quasispecies between the mice presenting the mutation. The presence of the different SNVs in the experimental conditions were highlighted by using a Heatmap representation (R software [304]). Colours indicate the relative frequency of each SNV, which were rearranged using a dendrogram algorithm that

shows the relationship between mutations. In addition, to identify a pattern of mutations that might be related to intracellular mutagenic enzymes, the number of transitions (A \leftrightarrow G and C \leftrightarrow T) and transversions (change of purine - A or G- to a pyrimidine -C or T- and vice versa) were calculated by counting the number of positions that presented the specific type of change.

2. Results

2.1. HBV expression in superinfected mice

HBVtg mice express high level of hepatitis B virus. It is well known that HDV infection strongly inhibits the replication of its helper virus. To inspect the effect of HDV expression on HBV replication in our superinfection mouse model, we analyzed different HBV serological and intrahepatic replication markers at different timepoints. Moreover, to inspect how type I Interferon could be involved in the mechanism of inhibition, the viral markers were considered also in knock-out for the IFN α / β pathway mice.

2.1.1. Serological evaluation of HBV replication

To evaluate HBV expression in the superinfected mice model and to relate it with the activation of interferon, a plasma sample was weekly collected up to the end of the experiment (21 dpi) per each mouse.

Similar HBV DNA levels were observed in both monoinfected conditions (HBVtg Luc and HBVtg x IFN α / β R KO Luc) over time. As expected, in presence of HDV, a strong inhibition of HBV DNA was observed in wild-type mice (HBVtg HDV) with an average of log₁₀ of viral genome/ml -vg/mL- of 6.19 ± 0.5 , 5.56 ± 0.49 and 5.33 ± 0.5 at respectively 7, 14 and 21 dpi, versus 6.85 ± 0.85 , 6.56 ± 0.8 and 6.5 ± 0.9 for HBVtg Luc mice (Figure 56). Such inhibition was not observed when suppressing interferon pathway (HBVtg x IFN α / β R KO HDV). Notably, the KO monoinfected mice (HBVtg x IFN α / β R KO Luc) presented higher HBV DNA concentrations in plasma that those HBVtg Luc mice (1.1, 1.13 and 1.18 log₁₀ of difference between HBVtg x IFN α / β R Luc and HBVtg Luc at 7, 14 and 21 dpi).

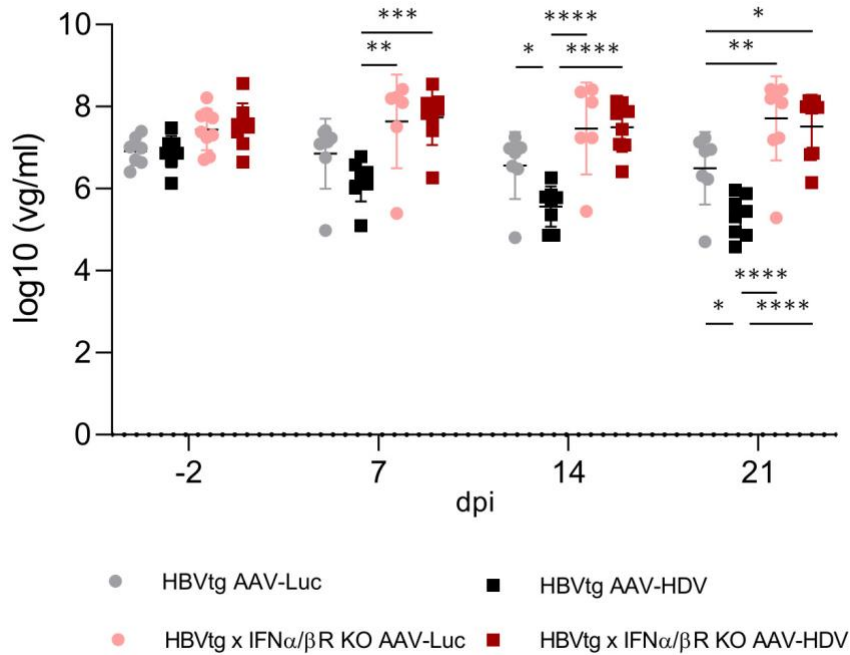


Figure 56. HBV DNA in mice' s plasma. The graph shows the HBV DNA levels (vg/ml) in plasma of each group of mice at basal (before AAV injection), 7-, 14- and 21-days post injection. Each group is represented by a different color (grey and black for WT mice, respectively monoinfected AAV-Luc and superinfected AAV-HDV; pink and red for KO mice respectively monoinfected and superinfected). The central line shows the group's mean. The standard deviation is also reported. P-values are shown as asterisks (* for \leq 0.05, ** for \leq 0.01, *** for \leq 0.001, **** for \leq 0.0001) and were obtained by using an Anova test with the Tukey analysis for multiple comparison after applying the Shapiro Wilk test ($p \geq 0.05$).

Likewise, in HBVtg HDV mice we observed a strong suppression of the HBV RNA levels (Figure 57). Nevertheless, unlike the HBV DNA, in HBVtg HDV mice the HBV RNA was very low even at 7dpi ($0.75 \pm 1.1 \log_{10}$ cps/mL) and maintained this trend up to 21dpi ($0.49 \pm 1.2 \log_{10}$ cps/mL). Monoinfected mice, both WT and KO, presented similar level of circulating HBV RNA in all the explored timepoints. Differently, HBVtg x IFN α / β R KO HDV seemed to present a quite smaller amount of HBV RNA in plasma, especially at 21dpi, although the difference was not statistically significant.

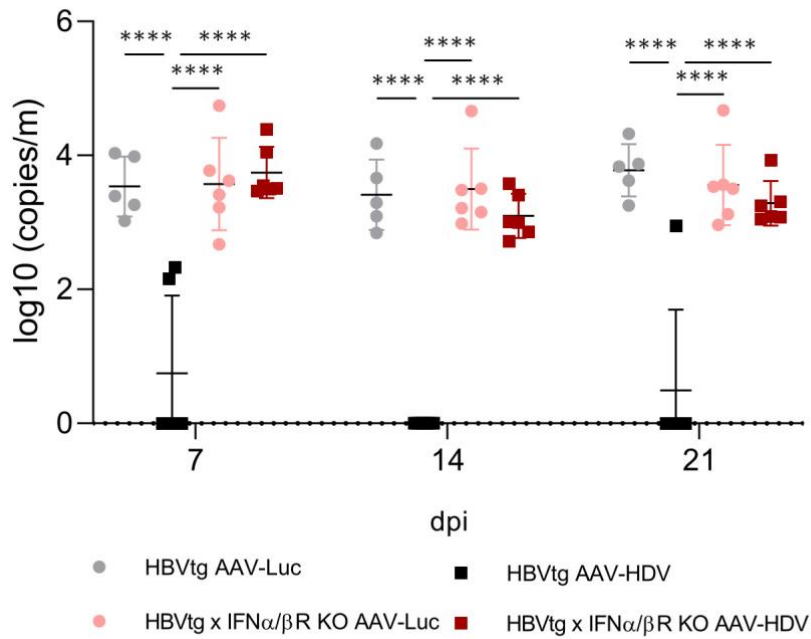


Figure 57. HBV RNA in plasma samples. The graph shows the HBV RNA levels (log₁₀ cp/ml) in plasma of each group of mice at 7, 14 and 21 dpi. Each group is represented by a different color (grey and black for WT mice, respectively monoinfected- AAV-Luc- and superinfected -AAV-HDV; pink and red for KO mice respectively monoinfected and superinfected). The central line shows the group's mean. The standard deviation is also reported. P-values are shown as asterisks (* for ≤0.05, ** for ≤0.01, *** for ≤0.001, **** for ≤0.0001) and were obtained by using an Anova test with the Tukey analysis for multiple comparison after applying the Shapiro Wilk test (p ≥ 0.05).

It must be emphasized that none of the mice had liver damage. The level of ALT transaminase was tested by Dr. Gloria Gonzalez's group and no differences were found between the groups over the time.

2.1.2. Intrahepatic HBV expression

To test the viral intrahepatic expression, the HBV pgRNA was quantified at 7 and 21 dpi by qPCR as described in Chapter 5, section 1.3.

At 7 dpi, the mean values among all groups of mice oscillated from 5 to 5.5 log₁₀ cp/mL. For HBVtg Luc mice, a significant difference was observed between IFNα/β R WT and KO mice, although the average concentrations were quite similar (HBVtg Luc: 5.14 ± 0.2 cps/mL and HBVtg x IFNα/βR KO: 5.43 ± 0.19 cps/mL). The same situation was observed for HBVtg HDV mice, where a little increased of the pgRNA

titer was detected in IFN α / β R KO mice related to the WT (respectively 5 ± 0.15 cps/mL and 5.34 ± 0.16 cps/mL for HBVtg X IFN α / β R WT HDV and KO) (Figure 58, panel A). If considering monoinfected versus superinfected mice, no relevant differences were observed for both WT and KO mice.

At 21dpi (Figure 58, panel B) we didn't observe any difference in the intrahepatic concentration of pgRNA between HBV monoinfected transgenic mice (HBVtg LUC) WT or KO for type I interferon signaling (5.28 ± 0.27 and 5.25 ± 0.39 for HBVtg Luc respectively WT and KO). When comparing HBVtg Luc mice with HBVtg HDV mice, a decrease of around 1 log₁₀ cps/mL was detected (pgRNA titer of 5.3 ± 0.3 and 4.42 ± 0.31 log₁₀ cp/mL for HBVtg Luc and HBVtg HDV). Notably, this reduction was reverted in superinfected mice knock-out for type I interferon (HBVtg x IFN α / β R KO HDV), where the level of pgRNA was the same that in those monoinfected mice (5.3 ± 0.38 log₁₀ cp/mL).

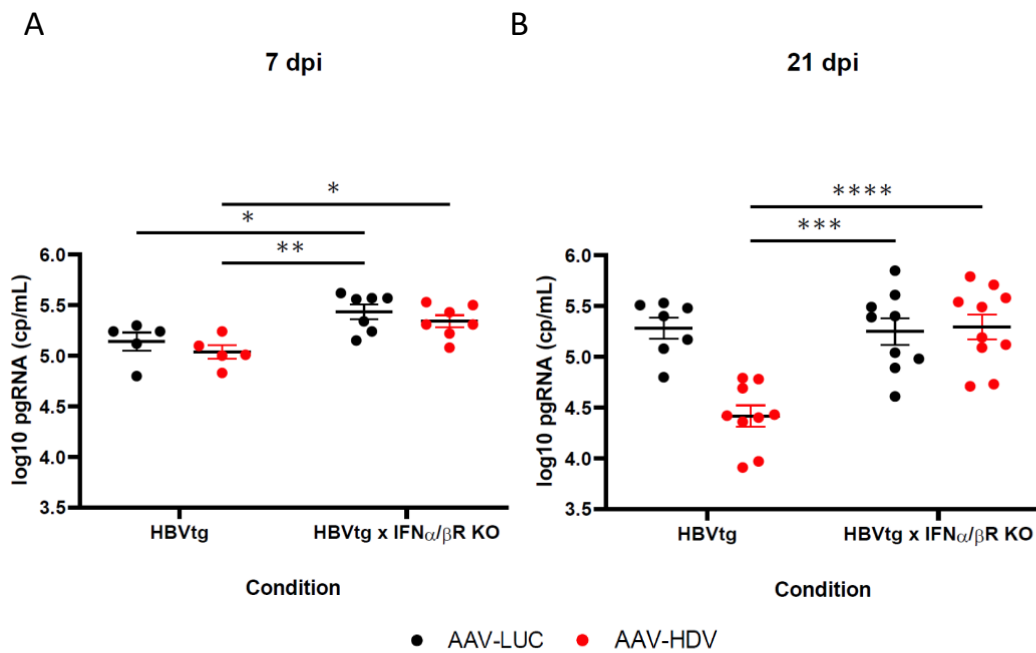


Figure 58. Intrahepatic HBV pgRNA expression in transgenic mice at 7 and 21 dpi. This graph shows the levels of the intrahepatic pgRNA (log₁₀ cp/mL) in the different groups of mice at 7 (panel A) and 21 dpi (panel B). Monoinfected mice (AAV-Luc) are represented by black dots whereas superinfected mice (AAV-HDV) by red dots. P-values is reported as asterisks ($p \leq 0.05$ and $** p \leq 0.01$) and was obtained by an Anova test with the Tukey analysis for multiple comparison after applying the Shapiro Wilk test ($p = 0.05$).

2.2. Analysis of the intrahepatic HBV RNA quasispecies in SRT at 7 and 21 dpi

HBV surface antigen is a key viral protein involved in HBV replication and it completely overlaps with the viral polymerase gene, so mutations acquired in these genes could potentially impact both the release of viral particles and the intracellular viral replication. As explained in the Chapter 5, section 1.4.1, here we amplified, at both 7 and 21dpi, a region encompassing the nucleotide positions 615 to 969 of the P gene of HBV genome, that contains the active domain of the RT, and that overlaps with the nt 615 to 835 of the S gene (corresponding to the last 72 amino acids - aa 155 – 227).

At 7dpi, the amplicon was correctly amplified from the liver of 22 sacrificed mice as follows: 5 HBVtg Luc, 4 HBVtg HDV, 6 HBVtg x IFN α / β R KO HDV and 7 HBVtg x IFN α / β R KO Luc (Table 15). After applying the required quality filters, a median (Q1-Q3) of 16938.5 (31631.25- 48575.75) of reads were obtained. To evaluate the evolution of the changes observed in this region at 7 dpi, the intrahepatic HBV RNA was also analyzed at 21 dpi. For this study, 5 mice were included for both HBVtg Luc and HBVtg HDV type mice and 4 mice were included for both knock-out conditions. A median (Q1-Q3) of 35292.5 (56025-91317.5) reads were achieved.

Table 15. Scheme of the mice included in the study per condition and timepoint.

Mouse strain	AAV injection (5×10^{10} vg of AAV/mouse)	n (Mice that passed filter)	
		7 dpi	21 dpi
HBVtg	AAV-LUC	5	5
	AAV-HDV	4	5
HBVtg IFN- α / β R KO	AAV-LUC	7	4
	AAV-HDV	6	4

When comparing the viral quasispecies in terms of RHL, no significant differences were observed between the different experimental conditions, although the superinfected (AAV-HDV) mice (both HBVtg and HBVtg x IFN α / β R KO) appeared to show higher values related to the monoinfected (AAV-Luc) mice (Figure 59).

Notably, the median RHL was slightly higher in HDV wild-type mice than in the knock-out (median (Q1-Q3) 11.6 (11.01 – 12.09) for HBVtg HDV mice and 11.46 (8.87 – 13.26) for HBVtg x IFN α / β R KO HDV mice). A small difference was also observed between monoinfected wild-type and knock-out mice, being higher in the first one (median (Q1-Q3) 11.38 (9.67 – 11.63) in HBVtg Luc vs. 10.79 (10.69 – 11.54) HBVtg x IFN α / β R KO Luc).

At 21 dpi, the complexity of the HBV RNA quasispecies in term of RHL at 1% was systematically higher than in 7 dpi. The median of RHL was slightly higher in superinfected than monoinfected mice, both WT and KO for type I IFN signaling (HBVtg HDV median (Q1-Q3): 13.94 (13.61 – 14.34), HBVtg IFN- α / β R KO HDV: 14.12 (13.97 – 14.27), HBVtg Luc: 13.66 (13.47 – 13.79) and HBVtg IFN- α / β R KO Luc: 13.01 (12.79 – 13.53), nevertheless, the differences were extremely limited and without statistical relevance after applying a Kruskal Wallis test.

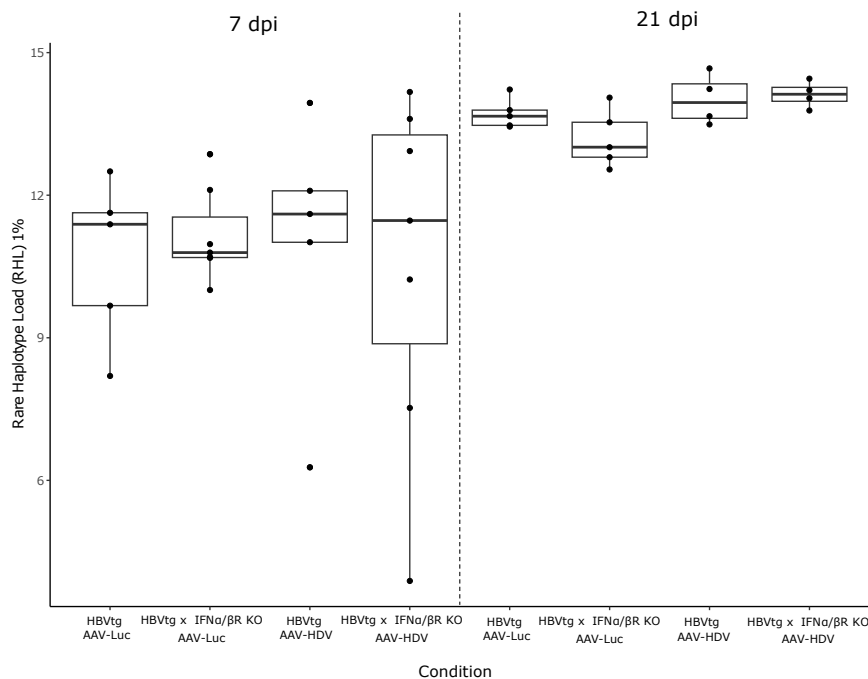


Figure 59. Rare Haplotype Load (RHL) at 1% of the intrahepatic HBV RNA SRT region. The boxplot shows the median and Q1 and Q3 of the RHL per each experimental condition. Each point represents a mouse. On the left RHL at 7 dpi and on the right at 21 dpi. The RHL value per each mouse in the group is also reported with dot. P-value was calculated by applying a Dunn Test for pairwise multiple comparisons with the Holm correction. The Kruskal Wallis test showed a non-significant statistic.

After observing the complexity of the different groups by RHL, we analyzed, at both timepoints, the presence of SNVs in the region corresponding to the HBV genome positions 615 and 629. At 7 dpi, we observed 27 mutated positions in the amplified region (considering single positions involved in mutations occurring in at least two mice), whereas at 21 dpi the number of mutations was reduced to 17 SNVs. At both 7 and 21 dpi, 5 SNVs were detected in all conditions (monoinfected and superinfected mice both WT and KO- Figure 60, panel A and B). Of them, the C628T, G750A and C928T mutations were identified in all conditions at both timepoints (in bold in Figure 60).

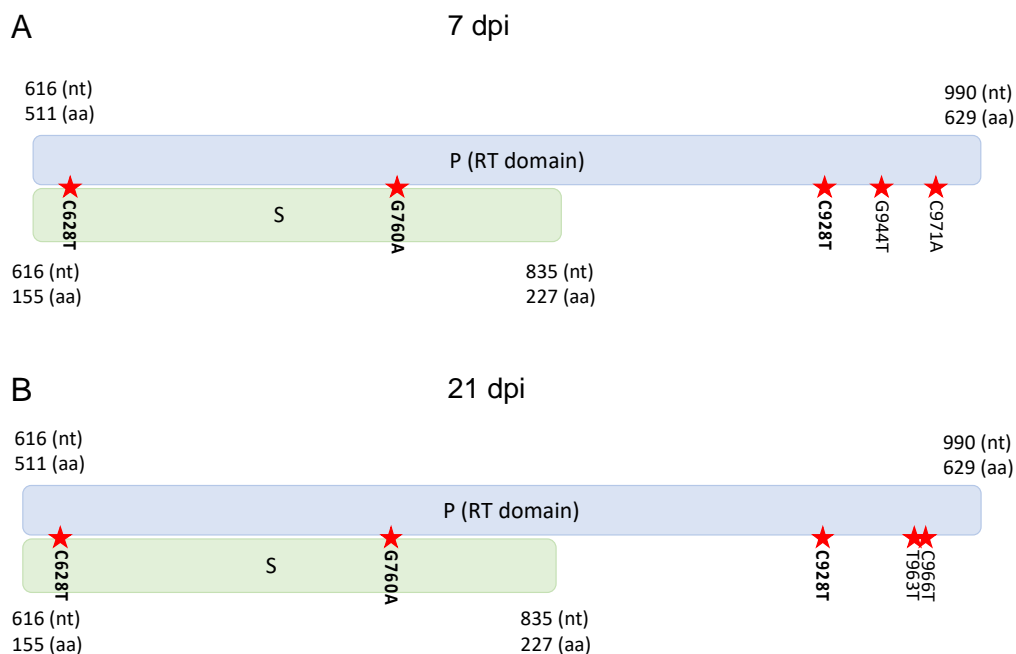


Figure 60. Mutations in the HBV RNA in the region corresponding to the SRT at 7 and 21 dpi. The mutations observed in all the experimental conditions are reported as stars. Their positions are also reported. Those SNVs that interested both 7 and 21 dpi are shown in bold.

Of these SNVs cited above, the variations C628T and G760A might introduce respectively an aminoacidic change (R515W and A559T). Another mutation observed at both timepoints, the C928T mutation, upon translation led to a stop codon at the Polymerase in position 615 (Q615*). Among the other variations that interested just a timepoint, some of them might also introduce aa changes, such as

the G944T (C620F aa substitution) at 7 dpi and the T963G (I626M aa change) at 21dpi.

Among the mutations that not affecting all groups of mice, 22 of them were found at 7 dpi, while 12 were identified at 21 dpi. At 7 dpi, most SNVs were observed at a higher frequency in superinfected mice both WT and IFN- α / β R KO. Some of them clustered together such as the mutations G685A, C919T, C864T, T813C, C910T and T812C in HBVtg HDV, and T711C, G670A, C761T and A834T in HBVtg IFN- α / β R KO HDV mice (Figure 61, panel A).

At 21 dpi, the SNVs were widely and quite randomly distributed in all groups. Some SNVs involved three of the four groups, such as the mutations T698G, C921T and A616C in HBVtg Luc and in HBVtg IFN- α / β R KO Luc mice, the C864T, C866T, C919T and C728T in both IFN α / β R WT and KO HBVtg Luc, and the C971A, C958T, C955T in HBVtg Luc mice but also perceived in HBVtg HDV and in HBVtg IFN- α / β R KO HDV mice (Figure 61, panel B). Notably, some mutations were detected at both timepoints (in bold in the heatmaps of Figure 61).

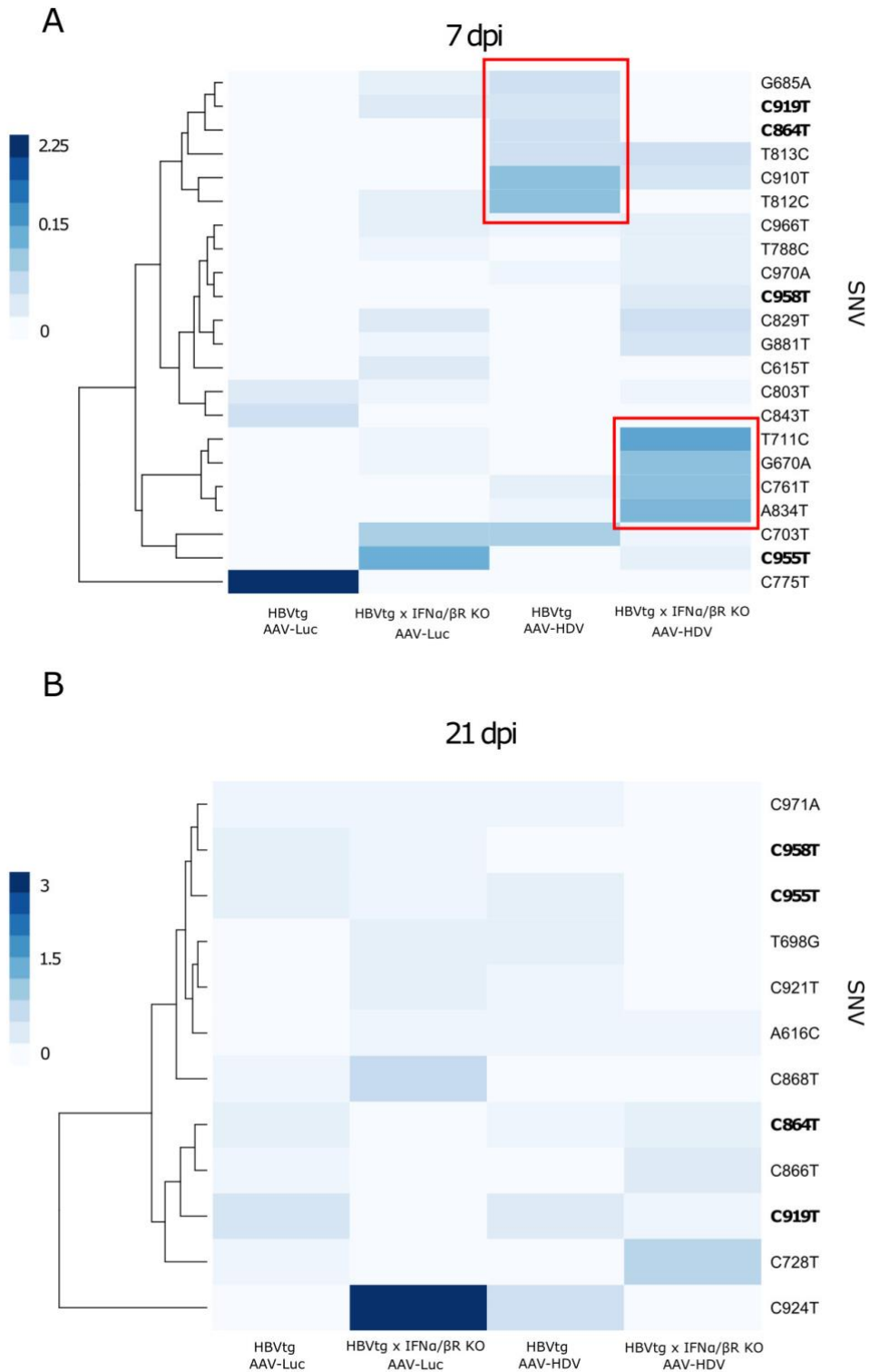


Figure 61. Heatmap of the SNVs observed in the intrahepatic HBV RNA at SRT QS. The heatmap shows the relative frequency of each SNVs observed in the SRT (the darker blue, the highest frequency) related to the different experimental conditions (heatmap's columns) at 7 (panel A) and 21 dpi (panel B). The relative frequency is the result of the average of the frequency of each change considering all the mice belonging to each experimental condition. The position and type of the SNV is reported on the row (on the right), together with the result of applying a dendrogram algorithm (on the left) that shows how the SNVs are related to each other. Mutations observed in both timepoints are reported in bold.

We also analyzed the effects of the above-mentioned mutations regarding amino acid changes in both S and P (RT) ORF (Table 16 and 17). By considering the S ORF, at 7 dpi, we found three non-synonymous substitutions in the final portion of the S protein (P203S, F220L and F220S), all in the transmembrane domain 4 (TM4). All of them presented higher frequency in superinfected mice: the P203S and F220S substitutions in HBVtg IFN- α / β R KO HDV mice, while the F220L in HBVtg HDV. In the RT portion, we identified 8 aminoacidic changes: 7/8 located in the RT palm and one in the thumb domains. Notably, 7/8 changes presented higher frequency in superinfected mice both HBVtg HDV and HBVtg IFN- α / β R KO HDV mice (Table16).

Table 16. The amino acid changes identified in HBV RNA in the SRT region at 7 dpi. This table shows the non-synonymous mutations that give rise to an amino acid change in both S and RT proteins. In addition, the domain in which they are located and the groups that present them are represented (indicating in red the group in which they are most frequently present). The changes that were observed at both 7 and 21dpi are reported in bold. Syn = Synonymous mutations (mutations that interested the third nucleotide of the codon without introducing changes in the encoded amino acid).

HBV SRT mutations at 7dpi			
7 dpi Surface (S) mutations			
SNV (nt)	Post-translational change (Aa)	Structural/functional Domain	Most frequent condition
C761T	P203S	TM4	HBVtg x IFN α / β R HDV
T812C	F220L	TM4	HBVtg HDV
T813C	F220S	TM4	HBVtg x IFN α / β R HDV
7 dpi Polymerase (RT) mutations			
SNV (nt)	Post-translational change (Aa)	Structural/functional Domain	Most frequent condition
G685A	A534T	Palm	HBVtg HDV
C703T	R540C	Palm/ B domain	HBVtg x IFN α / β R Luc
C761T	A559V	Palm/ C domain	HBVtg x IFN α / β R HDV
T812C	P576P	Palm	HBVtg HDV
A834T	L583F	Palm/ E domain	HBVtg x IFN α / β R HDV
C910T	P609S	Palm	HBVtg HDV
C919T	H612Y	Palm	HBVtg HDV
C970A	P629S	Thumb	HBVtg x IFN α / β R HDV

At 21 dpi, we found only a substitution (W182G) in the transmembrane domain 3 (TM3) of the S protein that corresponded to the SNV T698G. In the RT protein we found just 5 amino acid changes, less than those identified at 7dpi. These changes are reported in Table 17, together with the RT domain in which they are located, as well as the group in which they were found at higher frequency. Notably, 3 of these

5 mutations were found at higher frequency in superinfected mice, 2/3 in HBVtg HDV and the remaining in HBVtg IFN- α / β R KO HDV.

Table 17. Most relevant amino acid substitutions identified in the HBV RNA in the SRT region at 21 dpi. This table shows the non-synonymous mutations that give rise to an amino acid change in both S and RT proteins. In addition, the domain in which they are located and the groups that present them are represented (indicating in red the group in which they are most frequently present). The changes that were observed at both 7 and 21dpi are reported in bold. Syn = Synonymous mutations (mutations that interested the third nucleotide of the codon without introducing changes in the encoded amino acid).

HBV SRT mutations at 21 dpi			
21 dpi Surface (S) mutations			
SNV (nt)	Post-translational change (Aa)	Structural/functional Domain	Most frequent condition
T698G	W182G	TM3	HBVtg HDV
21 dpi Polymerase (RT) mutations			
SNV (nt)	Post-translational change (Aa)	Structural/functional Domain	Most frequent condition
A616C	I511L	Finger/ B domain	HBVtg x IFN α / β R HDV
T698G	V538G	Palm	HBVtg HDV
C919T	H612Y	Palm	HBVtg Luc
C955T	L624F	Thumb	HBVtg HDV
C971A	P629H	Thumb	HBVtg x IFN α / β R Luc

To try to determine the source of this variability, the SNVs were grouped considering the type of nucleotide change.

At 7 dpi, a wide variety of types of changes per position were observed (Figure 62, panel A), although the most prevalent were C \rightarrow T transitions, with a total of 33 among all the groups. HBVtg IFN- α / β R KO, both monoinfected and superinfected presented higher number of C to T transition per position than WT mice (10 SNVs in both HBVtg IFN- α / β R KO Luc and HBVtg IFN- α / β R KO HDV KO mice versus 5 and 8 in HBVtg Luc and HBVtg HDV mice, respectively).

At 21 dpi, we generally observed an increase of the number of positions involved in SNV with, again, a predominance of C \rightarrow T changes in all the experimental conditions

(Figure 62, panel B). In this case, these transitions mainly involved HBVtg Luc than the others, although the number of positions with this change were quite similar between the groups (10 positions in HBVtg Luc vs 8, 8, 7 for respectively HBVtg IFN- α / β R KO Luc, HBVtg HDV and HBVtg IFN- α / β R KO HDV).

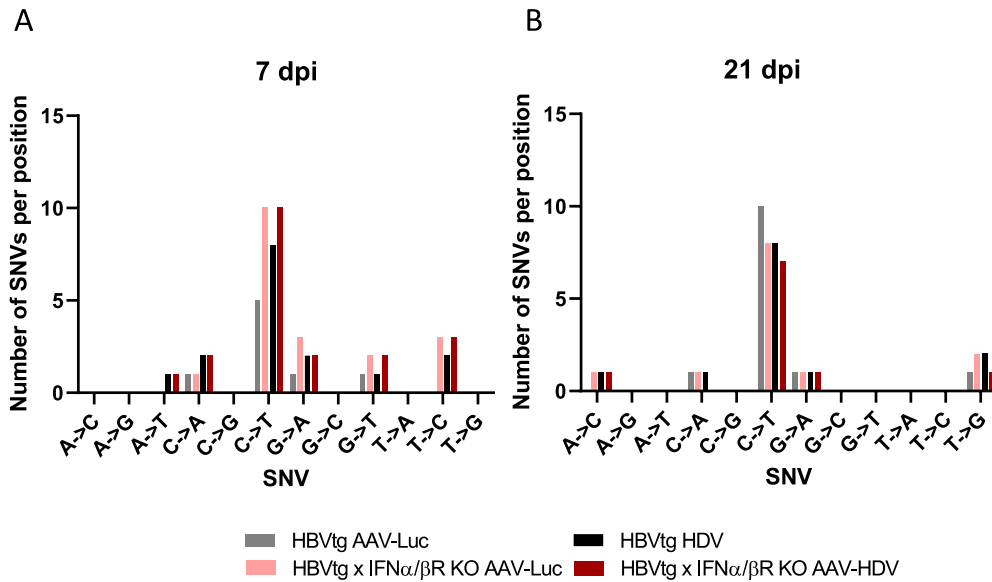


Figure 62. Types of changes in the HBV RNA QS of the SRT. This graph represents the number of the unique mutated positions per each type of mutation at 7 (panel A) and 21 dpi (panel B). Each experimental condition is highlighted by different colors. The different types of variation are reported on the x-axis.

2.3. Analysis of the intrahepatic *HBV X* gene quasispecies at 7 and 21 dpi

Since the *HBX* gene encodes for a key protein of the HBV life cycle (Chapter 1, section 1.2.4) involved in both viral replication and disease progression, we would like to inspect if the activation of type I Interferon (and the consequent activation of the IFN-dependent mutagenic enzymes), in presence of HDV could be a factor increasing HBx variability.

The *HBX* was sequenced from the intrahepatic RNA in two overlapping amplicons (named 5' and 3'-*HBX*). At 7 dpi, a median (Q1-Q3) of 17934.25 (50437-32502.75) of reads were obtained for the 5'-*HBX*, whereas a median (Q1-Q3) of 53521.75 (95739.25-42217.5) of reads were achieved in the 3'-end from a total of 23 mice (the number of mice per each amplicon and experimental condition is reported in Table 18).

As done for the SRT, and as the effect of inhibition of the HBV replication due to HDV infection was higher at later timepoints, the quasispecies of the virus was also analyzed at 21dpi. A median (Q1-Q3) of 46731.5 (63832.25-17100.75) reads were obtained after passing the quality filters for the 5'-*HBX*, while a median (Q1-Q3) of 41412.25 (77884-36471.75) reads were achieved for the 3'-end. The number of mice included in this second part of the study are reported in Table 18.

Table 3. Number of mice used to analysis the quasispecies of the intrahepatic HBV RNA in the HBX at both 7 and 21dpi. The table shows the number of mice per each experimental condition for which HBV RNA was optimally amplified and sequenced per each experimental condition.

Mouse strain	AAV injection (5×10^{10} vg of AAV/mouse)	n (Mice that passed filter)			
		7 dpi		21 dpi	
		5HBX	3HBX	5HBX	3HBX
HBVtg	AAV-LUC	5	4	8	8
	AAV-HDV	4	4	8	8
HBVtg IFN- α / β R KO	AAV-LUC	7	7	9	9
	AAV-HDV	7	5	9	9

As for the SRT, we first studied the quasispecies by considering how much it was enriched of haplotypes at a frequency below 1% (RHL) at both timepoints (Figure 63). At 7dpi, HBVtg mice, both mono and superinfected, presented a median (Q1-Q3) value of RHL in the 5'-HBX of 17.49 (17.23 – 17.56) and 17.92 (17.84 – 17.96), respectively. Notably, the presence of HDV seemed to be a determinant factor in increasing even more the RHL, as shown by the higher RHL value of in HBVtg x IFN α / β R KO HDV mice related to in HBVtg x IFN α / β R KO Luc (median (Q1-Q3) RHL 17.39 (16.81 – 17.62) and 16.28 (16.26 – 16.63) for KO mice, respectively super and monoinfected). Curiously, at 21 dpi QS complexity decreases in all groups (Figure 63), especially in HBVtg HDV with a median (Q1-Q3) of RHL of 11.83 (10.08 – 12.6). Of note, by applying a Kruskal Wallis test, differences between the different experimental groups were not statistically relevant.

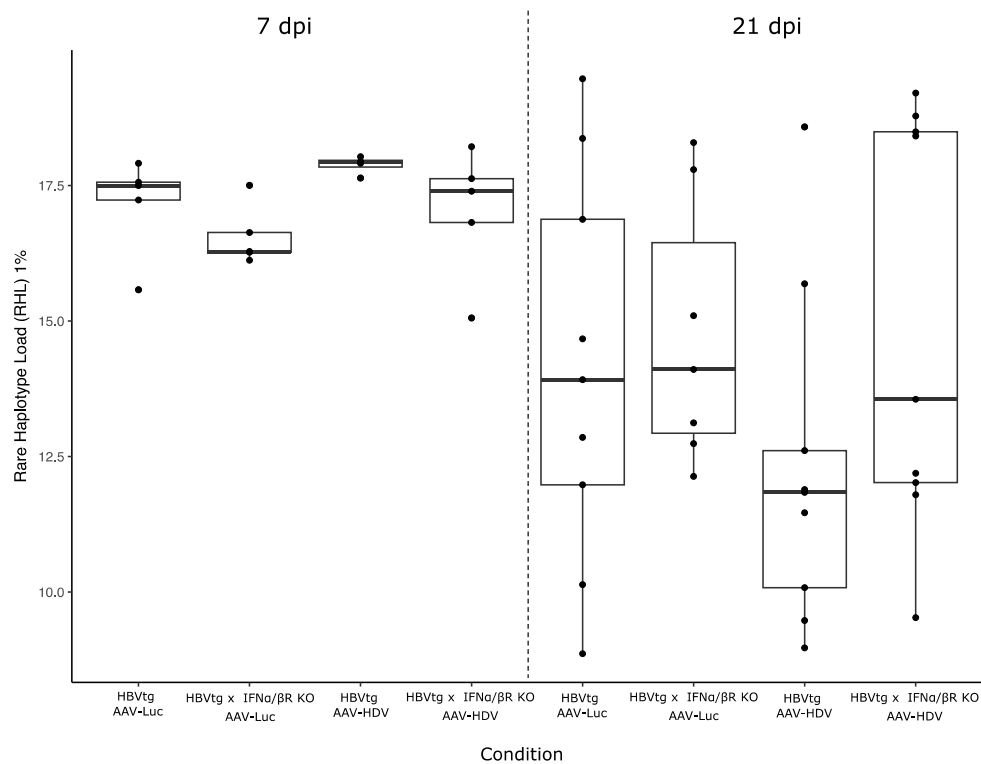


Figure 63. Rare haplotype load of the 5'-HBX quasispecies at 7 and 21 dpi. The boxplot shows the RHL at 1% at 7dpi (left) and 21 dpi (right) of the quasispecies of the HBX 5'-end in the four experimental conditions. The RHL value per each mouse in the group is also reported with dot. P-value was calculated by applying a Dunn Test for pairwise multiple comparisons with the Holm correction. The Kruskal Wallis test showed a non-significant statistic.

In the 3'-HBX region, the RHL was quite similar between the groups at 7dpi. As seen for the 5'-HBX, at 21dpi a reduction in the RHL index for all the groups was detected, without big difference between them (Figure 64). Notably, if we compare both amplicons, at both timepoints the RHL was lower in 3'-HBX than 5'-HBX.

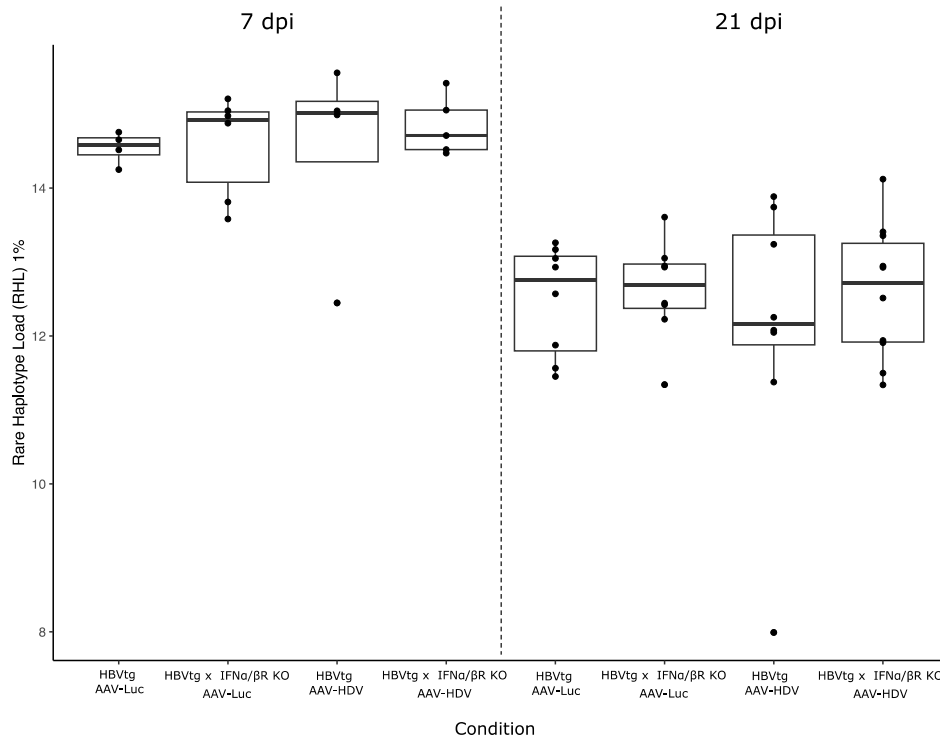


Figure 64. Rare haplotype load of the 3' HBX quasispecies at 7 and 21 dpi. The boxplot shows the RHL at 1% at 7dpi (left) and 21 dpi (right) of the 3' end HBX quasispecies in the four experimental conditions. The RHL value per each mouse in the group is also reported with dot. P-value was calculated by applying a Dunn Test for pairwise multiple comparisons with the Holm correction. The Kruskal Wallis test showed a non-significant statistic.

As previously explained in chapter 5, section 1.2.2, the SNVs were checked by aligning the haplotype's sequences with the master, and to avoid biased results due to the presence of random mutations, just those SNVs that were observed in at least two mice were considered. Of note, the SNVs were analyzed considering the entire gene (5' plus 3'-HBX).

A total of 15 SNVs were identified at 7 dpi, while the number of mutations reached a total number of 35 at 21 dpi. Of these observed changes, at 7 dpi 9/15 were common to all experimental conditions (Figure 65, panel A), and at 21 dpi, 15/35 were common to all four groups of animals (Figure 65, panel B). Interestingly, 7

SNVs were maintained at both timepoints. Of them, 2 fell on the X promoter (C1261T and G1355A) and other two on the core/preCore promoter (G1606A and G1803A). At 21 dpi, we also observed other synonymous mutations that involved the X (C1340T) and the core promoter (C1684T).

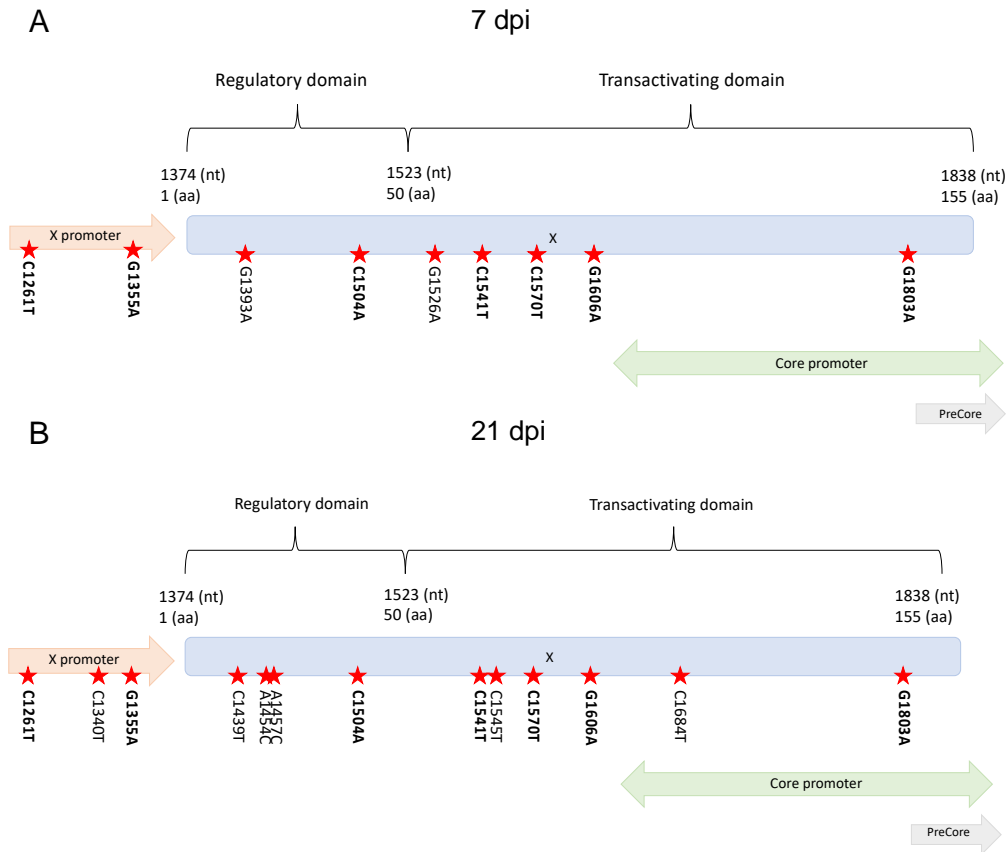


Figure 65. Mutations in X gene at 7 and 21 dpi. In this graph are represented the mutations present in all experimental conditions and in which position of the RT and S gene they are found. The mutations in bold are those identified at both 7 and 21 dpi.

As previously commented, the HBx is a key protein in viral replication since it acts as trans-activation factor of viral transcription. Aminoacidic changes in this protein could potentially impact the efficiency of the viral replication. Moreover, the *HBX* nucleotide sequence includes some genomic regions that are essential for the viral replication (Chapter 1, section 1.2.4). The SNVs leading to post translational modifications were analyzed in depth to inspect the functional effects they might have on viral replication.

Related to the mutations that affected all groups of mice, at 7dpi we found a total of 5 amino acid changes (C7T, A44V, A66V, R78H and A144S, corresponding to SNVs G1393A, C1504T, C1570T, G1606A and G1803A), two in HBx regulatory domain, and the other three in the trans-activating domain. Notably, apart from the C7T substitution (G1393A), all the others were maintained at 21 days after AAV injection. At this last timepoint, two more amino acidic changes were detected, the L58F and S104L, corresponding to the SNVs C1545T and C1684T.

When considering those SNVs that were not shared by all the groups of mice, we identified 6/15 SNVs at 7dpi and 20/35 at 21 dpi. Among all the mutations, just one was observed at both timepoints (C1451T). At 7 dpi, just one mutation (C1439T) was detected in HBVtg IFN- α / β R KO HDV mice, differently from the other groups that shared all the other 5 mutations with different relative frequency (Figure 66, panel A).

At 21 dpi, we detected 18 SNVs, each one differently represented in a group related to the other (Figure 66, panel B) and some of them were forming patterns of mutations. Of note, the most frequent variations were especially observed in superinfected mice, both WT and KO, especially the group of mutations C1912T+C1909T+C1575T and T1346C, T1293C and C1562A, were mainly observed in HBVtg HDV mice, and the pattern C1391T+T1406C+C1562T and C1451T, T1488C, C1408T and C1305T were featured in HBVtg IFN- α / β R KO HDV mice (Red boxes in figure 66, panel B).

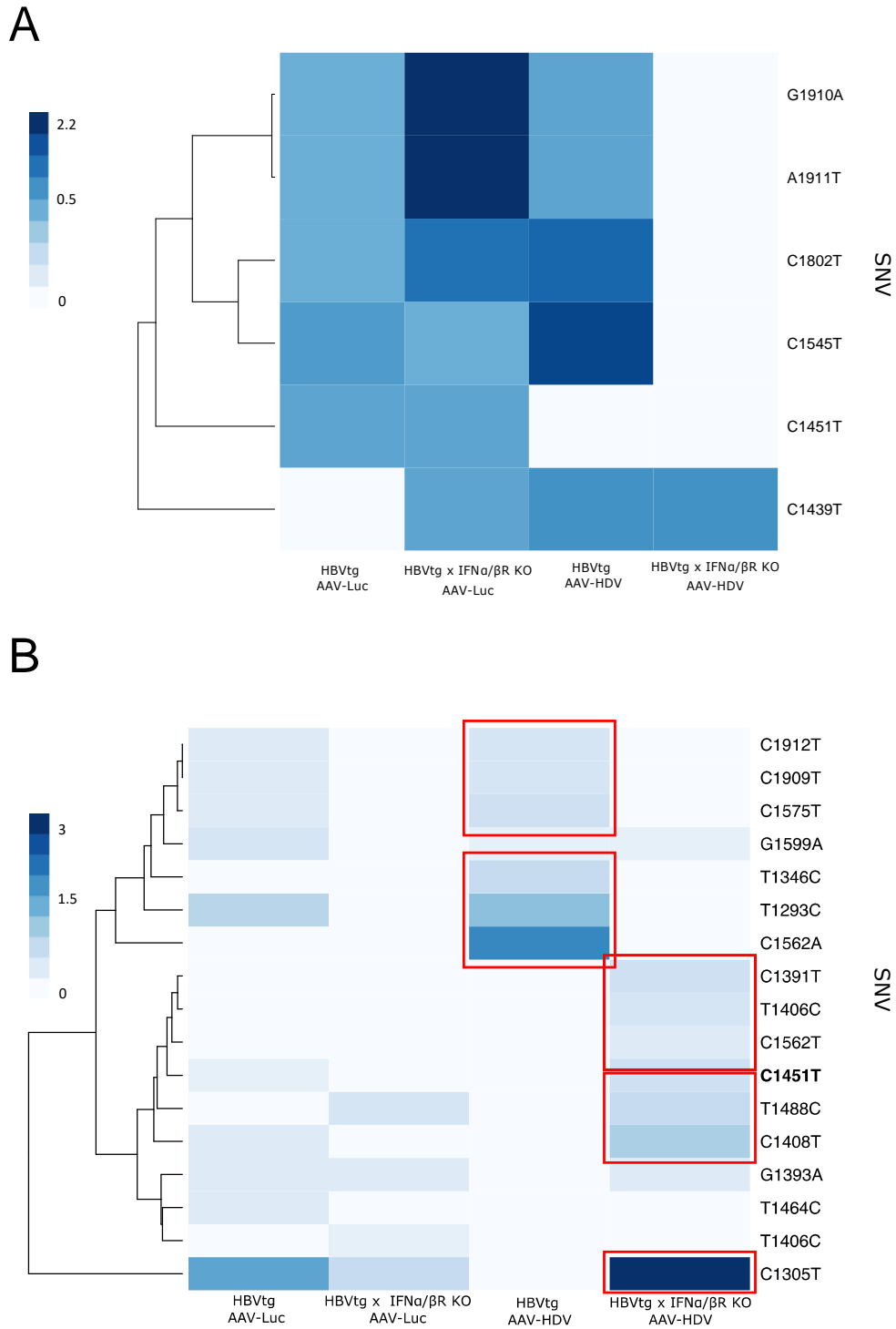


Figure 66. Heatmap of the SNVs observed in HBX at 7 and 21 dpi. The heatmap shows the relative frequency of each SNVs observed in the HBX (the darker blue, the highest frequency) related to the different experimental conditions (heatmap's columns) at 7 (panel A) and 21 dpi (panel B). The relative frequency is the result of the average of the frequency of each change considering all the mice belonging to each experimental condition. The position and type of the SNV is reported on the row (on the right), together with the result of applying a dendrogram algorithm (on the left) that shows how the SNVs are related to each other's. The red box shows the most important SNVs in the superinfected groups.

To inspect if the above-mentioned mutations might have a functional impact into the viral protein activity, the introduction of amino acid substitution was also studied at both timepoints (Tables 19 and 20). Moreover, the presence of synonymous variations that could affect functional or regulatory domains of the HBV expression was also reported.

Among the SNVs at 7 dpi, all were synonymous mutations, except for the C1545T that introduced a change from lysin to phenylalanine in position 58 (L58F) and involved the transactivating domain of the protein. This change was mainly observed in HBVtg HDV mice. the other three synonymous mutations were detected in core promoter (C1802T, G1910A and A1911T), las two of them also involved preCore promoter and had higher frequencies in HBVtg IFN- α / β R KO Luc mice (Table 19).

Table 19. Most relevant mutations identified in HBV RNA in the X region at 7 dpi. This table shows the mutations present in the HBV promoters as well as the non-synonymous mutations that give rise to an amino acid change in the region. Corresponding with the full HBX. In addition, the domain in which they are located and the groups that present higher relative frequencies are reported. Syn = Synonymous mutations (mutations that interested the third nucleotide of the codon without introducing changes in the encoded amino acid).

7dpi X mutations (HBX)

SNV (nt)	Post-translational change (Aa)	Structural/functional Domain	Most frequent condition
C1545T	L58F	Transactivating domain	HBVtg HDV
C1802T	Syn	Transactivating domain, CCAAT/enhancer-binding protein- α (C/EBP α) domain, Core promoter	HBVtg HDV
G1910A	Syn	preCore, Core promoter	HBVtg x IFN α / β R Luc
A1911T	Syn		HBVtg x IFN α / β R Luc

At 21 dpi, most mutations were synonymous, so did not change the encoded amino acid. However, 8 of the 18 could have a functional impact in HBV replication and 7 showed higher frequency in HBVtg HDV mice. All these changes are reported in Table 20.

Of them, four mutations were synonymous and two were positioned in the X promoter (T1293C and T1346C) and two in the preCore and Core promoter (C1909T and C1912T). The other 4 SNVs that might have an impact in HBV replication introduced amino acid changes. Of note, three of them had a higher frequency in HBVtg HDV mice (F63L, P68S and L89F), and took place in the transactivating domain of the proteins.

Table 20. Most relevant mutations identified in HBV RNA in the X region at 21 dpi. This table shows the mutations present in the HBV promoters as well as the non-synonymous mutations that give rise to an amino acid change. In addition, the domain in which they are located and the groups that present them are represented (indicating in red the group in which they are most frequently present). Syn: Synonymous mutations.

21 dpi X mutations (HBX)

SNV (nt)	Post-translational change (Aa)	Structural/functional Domain	Most frequent condition
T1293C	Syn	X promoter	HBVtg HDV
T1346C	Syn		HBVtg HDV
C1562A	F63L	Transactivating domain	HBVtg HDV
C1575T	P68S		HBVtg HDV
G1599A	A76T	Transactivating domain, CCAAT/enhancer-binding protein- α (C/EBP α) domain, DR2	HBVtg Luc
G1640T	L89F	Transactivating domain, CCAAT/enhancer-binding, protein- α (C/EBP α) domain, Enh II, Core promoter	HBVtg HDV
C1909T	Syn	preCore, Core promoter	HBVtg HDV
C1912T	Syn		HBVtg HDV

As done with the SRT region, and to inspect the possible source of variations, we also analyzed the type of changes. At this step, the HBX was again considered in its entirety.

At 7dpi, the most prevalent changes in the *HBX RNA* quasispecies were C \rightarrow T (with a total of 29 changes) and G \rightarrow A (24 total mutations) transitions. Curiously, the number of positions with G to A mutations was the same in all the groups, similarly to the C \rightarrow Ts, apart from the HBVtg x IFN α / β R KO HDV mice that had only 5 changes (Figure 67, panel A). Of note, changes that interested the portion of the amplicons that overlaps between 5' and 3'-end were referred once.

At 21dpi we observed a total of 45 SNVs per position in the full *HBX* region. When analyzing the type of the variations, we observed that most of them were again C→T and G→A transitions (52 positions mutated for C→T and 26 for G→A changes) (Figure 67, panel B), however we didn't observe any statistical difference between the groups.

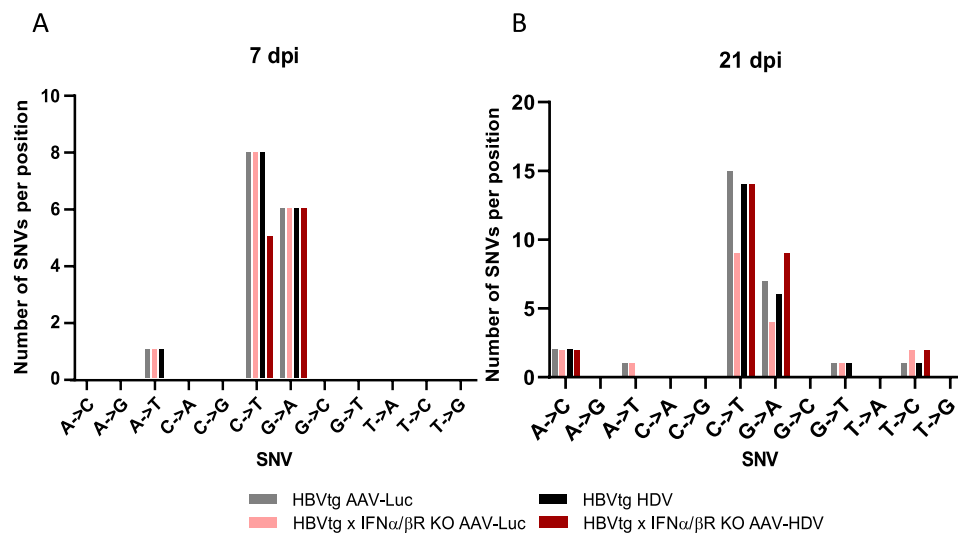


Figure 67. Types of SNVs in the HBX quasispecies at 7 and 21 dpi from intrahepatic HBV RNA. The graph shows the number of the total unique positions involved on each type of change along all the HBX sequence (both 5'- and 3'-HBX) at 7 (panel A) and 21 dpi (panel B). The different experimental conditions are represented by colors.

2.4. INF-I dependent HDV variability

As extensively described in the introduction (Chapter 1, section 3), HDV takes advantages of the activation of the IFN signaling, and the IFN-related expression of ADAR1 to edit *HDAG* mRNA to produce the large isoform of the protein. The proportion of the mutated codon defines the percentage of editing of HDV genomes. To inspect the role of type I interferon in HDV variability, we studied the intrahepatic HDV quasispecies at the two timepoints.

At 7dpi, the quasispecies of the 5'-*HDAG* were optimally amplified and analyzed by NGS in 4 HBVtg HDV-and 6 HBVtg IFN- α / β R KO HDV mice (Table 21). After passing the different quality filters, a median (Q1-Q3) of 7324.25 (20836.75-23512) reads were obtained.

To inspect deeply the effect of IFN signaling knock-out on HDAG variability, we analyzed also the 5'-*HDAG* from the intrahepatic HBV RNA at 21dpi. For both conditions, 6 mice passed the filter and were included in the study, with a median (Q1-Q3) of 21316 (5112.5-26428.5) reads.

Table 21. Scheme of the number of mice used for studying HDAG quasispecies at 7 and 21 dpi.

Mouse strain	AAV injection (5×10^{10} vg of AAV/mouse)	n (Mice that passed filter)	
		7 dpi	21 dpi
HBVtg	AAV-HDV	4	6
HBVtg IFN- α / β R KO	AAV-HDV	6	6

In terms of complexity, at both 7dpi and 21dpi, the complexity of intrahepatic HDV RNA in the 5'-*HDAG* tended to be lower in KO than in WT, especially at 21dpi, although the difference was not statistically relevant with a median (Q1-Q3) RHL at 1% of 24.19 (22.86 – 25.33) and 23.03 (21.92 – 23.89) for respectively HBVtg HDV and in HBVtg x IFN α / β R KO HDV mice at 7dpi, and median (Q1-Q3) of 25.4 (22.91 – 27.11) and 22.96 (19.27 -24.78) at 21dpi (Figure 68).

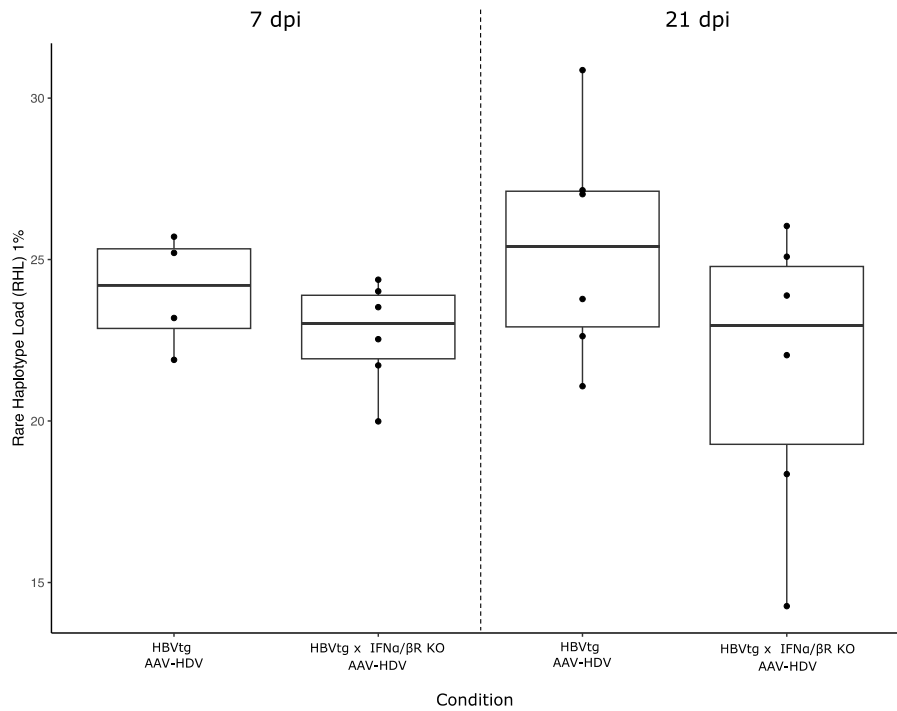


Figure 68. Rare haplotype load of the 5' HDAG quasispecies in HBVtg HDV mice at 7 and 21 dpi. The boxplot shows the RHL at 1% at 7dpi (left) and 21 dpi (right) of the 5' HDAG quasispecies in the two experimental conditions. The RHL value per each mouse in the group is also reported with dot. P-value was calculated by applying a Dunn Test for pairwise multiple comparisons with the Holm correction. The Kruskal Wallis test showed a non-significant statistic.

As explained in methods (Chapter 4, section 1.3.2.), the amplicon encompasses the stop codon for the S-HDAg thus allowing the quantification of the percentage of editing (the percentage of mutated mRNA, from the TAG to TGG, that allows the transduction of the L-HDAg) in the viral QS.

At 7dpi, the percentage of edited genome (deamination of the adenosine in codon 196) was quite similar between HBVtg HDV and in HBVtg x IFN α / β R KO HDV mice ($6.93\% \pm 1.87\%$ and $5.86\% \pm 2.07\%$ respectively) (Figure 69). Regarding the editing of HDV genome at 21 dpi, as expected, we observed a relevant statistical difference in the percentage of edited genome between HBVtg HDV (around 13% of the edited genome) and HBVtg x IFN α / β R KO HDV (8% of the genomes were edited).

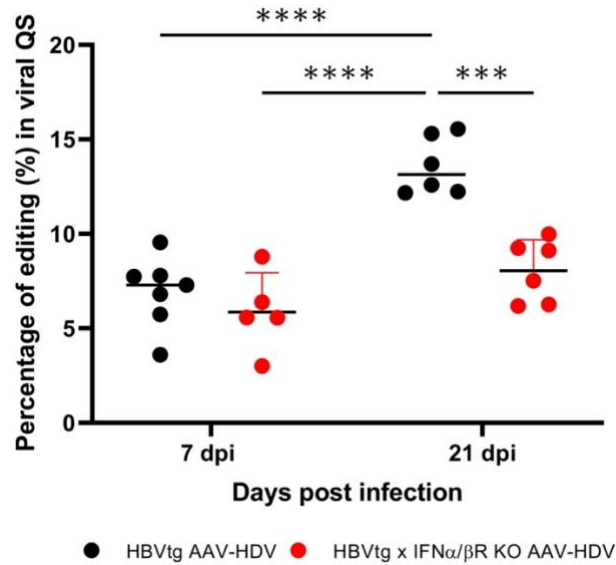


Figure 69. Editing of the HDV RNA in superinfected mice at 7 and 21dpi. The percentage of edited genome in viral QS was calculated by summing the relative frequencies of all the haplotypes carrying the edited adenosine in each mouse for both HBVtg HDV and HBVtg IFN α / β R KO HDV. All the data presented a normal distribution analysed by the Shapiro Wilk test (p value >0.05). P-values are reported as asterisks (* for ≤ 0.05 , ** for ≤ 0.01 , *** for ≤ 0.001 and **** for ≤ 0.0001) and calculated by applying an Anova test with a Tukey for multiple comparisons.

Analyzing the SNVs observed in this region at an early infection (7 dpi), a total of 64 SNVs were obtained. We also compared the detected SNVs between both groups at a late infection (21 dpi) when just 25 mutations were detected.

At 7 dpi, among all the SNVs, the 65% (42/64) were found in all experimental conditions, while at 21 dpi, the percentage of mutations shared by all the groups was 76% (19/25) (Figure 70, panel A for 7dpi and panel B for 21dpi). Notably, 10 SNVs (A341G, T344C, C390T, T397C, A408G, A418G, A432G, A458G, G497A and G555A-represented in bold in Figure 70) were identified at both timepoints.

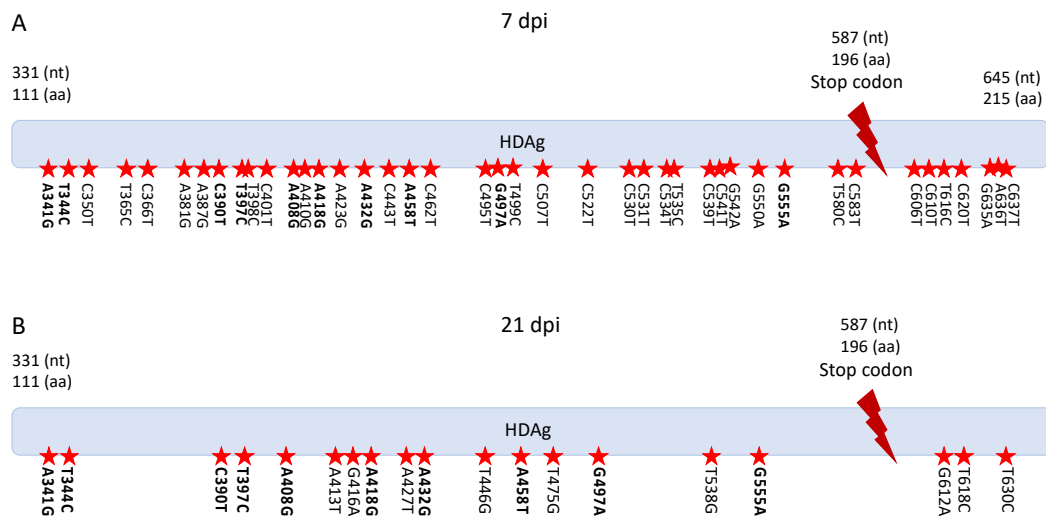


Figure 70. Mutations in 5'HDAG gene at 7 and 21 dpi. In this graph are represented the mutations present in all experimental conditions and in which position of the 5'HDAG gene they are found. The mutations in bold are those identified at both 7 and 21 dpi.

Of the 64 SNVs observed at 7 dpi, 22 of 64 showed a different incidence between superinfected WT and KO mice and formed pattern of mutations (Figure 71, panel A). Ten mutations were mainly observed in WT mice (G569A+A427G+G416A, C595T, G604A+G449A+G517A, A422G+T580A, G412A), whereas the other 12 especially concerned KO mice (C485T+C584T+C532T+C638T+C605T, C602T+C438T+C634T+A565G+A362T, G526A and T630C).

As previously commented, at 21 dpi just the 6 SNVs didn't involve both groups. Notably, 5 of 6 mutations were found in WT mice: A387G+T499C, A483G+A381G and T535C. The only variation spotted in both timepoints was the T580A, whose frequency was higher in WT than KO mice at 7dpi and vice-versa at 21dpi (Figure 72, panel B).

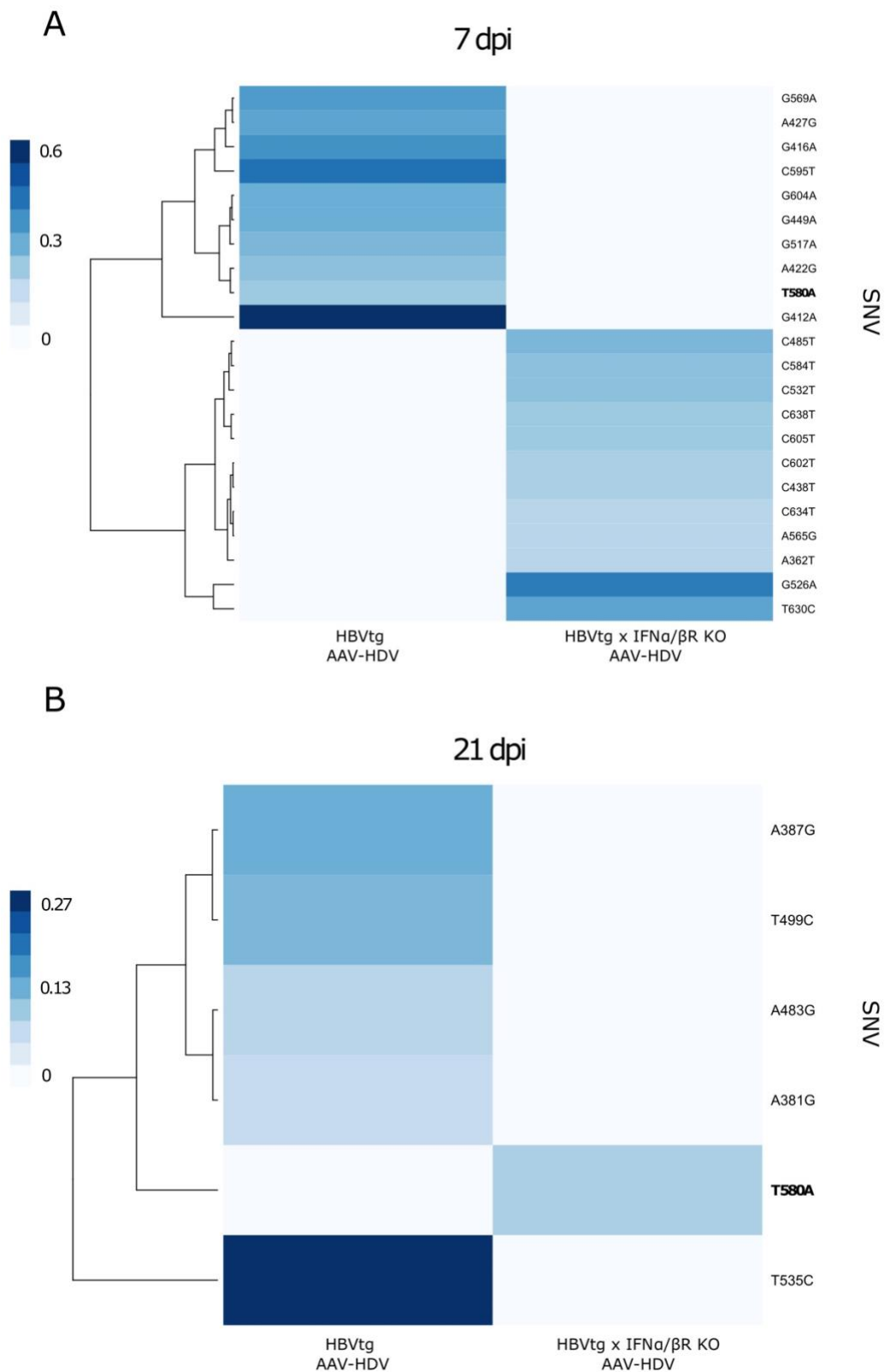


Figure 72. Heatmap of the SNVs observed in the intrahepatic HDV RNA at 5'HDAG QS. The heatmap shows the relative frequency of each SNVs observed in the 5'HDAG (the darker blue, the highest frequency) related to the different experimental conditions (heatmap's columns) at 7 (panel A) and 21 dpi (panel B). The relative frequency is the result of the average of the frequency of each change considering all the mice belonging to each experimental condition. The position and type of the SNV is reported on the row (on the right), together with the result of applying a dendrogram algorithm (on the left) that shows how the SNVs are related to each other's. Mutations observed in both timepoints are reported in bold.

Regarding the types of changes observed in the sequence at an early timepoint, most of them were again C→T transitions (23 and 31 for wild-type and knock-out mice, respectively), with a prevalence in HBVtg x IFN α / β R KO HDV mice. Of note, other types of change such as A → G, G→A and T→C changes were also detected (Figure 72, panel A).

At late timepoint (21dpi), the number of unique positions involved in changes was, as expected, lower. The incidence of C→T transitions was strongly reduced, whereas changes as A→G and T→C were increased (Figure 72, panel B). In most of them, the number of SNV per position was similar between both groups, excepting of the transition A→G that involved 9 unique positions in HBVtg HDV versus 6 in HBVtg x IFN α / β R KO HDV mice. It should be pointed out that, looking the prevalence of no-shared mutations between both groups, at 21dpi all the mutations observed in HBVtg HDV mice were A to G or T to C transitions, which are typically associated with ADAR1 activity.

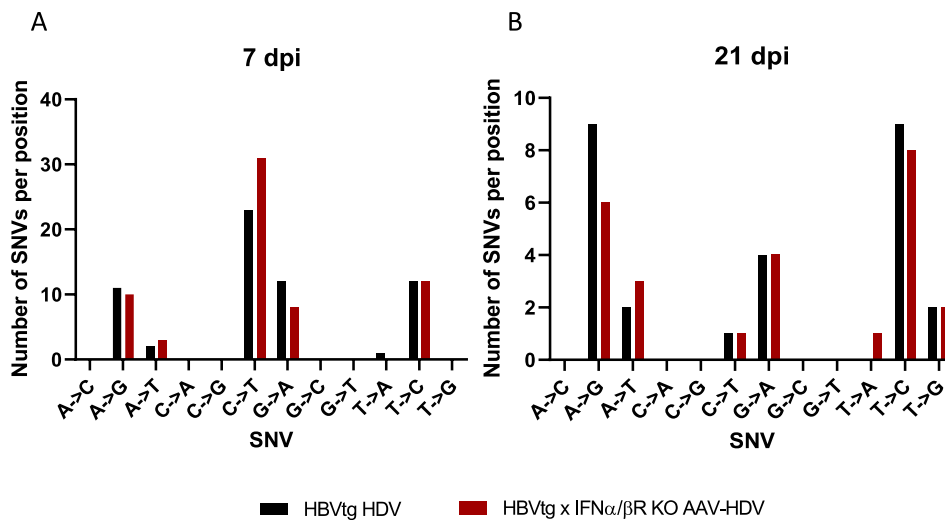


Figure 71. Types of SNVs in the intrahepatic HDV RNA in 5'HDAG quasispecies at 7 and 21 dpi. The graph shows the number of the total unique positions involved on each type of change along all the 5'HDAG sequence) at 7 (panel A) and 21 dpi (panel B). The different experimental conditions are represented by colors.

Discussion

Study 1. Inspecting HDV ribozyme conservation and comparison of quasispecies variability and evolution between ribozyme and 5'-HDAG.

HDV variability results in eight genotypes worldwide distributed with 20-40% divergence between them and up to 12% divergence within genotypes. Despite the hyper-conservation of the ribozyme, HDV is an extremely variable virus as previously described in Chapter 1, section 4.2, presenting an evolution rate between $9.5E-03$ to $1.2E-03$ substitutions/site/year [270].

The first part of this thesis has been focused on HDV variability in a highly functional region of the HDV genome, the ribozyme. As part of the “RNA world” hypothesis, which suggests that life on Earth began with a single RNA molecule, the sequences acting as ribozymes are considered to be an ancient system for transferring genetic information and catalyzing biochemical reactions [305]. This non-coding portion of the HDV RNA plays a key role in viral replication, as it is essential for the release of the genome and antigenome monomers from the concatemer formed during the rolling circle amplification of the HDV RNA. Based on this assumption, the HDV ribozyme may be an optimal target for new therapeutic strategies.

Due to the lack of viral polymerase, there is no specific antiviral therapy for HDV infection as previously commented in the introductory chapter (Chapter 1, section 2.5). The activation of the immune response through the pegylated IFN α has been the main antiviral strategy. However, only 25-40% of the treated patients reach a suppression of the HDV RNA after 1-2 years of treatment [306]. Furthermore, this suppression was not sustained, as reported in a long follow-up study (up to 4 years) after interferon interruption, which showed that about half of the patients who tested HDV RNA negative 6 months after interruption were HDV patients at the last follow-up visit [307]. Therefore, new treatment options against HDV, such as inhibition of HBsAg release, prenylation or viral entry, are currently under investigation [169]. Among them, the entry inhibitor Bulevirtide (BLV) has recently been approved for use alone or in combination with peg-IFN α in CHD patients with compensated liver disease [308,309]. Also, a recently published study showed the

first case of possible HDV cure in a cirrhotic patients treated with BLV monotherapy. However, none of these new therapeutic tools target the HDV life cycle. Although BLV seems to be a very promising drug in HDV treatment, since it blocks the further spread of the infection, it might not perturb the intracellular replication of the viral genome and the expression of delta antigen, especially in highly-viremic CHD patients who show few signs of liver damage [310].

Notably, a gene therapy approach may be a valuable strategy to promote HDV RNA elimination and block disease progression. Due to their intranuclear location, the genome and antigenome seem to be resistant to interfering RNA (siRNA) activity [311], whose silencing activity is developed in the cell cytoplasm. However, HDV genomes can be targeted by antisense oligonucleotides (ASOs), that are active within the nucleus [312]. Although inhibition of the expression of HBsAg through a siRNA may interfere with HDV infection [313], a combination of silencing molecules targeting both viruses may be a highly valuable therapeutic strategy. Unlike HBV, no HDV-specific silencing molecules have been reported to date [130]. The extreme variability of HDV RNA makes the design of an effective antisense oligonucleotide very difficult. To this end, the identification of highly conserved regions in the HDV genome to use as targets for gene silencing may be essential to obtain a valuable strategy regardless of the extreme variability of the virus.

Notably, the HDV genome has been mainly inspected using Sanger sequencing [314,315], which limits the study to the consensus sequence. Deep sequencing analysis is needed to have a full perspective of the viral conservation and variability. The first study included in this thesis aims to analyze hepatitis delta QS by next-generation sequencing. Of note, very few studies have focused on using this technique to study HDV QS.

In this study two longitudinal samples of 25 CHD patients were included and followed up for an average of 2.25 years. HDV was optimally extracted and sequenced from 19 patients. To ensure optimal quality of the sequencing, different quality filters (as reported in Chapter 4, section 1.4.1) were applied, thus resulting

in an average of 4550.5 reads per patient. This enables an extremely deep analysis of HDV QS in the ribozyme region.

The first step was to study QS conservation. As expected, the ribozyme was overall highly conserved (85% of nucleotides presented 100% of conservation), with a hyper-conserved region encompassing positions 715–745 (93.6% of this region presented 100% of conservation). This portion of HDV RNA involves the P1 domain, which is essential for ribozyme activity since the ribozyme cleaves at the base of the P1 helix [172].

To evaluate the evolution of ribozyme QS during follow-up, the two follow-up samples per patient were paired. Notably, when considering sequence distance or complexity indexes, no differences were observed between the two longitudinal samples, suggesting that the QS in this portion of the genome did not vary over time. This is consistent with the extremely important role that this non-coding RNA portion plays in viral replication.

However, a variable region between nt 739–769 was also identified (around 20% of nucleotides presented a conservation between 1.2 and 1.5 bits), although changes were observed in just a few nucleotide positions (specifically in six positions). This variable region falls into the P4 structural domain as observed in a previous study that reported a higher mutation level for both genomic and antigenomic RNA in this region [215]. In fact, the P4 domain is mainly involved in structural stability rather than self-cleavage activity. Studies have reported that the removal of the P4 domain doesn't inhibit ribozyme activity [179,182]. Although this domain is not directly involved in the self-cleavage process, it has been reported that the conservation of the nucleotide sequence is essential for a stable base pairing, which is a necessary condition for efficient self-cleavage activity [316].

A recently reported study suggested that the conservation of the sequence is a prerequisite to guarantee ribozyme activity under co-transcriptional conditions [317]. As expected, some positions are essential for the cleavage activity such as the uridine in position 23 of the antigenomic RNA, which forms a cross-link with the P1

stem, thus promoting the docking with the catalytic site. Mutations at this site, such as the U23A, can disrupt this cross-linking and potentially alter ribozyme folding [318]. Moreover, the C24 and 25 are involved in the formation of the P1.1 which appears to be involved in the trapping of the ribozyme cleavage site into the catalytic domain [318]. Another study published by Tanner et al. in 1994 proposed a model by introducing mutations that suggested that thymine 20, cytosine 21, and cytosine 75 in the HDV genome were close to the cleavable phosphate and were involved in coordinating the Mg²⁺ ion and cleavage catalysis [319]. However, mutations in the ribozyme could affect the cleavage activity differently [320].

Of note, the analysis of the full HDV genome in CHD patients by using 6 overlapping amplicons showed that several mutations can be found in the ribozyme sequence [215]. Here we identified a total of 48 mutations, 12 of which were found in at least six patients. The identification of mutations in such a conserved and active region of the HDV genome may suggest that they do not affect viral fitness. As previously reported, the observed mutations mainly involved the P4 domain (ribozyme positions 60, 61, 62, 64, 65 and 69) [315]. Of these 12 mutations, three presented a relative frequency of 1% or more and were maintained between the two follow-up samples. To assess whether these changes could impact viral replication, they were tested *in vitro*. These changes were the T23C in the L3 domain, and the C64d and T69C, both in the P4 domain.

The mutation T23C has been previously reported and the *in vitro* testing of the ribozyme self-cleavage activity in the presence of the T23C mutations showed a 10³-fold decrease in cleavage rate [320]. This mutation falls into the L3 domain, which is an important element in cleavage activity [182,313]. Notably, when we tested this mutation *in vitro*, we observed that it had a limited impact on viral replication. This is consistent with previously reported studies showing that this change might be less effective in interfering with ribozyme activity than others (e.g., at positions 20, 21, and 25) [315].

The mutations at positions 64 and 69, differently involved the P4 domain. The hyper-variable region observed in HDV QS also interested this domain. Higher

mutation levels in this domain have been previously reported [315]. As previously mentioned, this domain shows a stabilizing function of the ribozyme structure, and it is not directly involved in RNA catalysis. Although the presence of the T69C change did not affect viral replication, the deletion at position 64 strongly reduced HDV replication by more than 1 log IU/mL. Considering the non-essentiality of the P4 domain, the presence of this 1nt-long deletion in this portion could probably alter the secondary structure of the ribozyme and thus affect viral replication. Furthermore, this change was observed at a relatively high frequency (about 50%). The fact that it was observed and maintained at such a high frequency in viral QS suggests that this mutation may promote viral persistence, as has been reported for other viruses such as hepatitis C [321], and human immunodeficiency viruses [322].

In contrast to the ribozyme, the HDAG variability has been studied more thoroughly. The mutation rate for this region has been estimated to be around 1.4×10^{-2} and 1.5×10^{-3} substitutions/site/year [323]. This rate was also confirmed by a study published by our group in 2016, where the 5' extreme of the HDAG was analyzed by NGS (454 GS junior system), thus revealing a mutation rate of 9.5×10^{-3} to 1.2×10^{-3} substitutions/site/year [270]. Of note, if HDV acquires mutations mainly due to the lack of fidelity of the cellular RNA polymerase (as reported in Chapter 1, section 4.2), the ability of the viral genome to acquire mutations should be similar along the entire sequence. To evaluate this, we compared the QS and its evolution among the two functional portions of the viral genome, the ribozyme and the *HDAG* (at the 5'-end extreme). To date, few studies have analyzed HDV QS by NGS [324,325].

As expected, also considering the high degree of conservation of the ribozyme, and consistent with the role of the ribozyme in the HDV replication, the viral quasispecies were less complex and variable in this region related to the 5'-HDAG, especially when considering both complexity (Gini Simpson index and Shannon entropy) and diversity indexes (Mutation frequency and number of polymorphic sites). Notably, all indexes did not change during follow-up, suggesting that the QS remained quite stable in a state of stable HDV viremia.

To inspect and compare the acquisition of mutations in the two different regions, we calculated the evolution rate by considering the number of polymorphic sites normalized to the amplicon length and we related it to the years of follow-up. However, when we considered the evolution rate, we did not observe a significant difference between the different amplicons. This result would suggest that the acquisition of mutations *per se* is an event that is independent of the viral genomic region analyzed.

Interestingly, the evolution rate was inversely correlated with the years of follow-up, suggesting that a process of selection of the variants occurred over time. Notably, when the two amplicons were considered separately, this correlation was maintained and was stronger when considering the 5'-*HDAG*. In this region, a lower rate of evolution was observed at longer follow-up times. This is consistent with previous results published by our group, where an exponential decay with the time elapsed between samples was observed when analyzing the same amplicon during longer follow-up [270]. Of note, the *HDAG* quasispecies is subject to positive and/or negative selection processes as shown by the identification of several escape mutations in the HDAG protein, which are also associated with specific polymorphisms of the human leukocyte antigen complex (HLA) [216,326]. This negative relationship observed between the HDAG QS and the time might suggest that multiple selection processes occurred, resulting in a reduction in the evolution rate.

In contrast, ribozyme variability seems to be more stable over time, probably because the presence of variants in this portion could be mostly associated with fitness-based selection processes, and this region could tolerate mutations that have none or a limited impact on viral replication [327].

Study 2. In vivo study of the involvement of HDV in HBV variability: the role of type I interferon pathway.

As reported in the introduction, HBV is known as a "stealth" virus since it doesn't trigger an intense intracellular antiviral response. Although the hepatocytes express a broad range of PRRs (Pattern Recognition Receptors), studies suggest that HBV can escape this step because the rcDNA is released directly from the nucleocapsid into the nucleus and because the pgRNA (which presents an unusual stem-loop sequence as a packaging signal that could potentially be recognized by the PRR) is rapidly encapsidated. HDV, on the other hand, is a virus that markedly activates the intracellular innate immune response, and even takes advantage of this to produce the L-HDAg. HBV and HDV co-infection experiments conducted on humanized mice reported that HDV strongly activated the innate immunity as shown by the increased expression of several IFN-stimulated genes (ISGs), cytokines, and chemokines such as hIFN β and hIP10, together with the increased of HDV viral load [238]. Notably, in presence of HDV, an inhibition of HBV has been observed, and approximately 64% of HBV/HDV patients present a decrease in the replication of HBV and its intermediates [228]. These data have also been corroborated by in vitro experiments, liver biopsies, or humanized mice models [153,229–233]. The mechanism by which this inhibition is generated is still unclear and several possible scenarios have been suggested [328].

The upregulation of the IFN pathway in the context of HBV/HDV infection could potentially impact the replication of HBV. It has been reported that treatment with IFN α (which was the main therapeutic strategy for HBV years ago) could modify the epigenetic regulation of the cccDNA and thus affecting its transcription [329]. Several IFN α -induced proteins could be involved in such transcriptional control, such as signal transducer and activator of transcription 1 (STAT1), structural maintenance of chromosomes flexible hinge domain containing 1 (SMCHD1), or promyelocytic leukemia (PML), as shown by the restoration (even partial) of the cccDNA expression after downregulation of these proteins [330]. A study published by Dr Lucifora *et al.* reported that treatment with IFN α or the activation of the LTR β

could upregulate the expression of nuclear APOBEC (A3A and A3B), which, thanks to the interaction with the intranuclear HBcAg, could promote the deamination of the cccDNA (when it is single-stranded, during transcription), thus introducing uracils and double-stranded breaks into the DNA sequence that are recognized as errors, thus promoting cccDNA degradation [331]. Furthermore, a transcriptomic study suggested that this mechanism involves another ISG, the ISG20, an exonuclease that is observed mainly in acute and self-limiting HBV infection rather than in chronic hepatitis [332]. Nevertheless, the same group demonstrated *in vitro* that the presence of HDV and the consequent stimulation of the IFN pathway allowed the reduction of HBV expression (in terms of HBeAg, total viral DNA and pgRNA) without sensibly affecting the cccDNA pool [234], highlighting that the above-mentioned effects on cccDNA degradation are associated with higher concentrations of IFN- α , such as those resulting from treatment. Even then, IFN- α treatment leads to HBV DNA suppression in only 30-40% of the HBeAg-positive and about 20-40% of HBeAg-negative viremic patients, promoting HBsAg loss in only 10% of the patients treated with PegIFN- α and/or NUC administration after 5-year follow-up [333].

In addition to the effect on cccDNA suppression and/or degradation, the activation of APOBEC3 could introduce mutations into the HBV rcDNA. In HIV infection, it was reported that APOBEC3G (A3G) could introduce sublethal mutations in the cDNA during retrotranscription [334], thus contributing to HIV variability. Of note, both HIV and HBV replicate through a retrotranscription step. A similar effect of A3G has been observed in HBV infection, with the introduction of C \rightarrow T changes in the rcDNA during the retrotranscription step [335]. This editing might have important clinical implications. For example, an upregulation of APOBECs family members (especially A3G) has been detected in viral hepatitis-related cirrhotic patients [336], which could introduce hypermutations and contribute to viral evolution and potentially immune escape.

Moreover, HDV uses the ADAR1 enzyme to change the Amber/w codon and produce the L-HDAg. ADAR1 is another ISG and has an editing specificity on long double-stranded RNA (> 100 bp) [194]. As previously reported in the introduction (Chapter

1, section 1.3), HBV RNA is composed of three loops, including the epsilon loop which is essential for pgRNA encapsidation and retrotranscription. The introduction of mutations at this level might potentially affect the ability of HBV RNA to be introduced into new viral particles or to be retrotranscribed thus reducing the release of rcDNA. Indeed, it is still unknown whether the activation of these mutagenic enzymes, particularly ADAR1, could alter the variability of HBV and thus affect its replication.

The second part of this thesis has been focused on studying the role of HDV and the type I interferon pathway on HBV intrahepatic viral RNA variability.

To demonstrate the worth of our HBV/HDV system, we first analyzed HBV expression in the presence of HDV by quantifying both circulating and intrahepatic markers such as circulating HBV DNA and RNA and intracellular pgRNA. As expected, we observed that the presence of HDV affected the circulating HBV RNA and DNA levels, and this effect was reverted when type I interferon signaling was inhibited. This evidence suggests that HDV reduces the release of rcDNA in an IFN-dependent way. Notably, the extreme inhibition of circulating HBV RNA was congruent with the HBV DNA titer, as a difference of up to 3-log has been reported in CHB patients [135]. However, when we closely examined the expression of intrahepatic pgRNA we observed a slight reduction (less than 1 log) in WT mice superinfected with HDV compared to KO mice. However, this decline was very limited when compared to the reduction of circulating rcDNA. These results suggest that, in our system, the interaction between HDV and HBV might not be at the transcriptional level. However, it must be considered that the model used in this thesis cannot replicate the cccDNA, since the HBV DNA here was inserted into the mouse genome.

Once evaluated that our *in vivo* system was suitable for studying the interaction between HBV and HDV, we focused on studying if the presence of HDV could influence HBV variability.

This project was focused on the *SRT* and *HBX* genes. As explained in the previous chapters, (Chapter 5, section 1.4.1), the SRT interested the overlapping region between the active domain of the polymerase (RT) with the C-terminal region of the protein (S) (SRT region), because of their functions in viral replication and because mutations in one gene could be involved in the other [78]. The X gene (Chapter 1, section 1.2.4), instead, encodes for a protein that is essential for the transactivation of viral replication.

Looking at the SRT region, we observed that intrahepatic HBV RNA was more complex at 21 dpi than at 7dpi (higher rare haplotype load, or RHL). However, when we considered the different single nucleotide variations (SNVs), we observed a lower number of mutations in the QS at 21dpi rather than at 7dpi. This hypothetical incongruence could be explained by considering that the RHL measures complexity as much as the QS is enriched of haplotypes at low frequency, which means that the QS in this region might evolve by acquiring more haplotypes at low frequency but carrying few mutations. Some of these mutations were detected in all the experimental groups and at both timepoints. Among them, the C928T in the pol gene had also been reported in genotype C- infected CHB patients [337].

When considering the mutations that were detected differently between the experimental groups, we saw that at 7dpi superinfected (SI) WT mice presented more SNVs than the monoinfected (MI) mice. However, when we knocked down the type I IFN pathway, we didn't observe a decrease in the number of SNVs. Otherwise, KO mice presented even more mutations, with a different pattern between MI and SI mice. Notably, more than half of these mutations were C to T transitions, suggesting that other pathways rather than type I IFN could be involved in the activation of mutagenic enzymes at this timepoint. Notably, monoinfected WT mice have a complete predominance of this kind of transitions at this timepoint. Some of the mutations observed in the different groups could introduce amino acid changes in the S and/or RT, some of them of clinical interest. For example, a substitution at position 203 of the S gene (P203R) that has been associated with HCC in Vietnamese CHB patients [76], or the substitution in position 220 (F220L), which, together with other changes such as the V190A or S210N, determined a decrease in the secretion

of HBsAg *in vitro* and was associated with lower viremia (<10000 IU/mL) in CHB patients [338].

At 21dpi, the SNVs observed were less than 7dpi and they were generally distributed between the groups, especially when comparing WT MI and SI mice. However, the type of changes differed between HBVtg AAV-Luc and AAV-HDV since most of them were C to T transitions (7/8 mutated positions) in the first, whereas a more heterogeneous range (3/8) of changes characterized the second. The pattern of mutations changed when the type I interferon signaling was knocked down. This observation would hint that the changes observed at 21dpi in WT mice might potentially depend on the type I interferon activation. When comparing KO SI with MI mice, we observed two different groups of mutations in one group that were related to each other, but some of them were also observed in WT mice, thus indicating that the acquisition of these changes was not related to HDV expression or IFN I activation.

In the *HBX*, we observed a different trend than in the *SRT*. Both the 5' and 3' ends of the *HBX* revealed a decrease in the RHL at 21 dpi related to the early timepoint. Notably, the lowest value was seen in the 5'-*HBX* in SI WT mice, and a similar trend, although to a lesser extent, was detected in the 3'-*HBX* in the same group. In contrast to the RHL, the number of SNVs in the *HBX* increased strongly from 7 to 21dpi. This could indicate the existence of few haplotypes at low frequency, but highly mutated. A similar situation was described in a study published by our group when studying HBV DNA QS of chronically infected patients in the 5'-*HBX* region [96].

When analyzing the mutations found at 7 dpi, most of them appeared in almost all the groups, and some of them were also detected at 21dpi. This scenario changed at 21 dpi when more mutations were detected. When comparing wild type mice (HBVTg Luc vs HBVTg HDV), we found more changes along the entire gene in HBVTg Luc mice, with a prevalence of C->T transitions. However, in HBVTg HDV mice the mutations were more concentrated in the portion of the amplicon encoding for the transactivating domain of the HBx. Furthermore, the most relevant mutations observed implicated the HBVTg HDV mice that could introduce synonymous

variations involving enhancer or precore/core promoters that might potentially affect HBV expression [93]. Of them, just three were C->T transitions and the other mutations resulted in aa changes in the transactivating domain. Point mutations in the transactivating domain could affect the efficacy of HBx to promote HBV expression, as reported in a study that saw that substitutions introduced between nt 1554 and 1583 in the HBV genome could limit the transactivation of HBV expression with a substantial reduction of both intracellular total RNA, encapsidated rcDNA and HBsAg secretion [339]. Based on our results, the aa changes that presented a higher frequency in HBVtg HDV mice (F63L, P68S and L89F) were within a so-called Kunitz domain (aa 58 to 70). These domains are considered to be highly conserved because they protect HBx from cellular proteases-mediated degradation [340,341]. Of note, the effects of all the observed mutations on HBV expression remain to be elucidated.

Looking at the results viewed in HBV intrahepatic RNA QS altogether (in both *SRT* and *HBX*), we surprisingly detected a higher number of mutations and QS complexity than what was expected considering the model used. In fact, this *in vivo* model doesn't allow for the spread of infection (the exponential growth rate of the infection) associated with HBV re-infection. For this reason, and to avoid biased results, we focused on just those mutations that were present in at least two mice. Of note, some mutations were observed independently of the presence of HDV or a WT IFN pathway and interested both timepoints and sometimes were more frequent in different groups, such as the C1451T in the *SRT* highly frequent in HBVtg Luc at 7 dpi and in HBVtg IFN- α/β R KO HDV mice at 21dpi. Furthermore, most of these changes were C->T transitions, which are usually the result of deamination processes associated with APOBEC/AID family enzymes [342].

As previously commented, the mutagenic effects of cytosolic APOBEC (mainly A3G) on HBV rcDNA occur during pgRNA retrotranscription. In this study, we analyzed HBV QS at the transcription level, by focusing on the viral mRNAs prior to their eventual encapsidation. Of note, a quite recent *in vitro* study has reported that mouse APOBEC1 can introduce somatic mutations in animal genome without generating double-strand breaks due to its limited activity [343]. Like APOBEC, AID

can introduce C(G)->T(A) transitions in the cellular genome, an essential step during the development of antibody response [344]. Although APOBEC/AID seem to mainly edit ssDNA, the ability of A3A to introduce C->T changes into the mRNA has been demonstrated *in vitro* cultured monocytes and macrophages [345]. Moreover, a potential role of APOBEC in contributing to variability and evolution of SARS-CoV-2 (which is an RNA virus) has also been proposed [346–348]. Another *in vitro* study also reported that AID could promote HBV transcripts deamination and reduction [342,349]. The fact that all the C to T transitions involved all the experimental groups, might suggest that other pathways rather than type I interferon or HDV could be involved.

Notably, the image of HBV as a stealth virus is progressively changing. Indeed, a gene expression study of the liver of CHB patients versus uninfected showed that HBV infection increased the activation of the lymphotoxin β (LT β) signaling, which activates the non-canonical NF-k β pathway and, so the expression of APOBEC enzymes [350,351]. Moreover, an *in vitro* study suggested that HBx could directly enhance the NF-K β pathway [352,353]. This pathway is also involved in inducing the expression of APOBEC3B, which is also responsible for cccDNA degradation [354].

Moreover, other types of IFNs can induce the expression of the ISGs, and the activation of this pathway can also be related to the parenchymal context of liver tissue, such as the presence of resident immune cells. In fact, a study from Real C.I. *et al*, published in Scientific Report in 2016, showed that naked HBV particles (due to HBsAg deficiency) could stimulate liver Kupffer cells to promote immune responses, such as the expression of IFN1, IFIT1 and ISG15 [355]. Of note, the presence of naked capsid in the form of antibody complexes and carrying HBV RNA has been described in the plasma of CHB patients [356].

Another point to consider is that in our system the expression of HBV is slightly forced, since the integrated HBV genome presents 1.3 copies, with both enhancers I and II in duplicate, as well as the X gene which results in a high level of viral DNA replication in HBVtg mice with 1.3-fold integrated HBV genome [300]. This

construct might also explain why we observed more changes in HBX than SRT, since *HBX* transcripts, which are the shorter ones, could be expressed twice in this system. All these previously published data would suggest that all this high variability that we saw in all the experimental groups, and that was especially characterized by transitions associated with APOBEC/AID activity, could be related to the activation of these mutagenic enzymes in a type I-independent manner and that might directly involve HBV expression or the activation of different types of IFNs. Of note, a relatively large proportion of the SNVs observed in all the animals were repeated at both timepoints. This evidence would exclude that these changes were not an artifact and might suggest the existence of a hotspot of APOBEC/AID mutations in HBV transcripts.

However, in the presence of HDV, we didn't observe a strong increase in the proportion of C->T transitions corresponding to an augmented activation of the IFN signaling pathway. As previously reported, although the majority of SNVs identified in WT monoinfected mice were C->T, the typology of mutations in SI mice was more heterogeneous. This was even more evident when looking at *HBX*, where the not common change's types were in addition to C->T, also T->C and C->A mutations, and involved both SI mice (WT or KO). This observation would suggest that these changes might be produced by a different mechanism. Of note, it is well known that the interaction of S-HDAg with the cellular RNA pol II promotes RNA elongation and decreases polymerase fidelity [209,357]. The HBV genome is transcribed employing of the same cellular polymerase. The lack of fidelity of this polymerase could potentially introduce changes in HBV transcripts and contribute to the variability generated by APOBEC/AID enzymes.

Regarding HDV QS complexity, it was lower in KO mice than in WT at both timepoints, which suggest that this pathway might play a role in determining HDV variability. Firstly, this pathway induces the expression of ADAR1, which is responsible for the production of the long isoform of the delta antigen (L-HDAg). At 21 dpi, as expected, we observed a lower editing percentage in IFN-I pathway KO mice than in WT. The difference in the editing activity of this enzyme appears to be a remarkable effect of interferon. However, the limited but consistent percentage of

editing in KO HDV mice would suggest that another pathway rather than the interferon- α/β might be involved, such as type III IFB (IFN λ) [239,358,359]. Moreover, we must keep in mind that ADAR1 has several promoters, one interferon-inducible, and two that maintain its expression constitutively [360,361].

As commented in the introduction, the source of HDV variability is still not completely clear. Considering that the main source is the lack of fidelity of the cellular RNA pol II due to the interaction with S-HDAG, the reduction of the QS complexity at 21dpi might be related to the increase of the L-HDAG ratio, which limits the activity of the cellular polymerase. As reported in the introduction, the intranuclear level of L-HDAG is a determinant element to change the balance from viral RNA replication to viral particle release [170].

It is noteworthy that if we consider only the mutations that concerned WT versus KO mice, we saw that the number of SNVs decreased from 7 to 21 dpi. The difference in the type of changes between the two timepoints is outstanding. At 7 dpi, we observed mainly C \rightarrow T changes, whereas at 21 dpi the main type of changes were A \rightarrow G and T \rightarrow C transitions, which could correspond to ADAR1 activation. This is even more evident when we look at the non-common SNVs at 21dpi. At this timepoint, all the five SNVs observed in WT mice were A/T \rightarrow G/C transitions. Of note, ADAR1 edits viral antigenomic RNA when it is in the double-stranded rod-like structure. Although its highly specificity for the Amber/W site [362], the promiscuous editing in the non-Amber/W sites has been reported *in vitro*[363]. In this study the authors transfected HuH7 cells with plasmids expressing the genome and antigenome of HDV and then analyzed the variability by sequencing clones to identify the presence of A \rightarrow G mutations in non-Amber/W site, thus observing that promiscuous editing mainly interested edited sequence and were very infrequent (with 0.16 in genome and 0.32 in antigenome modified adenosine). Another study in woodchuck infected model showed that transitions probably associated with ADAR1 could appear at 73 weeks post-infection in the chronically infected animals, and that these changes were more frequent and occurred early in those animals that recovered from HDV infection [364], thus suggesting that the hyper-mutation of HDV RNA could affect the viral fitness thus promoting the control of the infection.

Remarkably, all the sequencing methods used in the above-mentioned studies were based on sequencing clones, which enables the analysis of just the most representative variants forming a QS.

Furthermore, although the main ADAR1 isoform that edits the Amber/W site is the constitutive and nuclear p110, an excess of the inducible isoform, for example due to an excessive stimulation of the IFNs pathway, could promote hypermutations on HDV [365]. The results on HDV QS would suggest that the role of ADAR1 in HDV replication might not be exclusively centered on editing and production of L-HDAg, if not also on viral QS variability. Again, since the model used in this project doesn't permit reinfection cycles, the mutations we observed remained at very low frequencies, and their impact on viral fitness cannot be evaluated in this system.

As deeply explained above in the introduction chapter, HDV can strongly activate the IFN response and promote the expression of the ISGs, but we didn't observe a strong increase in APOBEC or ADAR1-associated transitions in superinfected mice, which might suggest that this system was not the unique source of variability in the presence of delta virus.

Limitations of the studies

Based on the first study, the following limitations were identified:

- Although 25 patients were selected, only 19 of them were correctly amplified and passed the quality filters. It would be optimal to increase the number of patients and improve the amplification and purification of the genetic material to obtain a larger number of samples.
- All patients were infected with HDV genotype 1, which is the main genotype in infected patients in Spain [366]. The inclusion of more patients, with other origins, will be crucial to deeply inspect the ribozyme QS conservation, considering the high divergence among HDV genotypes and sub-genotypes.
- *In silico* modeling should be implemented to highlight the effects of the C64 deletion, T23C and T69C mutations on the secondary structure and consequently on the function of the ribozyme.
- Due to the difficulty in obtaining samples that amplified and passed the filters, it was also difficult to find paired samples for both HDV genomic regions analyzed, thus requiring further studies with more samples to provide a more in-in-depth assessment of the association between viral activity, quasispecies evolution, and immune response and the impact on disease progression and infection.

According to the second study, several limitations were identified:

- The transgenic mouse model used in this study is based on an integrated HBV genome that shows a high expression rate due to the presence of two copies of both expression enhancers. This might make it difficult to fully highlight the HDV-related inhibition of HBV, as a consistent HBV expression was present after 21dpi.
- Animals were sacrificed at 21dpi when HBV was still detectable. Extending the study to longer timepoints would help to determine the mechanism associated with HBV suppression and the specific role of HDV and IFN in HBV variability.
- Although the virus is produced following the stimulation of its own promoters, the viral particles released into the bloodstream cannot reinfect the cells and promote the spread of the infection since the hepatocytes do not express the receptor for the human hepatitis infection. Therefore, this system lacks expression of cccDNA. Moreover, the absence of reinfection processes impacts the frequency of the variants identified, since it is impossible to select the haplotypes forming the QS based on their fitness. Of note, the absence of reinfection could also be an advantage since it enables the detection of very low frequency and low fitness variants that might be lost after reinfection.
- In this study, the transgenic mouse model was generated with the IFN- α/β receptor inhibited (IFN- α/β R KO), thus inhibiting all the signaling associated with this receptor. However, the intracellular mutagenic enzyme can be stimulated by other pathways, including other types of IFN or another canonical or noncanonical signaling.

Conclusions

Regarding the first study we concluded:

1. As expected, considering its essential role in the viral life cycle, the ribozyme is overall highly conserved in the viral QS and did not change over time. The most conserved portion involved a domain that plays a direct role in the auto-catalysis process and may be a valuable target for the development of a new gene-silencing strategy against HDV.
2. The most variable region, however, involves a domain that is not essential from a functional point of view, but a deletion in this region may strongly impair viral replication. Considering its relatively high frequency, it may be a potential mechanism of viral persistence.
3. Considering data from previous reports, our results suggest that HDV quasispecies acquire mutations as a normal event during replication. However, the quasispecies evolve differently depending on the genomic portion.
4. In the case of the ribozyme, the quasispecies is less complex, and the variability is similar over time, as variants with lower fitness might rapidly disappear, leaving only those that have a limited effect on viral fitness.
5. In the case of HDAg, quasispecies are more variable, especially at shorter time-points, and their time-dependent decay might be the result of selection events that are likely to be related to both viral replication and immune pressure. Further studies are required to provide a more in-depth evaluation of the association between viral activity, quasispecies evolution, and immune response and its impact on disease progression.

Considering the second study we concluded:

1. Type I IFN-mediated inhibition of HBV in the presence of HDV was observed at circulating DNA and RNA levels, but such inhibition was not observed when intrahepatic RNA (pgRNA) was analyzed, suggesting a possible inhibition of HBV at the time of encapsidation.
2. The number of changes in both regions at the different timepoints was different, suggesting to us that each HBV region presents a different susceptibility to variability.
3. In general, high variability was found in the HBV RNA, with a high prevalence of C to T transitions, which are mutations typically introduced by APOBEC/AID family enzymes. This variability was not necessarily linked to the activation of the IFN-I pathway or to the presence of HDV, suggesting that HBV itself could activate an alternative pathway that downstream promotes the activation of this mutagenic enzyme.
4. Although HDV expression should strongly increase the activation of the IFN pathway, we didn't observe an increased proportion and frequency of the C->T transitions in superinfected mice (both WT and KO). Conversely, these two groups were the only ones to show alterations that weren't identified in the other groups, including alterations that couldn't be related to an intracellular mutagenic enzyme. These mutations could potentially be introduced by the same cellular RNA polymerase, which shows a reduced fidelity during transcription due to the presence of S-HDAg.
5. As expected, HDV is highly mutated at both timepoints, especially at 7 dpi. This is consistent with the higher expression of S-HDAg (less editing at this timepoint) which promotes RNA elongation and consequently the acquisition of mutations. However, at 21dpi, the most prevalent changes

were G to A transitions, which are associated with ADAR1 activity, thus suggesting that ADAR1, together with cellular RNA polymerase, ADAR1 might contribute to HDV protein and QS variability.

The NGS technique used for both studies, allows an in-depth study of the viral QS, and thanks to this platform we were able to perform an in-depth analysis of the viral quasispecies for both HBV and HDV. It allows us to study viral quasispecies, their evolution and their sequence conservation. This is extremely important for a highly variable virus such as HDV, especially when examining regions of the viral genome that are highly functional or under different selection pressures. In addition, this sequencing method also finds application in basic research, for example, to evaluate the acquisition of variation and its possible source, and to study mechanisms of interaction between viruses at the genetic level.

Futures perspectives

Concerning the first study, to have a complete spectrum of the conservation of the ribozyme, our attention is now focused on including HDV samples from other geographical areas, to compare the conservation of the region observed with genotype 1 with those of other genotypes. The hyper-conserved region observed in the HDV ribozyme could be a candidate for the design of an antisense oligonucleotide (ASO), as a new and direct antiviral strategy against HDV.

Finally, considering that the evolution rate of the QS in the *HDAG* changed depending on the follow-up time elapsed, the study of the evolution of the QS in this region could help to understand the mechanism (both virological and immunological) behind the fluctuations in HDV viremia observed in the CHD patients [367].

The model used in the second study of the thesis is a very helpful model to study the interaction between both viruses *in vivo*, however, longer timepoints are required to fully evaluate the role of HDV in HBV inhibition. In this sense, we have planned new experiments, in which the mice are sacrificed at HBV DNA undetectability, to evaluate the changes of intracellular pgRNA and try to determine the step in which HDV could inhibit HBV expression.

In addition, the analysis at later timepoints could allow the accumulation of variants thus elucidating the mechanism behind the variability of HBV in our system. To determine whether the interaction of HDAG with the RNA polymerase II could affect HBV variability, an *in vitro* assay is needed. In this experiment, the variability of HBV after natural infection will be studied by NGS in the presence of S-HDAG or L-HDAG alone.

Bibliography

1. Blumberg, B.S. A "New" Antigen in Leukemia Sera. *JAMA: The Journal of the American Medical Association* **1965**, *191*, 541, doi:10.1001/jama.1965.03080070025007.
2. Giles, J.P.; McCollum, R.W.; Berndtson, L.W.; Krugman, S. Viral Hepatitis. *New England Journal of Medicine* **1969**, *281*, 119–122, doi:10.1056/NEJM196907172810302.
3. Almeida, J.; Rubenstein, D.; Stott, E.J. New Antigen-Antibody System In Australia-Antigen-Positive Hepatitis. *The Lancet* **1971**, *298*, 1225–1227, doi:10.1016/S0140-6736(71)90543-5.
4. Dane, D.S.; Cameron, C.H.; Briggs, M. Virus-Like Particles In Serum Of Patients With Australia-Antigen-Associated Hepatitis. *The Lancet* **1970**, *295*, 695–698, doi:10.1016/S0140-6736(70)90926-8.
5. Galibert, F.; Mandart, E.; Fitoussi, F.; Tiollais, P.; Charnay, P. Nucleotide Sequence of the Hepatitis B Virus Genome (Subtype Ayw) Cloned in E. Coli. *Nature* **1979**, *281*, 646–650, doi:10.1038/281646a0.
6. Gusf, I.D.; Burrell, C.J.; Coulepis, A.G.; Robinson, W.S.; Zuckerman, A.J. Taxonomic Classification of Human Hepatitis B Virus. *Intervirology* **1986**, *25*, 14–29, doi:10.1159/000149651.
7. Trépo, C.; Chan, H.L.Y.; Lok, A. Hepatitis B Virus Infection. *The Lancet* **2014**, *384*, 2053–2063, doi:10.1016/S0140-6736(14)60220-8.
8. Zheng, X.; Li, X.; Liu, J.; Shi, L.; Wang, H.; Tian, K.; Pan, X. Horizontal Transmission Might Be a Common Route of Hepatitis B Virus Exposure in Highly Endemic Areas. *J Med Virol* **2022**, *94*, 4983–4992, doi:10.1002/jmv.27905.
9. World Health Organization (WHO) HBV (24 June 2022). <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b>.
10. Te, H.S.; Jensen, D.M. Epidemiology of Hepatitis B and C Viruses: A Global Overview. *Clin Liver Dis* **2010**, *14*, 1–21, doi:10.1016/j.cld.2009.11.009.
11. Papastergiou, V.; Lombardi, R.; MacDonald, D.; Tsochatzis, E.A. Global Epidemiology of Hepatitis B Virus (HBV) Infection. *Curr Hepatol Rep* **2015**, *14*, 171–178, doi:10.1007/s11901-015-0269-3.
12. Jefferies, M.; Rauff, B.; Rashid, H.; Lam, T.; Rafiq, S. Update on Global Epidemiology of Viral Hepatitis and Preventive Strategies. *World J Clin Cases* **2018**, *6*, 589–599, doi:10.12998/wjcc.v6.i13.589.
13. Schweitzer, A.; Horn, J.; Mikolajczyk, R.T.; Krause, G.; Ott, J.J. Estimations of Worldwide Prevalence of Chronic Hepatitis B Virus Infection: A Systematic Review of Data Published between 1965 and 2013. *The Lancet* **2015**, *386*, 1546–1555, doi:10.1016/S0140-6736(15)61412-X.

14. Nelson, P.K.; Mathers, B.M.; Cowie, B.; Hagan, H.; des Jarlais, D.; Horyniak, D.; Degenhardt, L. Global Epidemiology of Hepatitis B and Hepatitis C in People Who Inject Drugs: Results of Systematic Reviews. *The Lancet* **2011**, *378*, 571–583, doi:10.1016/S0140-6736(11)61097-0.
15. Urban, S.; Bartenschlager, R.; Kubitz, R.; Zoulim, F. Strategies to Inhibit Entry of HBV and HDV Into Hepatocytes. *Gastroenterology* **2014**, *147*, 48–64, doi:10.1053/j.gastro.2014.04.030.
16. van Damme, E.; Vanhove, J.; Severyn, B.; Verschueren, L.; Pauwels, F. The Hepatitis B Virus Interactome: A Comprehensive Overview. *Front Microbiol* **2021**, *12*, doi:10.3389/fmicb.2021.724877.
17. Schädler, S.; Hildt, E. HBV Life Cycle: Entry and Morphogenesis. *Viruses* **2009**, *1*, 185–209, doi:10.3390/v1020185.
18. Araujo, N.M.; Vianna, C.O.A.; Moraes, M.T.B.; Gomes, S.A. Expression of Hepatitis B Virus Surface Antigen (HBsAg) from Genotypes A, D and F and Influence of Amino Acid Variations Related or Not to Genotypes on HBsAg Detection. *Brazilian Journal of Infectious Diseases* **2009**, *13*, doi:10.1590/S1413-86702009000400005.
19. Chai, N.; Chang, H.E.; Nicolas, E.; Han, Z.; Jarnik, M.; Taylor, J. Properties of Subviral Particles of Hepatitis B Virus. *J Virol* **2008**, *82*, 7812–7817, doi:10.1128/JVI.00561-08.
20. Liu, C.-C.; Yansura, D.; Levinson, A.D. Direct Expression of Hepatitis B Surface Antigen in Monkey Cells from an SV40 Vector. *DNA* **1982**, *1*, 213–221, doi:10.1089/dna.1.1982.1.213.
21. Heermann, K.H.; Goldmann, U.; Schwartz, W.; Seyffarth, T.; Baumgarten, H.; Gerlich, W.H. Large Surface Proteins of Hepatitis B Virus Containing the Pre-S Sequence. *J Virol* **1984**, *52*, 396–402, doi:10.1128/jvi.52.2.396-402.1984.
22. Watanabe, T.; Sorensen, E.M.; Naito, A.; Schott, M.; Kim, S.; Ahlquist, P. Involvement of Host Cellular Multivesicular Body Functions in Hepatitis B Virus Budding. *Proceedings of the National Academy of Sciences* **2007**, *104*, 10205–10210, doi:10.1073/pnas.0704000104.
23. Jiang, B.; Himmelsbach, K.; Ren, H.; Boller, K.; Hildt, E. Subviral Hepatitis B Virus Filaments, like Infectious Viral Particles, Are Released via Multivesicular Bodies. *J Virol* **2016**, *90*, 3330–3341, doi:10.1128/JVI.03109-15.
24. Glebe, D.; Goldmann, N.; Lauber, C.; Seitz, S. HBV Evolution and Genetic Variability: Impact on Prevention, Treatment and Development of Antivirals. *Antiviral Res* **2021**, *186*, 104973, doi:10.1016/j.antiviral.2020.104973.
25. Hu, J.; Liu, K. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. *Viruses* **2017**, *9*, 56, doi:10.3390/v9030056.

26. Giersch, K.; Allweiss, L.; Volz, T.; Dandri, M.; Lütgehetmann, M. Serum HBV PgRNA as a Clinical Marker for CccDNA Activity. *J Hepatol* **2017**, *66*, 460–462, doi:10.1016/j.jhep.2016.09.028.
27. Beck, J. Hepatitis B Virus Replication. *World J Gastroenterol* **2007**, *13*, 48, doi:10.3748/wjg.v13.i1.48.
28. Okamoto, H.; Tsuda, F.; Sakugawa, H.; Sastrosoewignjo, R.I.; Imai, M.; Miyakawa, Y.; Mayumi, M. Typing Hepatitis B Virus by Homology in Nucleotide Sequence: Comparison of Surface Antigen Subtypes. *Journal of General Virology* **1988**, *69*, 2575–2583, doi:10.1099/0022-1317-69-10-2575.
29. Sattler, F.; Robinson, W.S. Hepatitis B Viral DNA Molecules Have Cohesive Ends. *J Virol* **1979**, *32*, 226–233, doi:10.1128/jvi.32.1.226-233.1979.
30. Doitsh, G.; Shaul, Y. Enhancer I Predominance in Hepatitis B Virus Gene Expression. *Mol Cell Biol* **2004**, *24*, 1799–1808, doi:10.1128/MCB.24.4.1799-1808.2004.
31. Minor, M.; Slagle, B. Hepatitis B Virus HBx Protein Interactions with the Ubiquitin Proteasome System. *Viruses* **2014**, *6*, 4683–4702, doi:10.3390/v6114683.
32. Luangsay, S.; Zoulim, F. Structure and Molecular Virology. In *Viral Hepatitis*; John Wiley & Sons, Ltd: Oxford, UK, 2013; pp. 63–80.
33. Seeger, C.; Mason, W.S. Molecular Biology of Hepatitis B Virus Infection. *Virology* **2015**, *479–480*, 672–686, doi:10.1016/j.virol.2015.02.031.
34. Glebe, D.; König, A. Molecular Virology of Hepatitis B Virus and Targets for Antiviral Intervention. *Intervirolgy* **2014**, *57*, 134–140, doi:10.1159/000360946.
35. Quarleri, J. Core Promoter: A Critical Region Where the Hepatitis B Virus Makes Decisions. *World J Gastroenterol* **2014**, *20*, 425, doi:10.3748/wjg.v20.i2.425.
36. Bock, C.T.; Schwinn, S.; Locarnini, S.; Fyfe, J.; Manns, M.P.; Trautwein, C.; Zentgraf, H. Structural Organization of the Hepatitis B Virus Minichromosome. *J Mol Biol* **2001**, *307*, 183–196, doi:10.1006/jmbi.2000.4481.
37. Li, H.-C.; Huang, E.-Y.; Su, P.-Y.; Wu, S.-Y.; Yang, C.-C.; Lin, Y.-S.; Chang, W.-C.; Shih, C. Nuclear Export and Import of Human Hepatitis B Virus Capsid Protein and Particles. *PLoS Pathog* **2010**, *6*, e1001162, doi:10.1371/journal.ppat.1001162.
38. Melegari, M.; Wolf, S.K.; Schneider, R.J. Hepatitis B Virus DNA Replication Is Coordinated by Core Protein Serine Phosphorylation and HBx Expression. *J Virol* **2005**, *79*, 9810–9820, doi:10.1128/JVI.79.15.9810-9820.2005.

39. Heger-Stevic, J.; Zimmermann, P.; Lecoq, L.; Böttcher, B.; Nassal, M. Hepatitis B Virus Core Protein Phosphorylation: Identification of the SRPK1 Target Sites and Impact of Their Occupancy on RNA Binding and Capsid Structure. *PLoS Pathog* **2018**, *14*, e1007488, doi:10.1371/journal.ppat.1007488.
40. Cole, A.G. Modulators of HBV Capsid Assembly as an Approach to Treating Hepatitis B Virus Infection. *Curr Opin Pharmacol* **2016**, *30*, 131–137, doi:10.1016/j.coph.2016.08.004.
41. Chen, M.T.; Billaud, J.-N.; Sällberg, M.; Guidotti, L.G.; Chisari, F. v.; Jones, J.; Hughes, J.; Milich, D.R. A Function of the Hepatitis B Virus Precore Protein Is to Regulate the Immune Response to the Core Antigen. *Proceedings of the National Academy of Sciences* **2004**, *101*, 14913–14918, doi:10.1073/pnas.0406282101.
42. Ito, K.; Kim, K.-H.; Lok, A.S.-F.; Tong, S. Characterization of Genotype-Specific Carboxyl-Terminal Cleavage Sites of Hepatitis B Virus e Antigen Precursor and Identification of Furin as the Candidate Enzyme. *J Virol* **2009**, *83*, 3507–3517, doi:10.1128/JVI.02348-08.
43. Schödel, F.; Peterson, D.; Milich, D. Hepatitis B Virus Core and e Antigen: Immune Recognition and Use as a Vaccine Carrier Moiety. *Intervirolgy* **1996**, *39*, 104–110, doi:10.1159/000150481.
44. Kramvis, A.; Kostaki, E.-G.; Hatzakis, A.; Paraskevis, D. Immunomodulatory Function of HBeAg Related to Short-Sighted Evolution, Transmissibility, and Clinical Manifestation of Hepatitis B Virus. *Front Microbiol* **2018**, *9*, doi:10.3389/fmicb.2018.02521.
45. Bozkaya, H.; Ayola, B.; Lok, A.S.F. High Rate of Mutations in the Hepatitis B Core Gene during the Immune Clearance Phase of Chronic Hepatitis B Virus Infection. *Hepatology* **1996**, *24*, 32–37, doi:10.1002/hep.510240107.
46. Roseman, A.M.; Borschukova, O.; Berriman, J.A.; Wynne, S.A.; Pumpens, P.; Crowther, R.A. Structures of Hepatitis B Virus Cores Presenting a Model Epitope and Their Complexes with Antibodies. *J Mol Biol* **2012**, *423*, 63–78, doi:10.1016/j.jmb.2012.06.032.
47. Zhang, Z.-H. Genetic Variation of Hepatitis B Virus and Its Significance for Pathogenesis. *World J Gastroenterol* **2016**, *22*, 126, doi:10.3748/wjg.v22.i1.126.
48. Locarnini, S.; McMillan, J.; Bartholomeusz, A. The Hepatitis B Virus and Common Mutants. *Semin Liver Dis* **2003**, *23*, 005–020, doi:10.1055/s-2003-37587.
49. Nassal, M. Hepatitis B Viruses: Reverse Transcription a Different Way. *Virus Res* **2008**, *134*, 235–249, doi:10.1016/j.virusres.2007.12.024.

50. Jones, S.A.; Hu, J. Hepatitis B Virus Reverse Transcriptase: Diverse Functions as Classical and Emerging Targets for Antiviral Intervention. *Emerg Microbes Infect* **2013**, *2*, 1–11, doi:10.1038/emi.2013.56.
51. Radziwill, G.; Tucker, W.; Schaller, H. Mutational Analysis of the Hepatitis B Virus P Gene Product: Domain Structure and RNase H Activity. *J Virol* **1990**, *64*, 613–620, doi:10.1128/jvi.64.2.613-620.1990.
52. Das, K.; Xiong, X.; Yang, H.; Westland, C.E.; Gibbs, C.S.; Sarafianos, S.G.; Arnold, E. Molecular Modeling and Biochemical Characterization Reveal the Mechanism of Hepatitis B Virus Polymerase Resistance to Lamivudine (3TC) and Emtricitabine (FTC). *J Virol* **2001**, *75*, 4771–4779, doi:10.1128/JVI.75.10.4771-4779.2001.
53. Kamer, G.; Argos, P. Primary Structural Comparison of RNA-Dependent Polymerases from Plant, Animal and Bacterial Viruses. *Nucleic Acids Res* **1984**, *12*, 7269–7282, doi:10.1093/nar/12.18.7269.
54. Melegari, M.; Scaglioni, P.P.; Wands, J.R. Hepatitis B Virus Mutants Associated with 3TC and Famciclovir Administration Are Replication Defective. *Hepatology* **1998**, *27*, 628–633, doi:10.1002/hep.510270243.
55. Garcia-Garcia, S.; Cortese, M.F.; Rodríguez-Algarra, F.; Tabernero, D.; Rando-Segura, A.; Quer, J.; Buti, M.; Rodríguez-Frías, F. Next-Generation Sequencing for the Diagnosis of Hepatitis B: Current Status and Future Prospects. *Expert Rev Mol Diagn* **2021**, *21*, 381–396, doi:10.1080/14737159.2021.1913055.
56. Clark, D.N.; Hu, J. Unveiling the Roles of HBV Polymerase for New Antiviral Strategies. *Future Virol* **2015**, *10*, 283–295, doi:10.2217/fvl.14.113.
57. Lampertico, P.; Agarwal, K.; Berg, T.; Buti, M.; Janssen, H.L.A.; Papatheodoridis, G.; Zoulim, F.; Tacke, F. EASL 2017 Clinical Practice Guidelines on the Management of Hepatitis B Virus Infection. *J Hepatol* **2017**, *67*, 370–398, doi:10.1016/j.jhep.2017.03.021.
58. EASL Clinical Practice Guidelines: Management of Chronic Hepatitis B Virus Infection. *J Hepatol* **2012**, *57*, 167–185, doi:10.1016/j.jhep.2012.02.010.
59. Lee, H.W.; Lee, J.S.; Ahn, S.H. Hepatitis B Virus Cure: Targets and Future Therapies. *Int J Mol Sci* **2020**, *22*, 213, doi:10.3390/ijms22010213.
60. Langley, D.R.; Walsh, A.W.; Baldick, C.J.; Eggers, B.J.; Rose, R.E.; Levine, S.M.; Kapur, A.J.; Colonno, R.J.; Tenney, D.J. Inhibition of Hepatitis B Virus Polymerase by Entecavir. *J Virol* **2007**, *81*, 3992–4001, doi:10.1128/JVI.02395-06.
61. Ling, R.; Harrison, T.J. Functional Analysis of Mutations Conferring Lamivudine Resistance on Hepatitis B Virus. *Journal of General Virology* **1999**, *80*, 601–606, doi:10.1099/0022-1317-80-3-601.

62. Carman, W.F. The Clinical Significance of Surface Antigen Variants of Hepatitis B Virus. *J Viral Hepat* **1997**, *4*, 11–20, doi:10.1111/j.1365-2893.1997.tb00155.x.
63. Feng, H.; Hu, K. Structural Characteristics and Molecular Mechanism of Hepatitis B Virus Reverse Transcriptase. *Viol Sin* **2009**, *24*, 509–517, doi:10.1007/s12250-009-3076-6.
64. Short, J.M.; Chen, S.; Roseman, A.M.; Butler, P.J.G.; Crowther, R.A. Structure of Hepatitis B Surface Antigen from Subviral Tubes Determined by Electron Cryomicroscopy. *J Mol Biol* **2009**, *390*, 135–141, doi:10.1016/j.jmb.2009.04.059.
65. Eble, B.E.; MacRae, D.R.; Lingappa, V.R.; Ganem, D. Multiple Topogenic Sequences Determine the Transmembrane Orientation of the Hepatitis B Surface Antigen. *Mol Cell Biol* **1987**, *7*, 3591–3601, doi:10.1128/MCB.7.10.3591.
66. Gripon, P.; Le Seyec, J.; Rumin, S.; Guguen-Guillouzo, C. Myristylation of the Hepatitis B Virus Large Surface Protein Is Essential for Viral Infectivity. *Virology* **1995**, *213*, 292–299, doi:10.1006/viro.1995.0002.
67. Bruss, V.; Hagelstein, J.; Gerhardt, E.; Galle, P.R. Myristylation of the Large Surface Protein Is Required for Hepatitis B Virusin Vitro Infectivity. *Virology* **1996**, *218*, 396–399, doi:10.1006/viro.1996.0209.
68. Matlack, K.E.S.; Misselwitz, B.; Plath, K.; Rapoport, T.A. *BiP Acts as a Molecular Ratchet during Posttranslational Transport of Prepro- α Factor across the ER Membrane*; 1999; pp. 553–564;.
69. Lambert, C.; Prange, R. Chaperone Action in the Posttranslational Topological Reorientation of the Hepatitis B Virus Large Envelope Protein: Implications for Translocational Regulation. *Proceedings of the National Academy of Sciences* **2003**, *100*, 5199–5204, doi:10.1073/pnas.0930813100.
70. Awe, K.; Lambert, C.; Prange, R. Mammalian BiP Controls Posttranslational ER Translocation of the Hepatitis B Virus Large Envelope Protein. *FEBS Lett* **2008**, *582*, 3179–3184, doi:10.1016/j.febslet.2008.07.062.
71. Rapoport, T.A.; Matlack, K.E.S.; Plath, K.; Misselwitz, B.; Staack, O. Posttranslational Protein Translocation Across the Membrane of the Endoplasmic Reticulum. *Biol Chem* **1999**, *380*, 1143–1150, doi:10.1515/BC.1999.145.
72. Bruss, V. A Short Linear Sequence in the Pre-S Domain of the Large Hepatitis B Virus Envelope Protein Required for Virion Formation. *J Virol* **1997**, *71*, 9350–9357, doi:10.1128/jvi.71.12.9350-9357.1997.
73. Gerlich, W.H.; Heermann, K.-H.; Xuanyong, L. Functions of Hepatitis B Surface Proteins. In; 1992; pp. 129–132.

74. Churin, Y.; Roderfeld, M.; Roeb, E. Hepatitis B Virus Large Surface Protein: Function and Fame. *Hepatobiliary Surg Nutr* **2015**, *4*, 1–10, doi:10.3978/j.issn.2304-3881.2014.12.08.
75. Dobrica, M.-O.; Lazar, C.; Branza-Nichita, N. N-Glycosylation and N-Glycan Processing in HBV Biology and Pathogenesis. *Cells* **2020**, *9*, 1404, doi:10.3390/cells9061404.
76. Thi Cam Huong, N.; Trung, N.Q.; Luong, B.A.; Tram, D.B.; Vu, H.A.; Bui, H.H.; Pham Thi Le, H. Mutations in the HBV PreS/S Gene Related to Hepatocellular Carcinoma in Vietnamese Chronic HBV-Infected Patients. *PLoS One* **2022**, *17*, e0266134, doi:10.1371/journal.pone.0266134.
77. Liu, W.; Cai, S.; Pu, R.; Li, Z.; Liu, D.; Zhou, X.; Yin, J.; Chen, X.; Chen, L.; Wu, J.; et al. HBV PreS Mutations Promote Hepatocarcinogenesis by Inducing Endoplasmic Reticulum Stress and Upregulating Inflammatory Signaling. *Cancers (Basel)* **2022**, *14*, 3274, doi:10.3390/cancers14133274.
78. Chen, Y.C.; Delbrook, K.; Dealwis, C.; Mimms, L.; Mushahwar, I.K.; Mandeck, W. Discontinuous Epitopes of Hepatitis B Surface Antigen Derived from a Filamentous Phage Peptide Library. *Proceedings of the National Academy of Sciences* **1996**, *93*, 1997–2001, doi:10.1073/pnas.93.5.1997.
79. Torresi, J. The Virological and Clinical Significance of Mutations in the Overlapping Envelope and Polymerase Genes of Hepatitis B Virus. *Journal of Clinical Virology* **2002**, *25*, 97–106, doi:10.1016/S1386-6532(02)00049-5.
80. Torresi, J.; Earnest-Silveira, L.; Deliyannis, G.; Edgton, K.; Zhuang, H.; Locarnini, S.A.; Fyfe, J.; Sozzi, T.; Jackson, D.C. Reduced Antigenicity of the Hepatitis B Virus HBsAg Protein Arising as a Consequence of Sequence Changes in the Overlapping Polymerase Gene That Are Selected by Lamivudine Therapy. *Virology* **2002**, *293*, 305–313, doi:10.1006/viro.2001.1246.
81. Shirvani-Dastgerdi, E.; Winer, B.Y.; Celià-Terrassa, T.; Kang, Y.; Taberner, D.; Yagmur, E.; Rodríguez-Frías, F.; Gregori, J.; Luedde, T.; Trautwein, C.; et al. Selection of the Highly Replicative and Partially Multidrug Resistant RtS78T HBV Polymerase Mutation during TDF-ETV Combination Therapy. *J Hepatol* **2017**, *67*, 246–254, doi:10.1016/j.jhep.2017.03.027.
82. Belloni, L.; Pollicino, T.; De Nicola, F.; Guerrieri, F.; Raffa, G.; Fanciulli, M.; Raimondo, G.; Levrero, M. Nuclear HBx Binds the HBV Minichromosome and Modifies the Epigenetic Regulation of CccDNA Function. *Proceedings of the National Academy of Sciences* **2009**, *106*, 19975–19979, doi:10.1073/pnas.0908365106.
83. Fallot, G.; Neuveut, C.; Buendia, M.-A. Diverse Roles of Hepatitis B Virus in Liver Cancer. *Curr Opin Virol* **2012**, *2*, 467–473, doi:10.1016/j.coviro.2012.05.008.

84. Murakami, S.; Cheong, J.H.; Kaneko, S. Human Hepatitis Virus X Gene Encodes a Regulatory Domain That Represses Transactivation of X Protein. *J Biol Chem* **1994**, *269*, 15118–15123.
85. Kumar, V.; Jayasuryan, N.; Kumar, R. A Truncated Mutant (Residues 58-140) of the Hepatitis B Virus X Protein Retains Transactivation Function. *Proceedings of the National Academy of Sciences* **1996**, *93*, 5647–5652, doi:10.1073/pnas.93.11.5647.
86. Tang, H.; Delgermaa, L.; Huang, F.; Oishi, N.; Liu, L.; He, F.; Zhao, L.; Murakami, S. The Transcriptional Transactivation Function of HBx Protein Is Important for Its Augmentation Role in Hepatitis B Virus Replication. *J Virol* **2005**, *79*, 5548–5556, doi:10.1128/JVI.79.9.5548-5556.2005.
87. Pollicino, T.; Belloni, L.; Raffa, G.; Pediconi, N.; Squadrito, G.; Raimondo, G.; Levrero, M. Hepatitis B Virus Replication Is Regulated by the Acetylation Status of Hepatitis B Virus CccDNA-Bound H3 and H4 Histones. *Gastroenterology* **2006**, *130*, 823–837, doi:10.1053/j.gastro.2006.01.001.
88. Murphy, C.M.; Xu, Y.; Li, F.; Nio, K.; Reszka-Blanco, N.; Li, X.; Wu, Y.; Yu, Y.; Xiong, Y.; Su, L. Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep* **2016**, *16*, 2846–2854, doi:10.1016/j.celrep.2016.08.026.
89. Tang, H.; Oishi, N.; Kaneko, S.; Murakami, S. Molecular Functions and Biological Roles of Hepatitis B Virus x Protein. *Cancer Sci* **2006**, *97*, 977–983, doi:10.1111/j.1349-7006.2006.00299.x.
90. Kumar, V.; Sarkar, D.P. Hepatitis B Virus X Protein: Structure-Function Relationships and Role in Viral Pathogenesis. In; 2004; pp. 377–407.
91. Kidd-Ljunggren, K.; Oberg, M.; Kidd, A.H. The Hepatitis B Virus X Gene: Analysis of Functional Domain Variation and Gene Phylogeny Using Multiple Sequences. *Journal of General Virology* **1995**, *76*, 2119–2130, doi:10.1099/0022-1317-76-9-2119.
92. Ali, A. Hepatitis B Virus, HBx Mutants and Their Role in Hepatocellular Carcinoma. *World J Gastroenterol* **2014**, *20*, 10238, doi:10.3748/wjg.v20.i30.10238.
93. Kim, H.; Lee, S.-A.; Kim, B.-J. X Region Mutations of Hepatitis B Virus Related to Clinical Severity. *World J Gastroenterol* **2016**, *22*, 5467, doi:10.3748/wjg.v22.i24.5467.
94. Qadri, I.; Fatima, K.; Abdel-Hafiz, H. Hepatitis B Virus X Protein Impedes the DNA Repair via Its Association with Transcription Factor, TFIID. *BMC Microbiol* **2011**, *11*, 48, doi:10.1186/1471-2180-11-48.
95. Kwun, H.J. Natural Variants of Hepatitis B Virus X Protein Have Differential Effects on the Expression of Cyclin-Dependent Kinase Inhibitor P21 Gene. *Nucleic Acids Res* **2004**, *32*, 2202–2213, doi:10.1093/nar/gkh553.

96. Cortese, M.F.; González, C.; Gregori, J.; Casillas, R.; Carioti, L.; Guerrero-Murillo, M.; Riveiro-Barciela, M.; Godoy, C.; Sopena, S.; Yll, M.; et al. Sophisticated Viral Quasispecies with a Genotype-Related Pattern of Mutations in the Hepatitis B X Gene of HBeAg-ve Chronically Infected Patients. *Sci Rep* **2021**, *11*, 4215, doi:10.1038/s41598-021-83762-4.
97. Sureau, C.; Salisse, J. A Conformational Heparan Sulfate Binding Site Essential to Infectivity Overlaps with the Conserved Hepatitis B Virus A-Determinant. *Hepatology* **2013**, *57*, 985–994, doi:10.1002/hep.26125.
98. Yan, H.; Liu, Y.; Sui, J.; Li, W. NTCP Opens the Door for Hepatitis B Virus Infection. *Antiviral Res* **2015**, *121*, 24–30, doi:10.1016/j.antiviral.2015.06.002.
99. Watashi, K.; Urban, S.; Li, W.; Wakita, T. NTCP and Beyond: Opening the Door to Unveil Hepatitis B Virus Entry. *Int J Mol Sci* **2014**, *15*, 2892–2905, doi:10.3390/ijms15022892.
100. Urban, S.; Schulze, A.; Dandri, M.; Petersen, J. The Replication Cycle of Hepatitis B Virus. *J Hepatol* **2010**, *52*, 282–284, doi:10.1016/j.jhep.2009.10.031.
101. Kann, M. Intracellular Transport of Hepatitis B Virus. *World J Gastroenterol* **2007**, *13*, 39, doi:10.3748/wjg.v13.i1.39.
102. Wei, L.; Ploss, A. Mechanism of Hepatitis B Virus CccDNA Formation. *Viruses* **2021**, *13*, 1463, doi:10.3390/v13081463.
103. Xia, Y.; Guo, H. Hepatitis B Virus CccDNA: Formation, Regulation and Therapeutic Potential. *Antiviral Res* **2020**, *180*, 104824, doi:10.1016/j.antiviral.2020.104824.
104. Shon, J.K.; Shon, B.H.; Park, I.Y.; Lee, S.U.; Fa, L.; Chang, K.Y.; Shin, J.H.; Lee, Y.I. Hepatitis B Virus-X Protein Recruits Histone Deacetylase 1 to Repress Insulin-like Growth Factor Binding Protein 3 Transcription. *Virus Res* **2009**, *139*, 14–21, doi:10.1016/j.virusres.2008.09.006.
105. Salerno, D.; Chiodo, L.; Alfano, V.; Floriot, O.; Cottone, G.; Paturel, A.; Pallocca, M.; Plissonnier, M.-L.; Jeddari, S.; Belloni, L.; et al. Hepatitis B Protein HBx Binds the DLEU2 LncRNA to Sustain CccDNA and Host Cancer-Related Gene Transcription. *Gut* **2020**, *69*, 2016–2024, doi:10.1136/gutjnl-2019-319637.
106. Shen, C.; Feng, X.; Mao, T.; Yang, D.; Zou, J.; Zao, X.; Deng, Q.; Chen, X.; Lu, F. Yin-Yang 1 and HBx Protein Activate HBV Transcription by Mediating the Spatial Interaction of CccDNA Minichromosome with Cellular Chromosome 19p13.11. *Emerg Microbes Infect* **2020**, *9*, 2455–2464, doi:10.1080/22221751.2020.1840311.
107. López-Cabrera, M.; Letovsky, J.; Hu, K.-Q.; Siddiqui, A. Transcriptional Factor C/EBP Binds to and Transactivates the Enhancer Element II of the Hepatitis B Virus. *Virology* **1991**, *183*, 825–829, doi:10.1016/0042-6822(91)91019-D.

108. Mohd-Ismail; Lim; Gunaratne; Tan Mapping the Interactions of HBV CccDNA with Host Factors. *Int J Mol Sci* **2019**, *20*, 4276, doi:10.3390/ijms20174276.
109. Lim, C.S.; Brown, C.M. Hepatitis B Virus Nuclear Export Elements: RNA Stem-Loop α and β , Key Parts of the HBV Post-Transcriptional Regulatory Element. *RNA Biol* **2016**, *13*, 743–747, doi:10.1080/15476286.2016.1166330.
110. Kidd, A. A Revised Secondary Structure Model for the 3'-End of Hepatitis B Virus Pregenomic RNA. *Nucleic Acids Res* **1996**, *24*, 3295–3301, doi:10.1093/nar/24.17.3295.
111. Jones, S.A.; Hu, J. Hepatitis B Virus Reverse Transcriptase: Diverse Functions as Classical and Emerging Targets for Antiviral Intervention. *Emerg Microbes Infect* **2013**, *2*, 1–11, doi:10.1038/emi.2013.56.
112. le Pogam, S.; Chua, P.K.; Newman, M.; Shih, C. Exposure of RNA Templates and Encapsidation of Spliced Viral RNA Are Influenced by the Arginine-Rich Domain of Human Hepatitis B Virus Core Antigen (HBcAg 165-173). *J Virol* **2005**, *79*, 1871–1887, doi:10.1128/JVI.79.3.1871-1887.2005.
113. Tang, H.; McLachlan, A. A Pregenomic RNA Sequence Adjacent to DR1 and Complementary to Epsilon Influences Hepatitis B Virus Replication Efficiency. *Virology* **2002**, *303*, 199–210, doi:10.1006/viro.2002.1645.
114. Glebe, D.; Bremer, C. The Molecular Virology of Hepatitis B Virus. *Semin Liver Dis* **2013**, *33*, 103–112, doi:10.1055/s-0033-1345717.
115. Yang, W.; Summers, J. Illegitimate Replication of Linear Hepadnavirus DNA through Nonhomologous Recombination. *J Virol* **1995**, *69*, 4029–4036, doi:10.1128/jvi.69.7.4029-4036.1995.
116. Staprans, S.; Loeb, D.D.; Ganem, D. Mutations Affecting Hepadnavirus Plus-Strand DNA Synthesis Dissociate Primer Cleavage from Translocation and Reveal the Origin of Linear Viral DNA. *J Virol* **1991**, *65*, 1255–1262, doi:10.1128/jvi.65.3.1255-1262.1991.
117. Meier, M.-A.; Calabrese, D.; Suslov, A.; Terracciano, L.M.; Heim, M.H.; Wieland, S. Ubiquitous Expression of HBsAg from Integrated HBV DNA in Patients with Low Viral Load. *J Hepatol* **2021**, *75*, 840–847, doi:10.1016/j.jhep.2021.04.051.
118. Lambert, C.; Döring, T.; Prange, R. Hepatitis B Virus Maturation Is Sensitive to Functional Inhibition of ESCRT-III, Vps4, and Γ 2-Adaptin. *J Virol* **2007**, *81*, 9050–9060, doi:10.1128/JVI.00479-07.
119. Raimondo, G.; Locarnini, S.; Pollicino, T.; Levrero, M.; Zoulim, F.; Lok, A.S.; Allain, J.-P.; Berg, T.; Bertoletti, A.; Brunetto, M.R.; et al. Update of the Statements on Biology and Clinical Impact of Occult Hepatitis B Virus Infection. *J Hepatol* **2019**, *71*, 397–408, doi:10.1016/j.jhep.2019.03.034.

120. Pollicino, T.; Caminiti, G. HBV-Integration Studies in the Clinic: Role in the Natural History of Infection. *Viruses* **2021**, *13*, 368, doi:10.3390/v13030368.
121. Fattovich, G.; Bortolotti, F.; Donato, F. Natural History of Chronic Hepatitis B: Special Emphasis on Disease Progression and Prognostic Factors. *J Hepatol* **2008**, *48*, 335–352, doi:10.1016/j.jhep.2007.11.011.
122. Song, A.; Lin, X.; Chen, X. Functional Cure for Chronic Hepatitis B: Accessibility, Durability, and Prognosis. *Virol J* **2021**, *18*, 114, doi:10.1186/s12985-021-01589-x.
123. Tanaka, E.; Matsumoto, A. Guidelines for Avoiding Risks Resulting from Discontinuation of Nucleoside/Nucleotide Analogs in Patients with Chronic Hepatitis B. *Hepatology Research* **2014**, *44*, 1–8, doi:10.1111/hepr.12108.
124. Kim, G.-A.; Lee, H.C.; Kim, M.-J.; Ha, Y.; Park, E.J.; An, J.; Lee, D.; Shim, J.H.; Kim, K.M.; Lim, Y.-S. Incidence of Hepatocellular Carcinoma after HBsAg Seroclearance in Chronic Hepatitis B Patients: A Need for Surveillance. *J Hepatol* **2015**, *62*, 1092–1099, doi:10.1016/j.jhep.2014.11.031.
125. Papatheodoridis, G. v.; Chan, H.L.-Y.; Hansen, B.E.; Janssen, H.L.A.; Lampertico, P. Risk of Hepatocellular Carcinoma in Chronic Hepatitis B: Assessment and Modification with Current Antiviral Therapy. *J Hepatol* **2015**, *62*, 956–967, doi:10.1016/j.jhep.2015.01.002.
126. Yang, Y.-C.; Yang, H.-C. Recent Progress and Future Prospective in HBV Cure by CRISPR/Cas. *Viruses* **2021**, *14*, 4, doi:10.3390/v14010004.
127. van den Berg, F.; Limani, S.W.; Mnyandu, N.; Maepa, M.B.; Ely, A.; Arbuthnot, P. Advances with RNAi-Based Therapy for Hepatitis B Virus Infection. *Viruses* **2020**, *12*, 851, doi:10.3390/v12080851.
128. Gane, E.; Locarnini, S.; Lim, T.H.; Strasser, S.; Sievert, W.; Cheng, W.; Thompson, A.; Given, B.; Schluep, T.; Hamilton, J.; et al. Short-Term Treatment with RNA Interference Therapy, JNJ-3989, Results in Sustained Hepatitis B Surface Antigen Suppression in Patients with Chronic Hepatitis B Receiving Nucleos(t)ide Analogue Treatment. *J Hepatol* **2020**, *73*, S20, doi:10.1016/S0168-8278(20)30597-3.
129. Springer, A.D.; Dowdy, S.F. GalNAc-SiRNA Conjugates: Leading the Way for Delivery of RNAi Therapeutics. *Nucleic Acid Ther* **2018**, *28*, 109–118, doi:10.1089/nat.2018.0736.
130. Tan, Y.C.; Lee, G.H.; Huang, D.Q.; Lim, S.G. Future <sc>anti-HDV</Sc> Treatment Strategies, Including Those Aimed at <sc>HBV</Sc> Functional Cure. *Liver International* **2022**, doi:10.1111/liv.15387.
131. Fanning, G.C.; Zoulim, F.; Hou, J.; Bertoletti, A. Therapeutic Strategies for Hepatitis B Virus Infection: Towards a Cure. *Nat Rev Drug Discov* **2019**, *18*, 827–844, doi:10.1038/s41573-019-0037-0.

132. Mak, L.-Y.; Seto, W.-K.; Fung, J.; Yuen, M.-F. New Biomarkers of Chronic Hepatitis B. *Gut Liver* **2019**, *13*, 589–595, doi:10.5009/gnl18425.
133. van Bömmel, F.; Bartens, A.; Mysickova, A.; Hofmann, J.; Krüger, D.H.; Berg, T.; Edelmann, A. Serum Hepatitis B Virus RNA Levels as an Early Predictor of Hepatitis B Envelope Antigen Seroconversion during Treatment with Polymerase Inhibitors. *Hepatology* **2015**, *61*, 66–76, doi:10.1002/hep.27381.
134. Liu, Y.; Jiang, M.; Xue, J.; Yan, H.; Liang, X. Serum HBV RNA Quantification: Useful for Monitoring Natural History of Chronic Hepatitis B Infection. *BMC Gastroenterol* **2019**, *19*, 53, doi:10.1186/s12876-019-0966-4.
135. Cortese, M.F.; Riveiro-Barciela, M.; Taberner, D.; Rodriguez-Algarra, F.; Palom, A.; Sopena, S.; Rando-Segura, A.; Roade, L.; Kuchta, A.; Ferrer-Costa, R.; et al. Standardized Hepatitis B Virus RNA Quantification in Untreated and Treated Chronic Patients: A Promising Marker of Infection Follow-Up. *Microbiol Spectr* **2022**, *10*, doi:10.1128/spectrum.02149-21.
136. Kimura, T.; Ohno, N.; Terada, N.; Rokuhara, A.; Matsumoto, A.; Yagi, S.; Tanaka, E.; Kiyosawa, K.; Ohno, S.; Maki, N. Hepatitis B Virus DNA-Negative Dane Particles Lack Core Protein but Contain a 22-KDa Precore Protein without C-Terminal Arginine-Rich Domain. *Journal of Biological Chemistry* **2005**, *280*, 21713–21719, doi:10.1074/jbc.M501564200.
137. Kimura, T.; Rokuhara, A.; Sakamoto, Y.; Yagi, S.; Tanaka, E.; Kiyosawa, K.; Maki, N. Sensitive Enzyme Immunoassay for Hepatitis B Virus Core-Related Antigens and Their Correlation to Virus Load. *J Clin Microbiol* **2002**, *40*, 439–445, doi:10.1128/JCM.40.2.439-445.2002.
138. Tada, T.; Kumada, T.; Toyoda, H.; Kiriya, S.; Tanikawa, M.; Hisanaga, Y.; Kanamori, A.; Kitabatake, S.; Yama, T.; Tanaka, J. HBcrAg Predicts Hepatocellular Carcinoma Development: An Analysis Using Time-Dependent Receiver Operating Characteristics. *J Hepatol* **2016**, *65*, 48–56, doi:10.1016/j.jhep.2016.03.013.
139. Matsumoto, A.; Tanaka, E.; Minami, M.; Okanoue, T.; Yatsunami, H.; Nagaoka, S.; Suzuki, F.; Kobayashi, M.; Chayama, K.; Imamura, M.; et al. Low Serum Level of Hepatitis B Core-Related Antigen Indicates Unlikely Reactivation of Hepatitis after Cessation of Lamivudine Therapy. *Hepatology Research* **2007**, *37*, 661–666, doi:10.1111/j.1872-034X.2007.00094.x.
140. Seto, W.-K.; Wong, D.-H.; Chan, T.-Y.; Hwang, Y.-Y.; Fung, J.; Liu, K.S.-H.; Gill, H.; Lam, Y.-F.; Cheung, K.-S.; Lie, A.K.W.; et al. Association of Hepatitis B Core-Related Antigen With Hepatitis B Virus Reactivation in Occult Viral Carriers Undergoing High-Risk Immunosuppressive Therapy. *American Journal of Gastroenterology* **2016**, *111*, 1788–1795, doi:10.1038/ajg.2016.436.
141. Rizzetto, M.; Canese, M.G.; Arico, S.; Crivelli, O.; Trepo, C.; Bonino, F.; Verme, G. Immunofluorescence Detection of New Antigen-Antibody System (Delta/Anti-Delta) Associated to Hepatitis B Virus in Liver and in Serum of HBsAg Carriers. *Gut* **1977**, *18*, 997–1003, doi:10.1136/gut.18.12.997.

142. Rizzetto, M.; Hoyer, B.; Canese, M.G.; Shih, J.W.; Purcell, R.H.; Gerin, J.L. Delta Agent: Association of Delta Antigen with Hepatitis B Surface Antigen and RNA in Serum of Delta-Infected Chimpanzees. *Proceedings of the National Academy of Sciences* **1980**, *77*, 6124–6128, doi:10.1073/pnas.77.10.6124.
143. Rizzetto, M. The Delta Agent. *Hepatology* **2007**, *3*, 729–737, doi:10.1002/hep.1840030518.
144. World Health Organization (WHO) HDV (24 June 2022). <https://www.who.int/news-room/fact-sheets/detail/hepatitis-d>.
145. Palom, A.; Rando-Segura, A.; Vico, J.; Pacín, B.; Vargas, E.; Barreira-Díaz, A.; Rodríguez-Frías, F.; Riveiro-Barciela, M.; Esteban, R.; Buti, M. Implementation of Anti-HDV Reflex Testing among HBsAg-Positive Individuals Increases Testing for Hepatitis D. *JHEP Reports* **2022**, *4*, 100547, doi:10.1016/j.jhepr.2022.100547.
146. Stockdale, A.J.; Kreuels, B.; Henrion, M.Y.R.; Giorgi, E.; Kyomuhangi, I.; de Martel, C.; Hutin, Y.; Geretti, A.M. The Global Prevalence of Hepatitis D Virus Infection: Systematic Review and Meta-Analysis. *J Hepatol* **2020**, *73*, 523–532, doi:10.1016/j.jhep.2020.04.008.
147. Rizzetto, M.; Hamid, S.; Negro, F. The Changing Context of Hepatitis D. *J Hepatol* **2021**, *74*, 1200–1211, doi:10.1016/j.jhep.2021.01.014.
148. Hayashi, T.; Takeshita, Y.; Hutin, Y.J.-F.; Harmanci, H.; Easterbrook, P.; Hess, S.; van Holten, J.; Oru, E.O.; Kaneko, S.; Yurdaydin, C.; et al. The Global Hepatitis Delta Virus (HDV) Epidemic: What Gaps to Address in Order to Mount a Public Health Response? *Archives of Public Health* **2021**, *79*, 180, doi:10.1186/s13690-021-00693-2.
149. Negro, F. Hepatitis D Virus Coinfection and Superinfection. *Cold Spring Harb Perspect Med* **2014**, *4*, a021550–a021550, doi:10.1101/cshperspect.a021550.
150. Fattovich, G.; Boscaro, S.; Noventa, F.; Pornaro, E.; Stenico, D.; Alberti, A.; Ruol, A.; Realdi, G. Influence of Hepatitis Delta Virus Infection on Progression to Cirrhosis in Chronic Hepatitis Type B. *Journal of Infectious Diseases* **1987**, *155*, 931–935, doi:10.1093/infdis/155.5.931.
151. SARACCO, G. Serologic Markers with Fulminant Hepatitis in Persons Positive for Hepatitis B Surface Antigen. *Ann Intern Med* **1988**, *108*, 380, doi:10.7326/0003-4819-108-3-380.
152. Tseligka, E.D.; Clément, S.; Negro, F. HDV Pathogenesis: Unravelling Ariadne's Thread. *Viruses* **2021**, *13*, 778, doi:10.3390/v13050778.
153. Pollicino, T.; Raffa, G.; Santantonio, T.; Gaeta, G.B.; Iannello, G.; Alibrandi, A.; Squadrito, G.; Cacciola, I.; Calvi, C.; Colucci, G.; et al. Replicative and Transcriptional Activities of Hepatitis B Virus in Patients Coinfected with

- Hepatitis B and Hepatitis Delta Viruses. *J Virol* **2011**, *85*, 432–439, doi:10.1128/JVI.01609-10.
154. Ricco, G.; Popa, D.C.; Cavallone, D.; Iacob, S.; Salvati, A.; Tabacelia, D.; Oliveri, F.; Mascolo, G.; Bonino, F.; Yuan, Q.; et al. Quantification of Serum Markers of Hepatitis B (HBV) and Delta Virus (HDV) Infections in Patients with Chronic HDV Infection. *J Viral Hepat* **2018**, *25*, 911–919, doi:10.1111/jvh.12895.
 155. Moelling, K.; Broecker, F. Viroids and the Origin of Life. *Int J Mol Sci* **2021**, *22*, 3476, doi:10.3390/ijms22073476.
 156. Flores, R.; Ruiz-Ruiz, S.; Serra, P. Viroids and Hepatitis Delta Virus. *Semin Liver Dis* **2012**, *32*, 201–210, doi:10.1055/s-0032-1323624.
 157. Taylor, J.M. Hepatitis Delta Virus. *Virology* **2006**, *344*, 71–76, doi:10.1016/j.virol.2005.09.033.
 158. He, L.-F.; Ford, E.; Purcell, R.H.; London, W.T.; Phillips, J.; Gerin, J.L. The Size of the Hepatitis Delta Agent. *J Med Virol* **1989**, *27*, 31–33, doi:10.1002/jmv.1890270107.
 159. 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. Structure, Sequence and Expression of the Hepatitis Delta (δ) Viral Genome. *Nature* **1986**, *323*, 508–514, doi:10.1038/323508a0.
 160. Gudima, S.; Chang, J.; Moraleda, G.; Azvolinsky, A.; Taylor, J. Parameters of Human Hepatitis Delta Virus Genome Replication: The Quantity, Quality, and Intracellular Distribution of Viral Proteins and RNA. *J Virol* **2002**, *76*, 3709–3719, doi:10.1128/JVI.76.8.3709-3719.2002.
 161. Turon-Lagot, V.; Saviano, A.; Schuster, C.; Baumert, T.F.; Verrier, E.R. Targeting the Host for New Therapeutic Perspectives in Hepatitis D. *J Clin Med* **2020**, *9*, 222, doi:10.3390/jcm9010222.
 162. Zhang, Z.; Ni, Y.; Lempp, F.A.; Walter, L.; Mutz, P.; Bartenschlager, R.; Urban, S. Hepatitis D Virus-Induced Interferon Response and Administered Interferons Control Cell Division-Mediated Virus Spread. *J Hepatol* **2022**, *77*, 957–966, doi:10.1016/j.jhep.2022.05.023.
 163. Samuel, D.; Zignego, A.L.; Reynes, M.; Feray, C.; Arulnaden, J.L.; David, M.F.; Gigou, M.; Bismuth, A.; Mathieu, D.; Gentilini, P. Long-Term Clinical and Virological Outcome after Liver Transplantation for Cirrhosis Caused by Chronic Delta Hepatitis. *Hepatology* **1995**, *21*, 333–339.
 164. Mederacke, I.; Filmann, N.; Yurdaydin, C.; Bremer, B.; Puls, F.; Zacher, B.J.; Heidrich, B.; Tillmann, H.L.; Rosenau, J.; Bock, C.-T.; et al. Rapid Early HDV RNA Decline in the Peripheral Blood but Prolonged Intrahepatic Hepatitis Delta Antigen Persistence after Liver Transplantation. *J Hepatol* **2012**, *56*, 115–122, doi:10.1016/j.jhep.2011.06.016.

165. Giersch, K.; Helbig, M.; Volz, T.; Allweiss, L.; Mancke, L. v.; Lohse, A.W.; Polywka, S.; Pollok, J.M.; Petersen, J.; Taylor, J.; et al. Persistent Hepatitis D Virus Mono-Infection in Humanized Mice Is Efficiently Converted by Hepatitis B Virus to a Productive Co-Infection. *J Hepatol* **2014**, *60*, 538–544, doi:10.1016/j.jhep.2013.11.010.
166. Chou, H.-C.; Hsieh, T.-Y.; Sheu, G.-T.; Lai, M.M.C. Hepatitis Delta Antigen Mediates the Nuclear Import of Hepatitis Delta Virus RNA. *J Virol* **1998**, *72*, 3684–3690, doi:10.1128/JVI.72.5.3684-3690.1998.
167. Xia, Y.P.; Yeh, C.T.; Ou, J.H.; Lai, M.M. Characterization of Nuclear Targeting Signal of Hepatitis Delta Antigen: Nuclear Transport as a Protein Complex. *J Virol* **1992**, *66*, 914–921, doi:10.1128/jvi.66.2.914-921.1992.
168. Ding, B.; Kwon, M.-O.; Hammond, R.; Owens, R. Cell-to-Cell Movement of Potato Spindle Tuber Viroid. *The Plant Journal* **1997**, *12*, 931–936, doi:10.1046/j.1365-313X.1997.12040931.x.
169. Mentha, N.; Clément, S.; Negro, F.; Alfaiate, D. A Review on Hepatitis D: From Virology to New Therapies. *J Adv Res* **2019**, *17*, 3–15, doi:10.1016/j.jare.2019.03.009.
170. Zi, J.; Gao, X.; Du, J.; Xu, H.; Niu, J.; Chi, X. Multiple Regions Drive Hepatitis Delta Virus Proliferation and Are Therapeutic Targets. *Front Microbiol* **2022**, *13*, doi:10.3389/fmicb.2022.838382.
171. Peng, H.; Latifi, B.; Müller, S.; Lupták, A.; Chen, I.A. Self-Cleaving Ribozymes: Substrate Specificity and Synthetic Biology Applications. *RSC Chem Biol* **2021**, *2*, 1370–1383, doi:10.1039/D0CB00207K.
172. Webb, C.-H.T.; Lupták, A. HDV-like Self-Cleaving Ribozymes. *RNA Biol* **2011**, *8*, 719–727, doi:10.4161/rna.8.5.16226.
173. Taylor, J.M. Chapter 3 Replication of the Hepatitis Delta Virus RNA Genome. In; 2009; pp. 103–121.
174. Sureau, C.; Negro, F. The Hepatitis Delta Virus: Replication and Pathogenesis. *J Hepatol* **2016**, *64*, S102–S116, doi:10.1016/j.jhep.2016.02.013.
175. Reid, C.E.; Lazinski, D.W. A Host-Specific Function Is Required for Ligation of a Wide Variety of Ribozyme-Processed RNAs. *Proceedings of the National Academy of Sciences* **2000**, *97*, 424–429, doi:10.1073/pnas.97.1.424.
176. Sharmeen, L.; Kuo, M.Y.; Taylor, J. Self-Ligating RNA Sequences on the Antigenome of Human Hepatitis Delta Virus. *J Virol* **1989**, *63*, 1428–1430, doi:10.1128/jvi.63.3.1428-1430.1989.
177. Riccitelli, N.; Lupták, A. HDV Family of Self-Cleaving Ribozymes. In; 2013; pp. 123–171.

178. Ferré-D'Amaré, A.R.; Zhou, K.; Doudna, J.A. Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature* **1998**, *395*, 567–574, doi:10.1038/26912.
179. Thill, G.; Vasseur, M.; Tanner, N.K. Structural and Sequence Elements Required for the Self-Cleaving Activity of the Hepatitis Delta Virus Ribozyme. *Biochemistry* **1993**, *32*, 4254–4262, doi:10.1021/bi00067a013.
180. Chen, J.-H.; Yajima, R.; Chadalavada, D.M.; Chase, E.; Bevilacqua, P.C.; Golden, B.L. A 1.9 Å Crystal Structure of the HDV Ribozyme Precleavage Suggests Both Lewis Acid and General Acid Mechanisms Contribute to Phosphodiester Cleavage. *Biochemistry* **2010**, *49*, 6508–6518, doi:10.1021/bi100670p.
181. Ke, A.; Zhou, K.; Ding, F.; Cate, J.H.D.; Doudna, J.A. A Conformational Switch Controls Hepatitis Delta Virus Ribozyme Catalysis. *Nature* **2004**, *429*, 201–205, doi:10.1038/nature02522.
182. Been, M.D.; Wickham, G.S. Self-Cleaving Ribozymes of Hepatitis Delta Virus RNA. *Eur J Biochem* **1997**, *247*, 741–753, doi:10.1111/j.1432-1033.1997.00741.x.
183. Nishikawa, F.; Shirai, M.; Nishikawa, S. Site-Specific Modification of Functional Groups in Genomic Hepatitis Delta Virus (HDV) Ribozyme. *Eur J Biochem* **2002**, *269*, 5792–5803, doi:10.1046/j.1432-1033.2002.03280.x.
184. Deschenes, P. Formation of the P1.1 Pseudoknot Is Critical for Both the Cleavage Activity and Substrate Specificity of an Antigenomic Trans-Acting Hepatitis Delta Ribozyme. *Nucleic Acids Res* **2003**, *31*, 2087–2096, doi:10.1093/nar/gkg307.
185. Sharmeen, L.; Kuo, M.Y.; Dinter-Gottlieb, G.; Taylor, J. Antigenomic RNA of Human Hepatitis Delta Virus Can Undergo Self-Cleavage. *J Virol* **1988**, *62*, 2674–2679, doi:10.1128/jvi.62.8.2674-2679.1988.
186. Prody, G.A.; Bakos, J.T.; Buzayan, J.M.; Schneider, I.R.; Bruening, G. Autolytic Processing of Dimeric Plant Virus Satellite RNA. *Science (1979)* **1986**, *231*, 1577–1580, doi:10.1126/science.231.4745.1577.
187. Hutchins, C.J.; Rathjen, P.D.; Forster, A.C.; Symons, R.H. Self-Cleavage of plus and Minus RNA Transcripts of Avocado Sunblotch Viroid. *Nucleic Acids Res* **1986**, *14*, 3627–3640, doi:10.1093/nar/14.9.3627.
188. Forster, A.C.; Symons, R.H. Self-Cleavage of plus and Minus RNAs of a Virusoid and a Structural Model for the Active Sites. *Cell* **1987**, *49*, 211–220, doi:10.1016/0092-8674(87)90562-9.
189. Golden, B.L. Two Distinct Catalytic Strategies in the Hepatitis Delta Virus Ribozyme Cleavage Reaction. *Biochemistry* **2011**, *50*, 9424–9433, doi:10.1021/bi201157t.
190. Weiner, A.J.; Choo, Q.L.; Wang, K.S.; Govindarajan, S.; Redeker, A.G.; Gerin, J.L.; Houghton, M. 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* **1988**, *62*, 594–599, doi:10.1128/jvi.62.2.594-599.1988.
191. Thomas, J.M.; Beal, P.A. How Do ADARs Bind RNA? New Protein-RNA Structures Illuminate Substrate Recognition by the RNA Editing ADARs. *BioEssays* **2017**, *39*, 1–8, doi:10.1002/bies.201600187.
 192. Patterson, J.B.; Samuel, C.E. Expression and Regulation by Interferon of a Double-Stranded-RNA-Specific Adenosine Deaminase from Human Cells: Evidence for Two Forms of the Deaminase. *Mol Cell Biol* **1995**, *15*, 5376–5388, doi:10.1128/mcb.15.10.5376.
 193. Strehblow, A.; Hallegger, M.; Jantsch, M.F. Nucleocytoplasmic Distribution of Human RNA-Editing Enzyme ADAR1 Is Modulated by Double-Stranded RNA-Binding Domains, a Leucine-Rich Export Signal, and a Putative Dimerization Domain. *Mol Biol Cell* **2002**, *13*, 3822–3835, doi:10.1091/mbc.e02-03-0161.
 194. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu Rev Biochem* **2010**, *79*, 321–349, doi:10.1146/annurev-biochem-060208-105251.
 195. Polson, A.G.; Bass, B.L.; Casey, J.L. RNA Editing of Hepatitis Delta Virus Antigenome by DsRNA-Adenosine Deaminase. **1996**, *380*, 454–456, doi:doi.org/10.1038/380454a0.
 196. Casey, J.L.; Gerin, J.L. Hepatitis D Virus RNA Editing: Specific Modification of Adenosine in the Antigenomic RNA. *J Virol* **1995**, *69*, 7593–7600, doi:10.1128/jvi.69.12.7593-7600.1995.
 197. Mentha, N.; Clément, S.; Negro, F.; Alfaiate, D. A Review on Hepatitis D: From Virology to New Therapies. *J Adv Res* **2019**, *17*, 3–15, doi:10.1016/j.jare.2019.03.009.
 198. Polson, A.G.; Bass, B.L.; Casey, J.L. RNA Editing of Hepatitis Delta Virus Antigenome by DsRNA-Adenosine Deaminase. *Nature* **1996**, *380*, 454–456, doi:10.1038/380454a0.
 199. Chao, M.; Hsieh, S.Y.; Taylor, J. Role of Two Forms of Hepatitis Delta Virus Antigen: Evidence for a Mechanism of Self-Limiting Genome Replication. *J Virol* **1990**, *64*, 5066–5069, doi:10.1128/jvi.64.10.5066-5069.1990.
 200. Kuo, M.Y.; Chao, M.; Taylor, J. Initiation of Replication of the Human Hepatitis Delta Virus Genome from Cloned DNA: Role of Delta Antigen. *J Virol* **1989**, *63*, 1945–1950, doi:10.1128/jvi.63.5.1945-1950.1989.
 201. Huang, W.-H.; Mai, R.-T.; Wu Lee, Y.-H. Transcription Factor YY1 and Its Associated Acetyltransferases CBP and P300 Interact with Hepatitis Delta Antigens and Modulate Hepatitis Delta Virus RNA Replication. *J Virol* **2008**, *82*, 7313–7324, doi:10.1128/JVI.02581-07.

202. Yamaguchi, Y.; Mura, T.; Chanarat, S.; Okamoto, S.; Handa, H. Hepatitis Delta Antigen Binds to the Clamp of RNA Polymerase II and Affects Transcriptional Fidelity. *Genes to Cells* **2007**, *12*, 863–875, doi:10.1111/j.1365-2443.2007.01094.x.
203. Greco-Stewart, V.; Pelchat, M. Interaction of Host Cellular Proteins with Components of the Hepatitis Delta Virus. *Viruses* **2010**, *2*, 189–212, doi:10.3390/v2010189.
204. Yamaguchi, Y.; Filipovska, J.; Yano, K.; Furuya, A.; Inukai, N.; Narita, T.; Wada, T.; Sugimoto, S.; Konarska, M.M.; Handa, H. Stimulation of RNA Polymerase II Elongation by Hepatitis Delta Antigen. *Science (1979)* **2001**, *293*, 124–127, doi:10.1126/science.1057925.
205. Chang, F.L.; Chen, P.J.; Tu, S.J.; Wang, C.J.; Chen, D.S. The Large Form of Hepatitis Delta Antigen Is Crucial for Assembly of Hepatitis Delta Virus. *Proceedings of the National Academy of Sciences* **1991**, *88*, 8490–8494, doi:10.1073/pnas.88.19.8490.
206. Lazinski, D.W.; Taylor, J.M. Relating Structure to Function in the Hepatitis Delta Virus Antigen. *J Virol* **1993**, *67*, 2672–2680, doi:10.1128/jvi.67.5.2672-2680.1993.
207. Chang, M.F.; Chang, S.C.; Chang, C.I.; Wu, K.; Kang, H.Y. Nuclear Localization Signals, but Not Putative Leucine Zipper Motifs, Are Essential for Nuclear Transport of Hepatitis Delta Antigen. *J Virol* **1992**, *66*, 6019–6027, doi:10.1128/jvi.66.10.6019-6027.1992.
208. Alves, C.; Freitas, N.; Cunha, C. Characterization of the Nuclear Localization Signal of the Hepatitis Delta Virus Antigen. *Virology* **2008**, *370*, 12–21, doi:10.1016/j.virol.2007.07.034.
209. Yamaguchi, Y.; Mura, T.; Chanarat, S.; Okamoto, S.; Handa, H. Hepatitis Delta Antigen Binds to the Clamp of RNA Polymerase II and Affects Transcriptional Fidelity. *Genes to Cells* **2007**, *12*, 863–875, doi:10.1111/j.1365-2443.2007.01094.x.
210. Lee, C.-H.; Chang, S.C.; Wu, C.H.H.; Chang, M.-F. 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. *Journal of Biological Chemistry* **2001**, *276*, 8142–8148, doi:10.1074/jbc.M004477200.
211. Glenn, J.S.; Watson, J.A.; Havel, C.M.; White, J.M. Identification of a Prenylation Site in Delta Virus Large Antigen. *Science (1979)* **1992**, *256*, 1331–1333, doi:10.1126/science.1598578.
212. Chang, M.F.; Baker, S.C.; Soe, L.H.; Kamahora, T.; Keck, J.G.; Makino, S.; Govindarajan, S.; Lai, M.M. Human Hepatitis Delta Antigen Is a Nuclear Phosphoprotein with RNA-Binding Activity. *J Virol* **1988**, *62*, 2403–2410, doi:10.1128/jvi.62.7.2403-2410.1988.

213. Mu, J.-J.; Wu, H.-L.; Chiang, B.-L.; Chang, R.-P.; Chen, D.-S.; Chen, P.-J. Characterization of the Phosphorylated Forms and the Phosphorylated Residues of Hepatitis Delta Virus Delta Antigens. *J Virol* **1999**, *73*, 10540–10545, doi:10.1128/JVI.73.12.10540-10545.1999.
214. Bordier, B.B.; Ohkanda, J.; Liu, P.; Lee, S.-Y.; Salazar, F.H.; Marion, P.L.; Ohashi, K.; Meuse, L.; Kay, M.A.; Casey, J.L.; et al. In Vivo Antiviral Efficacy of Prenylation Inhibitors against Hepatitis Delta Virus. *Journal of Clinical Investigation* **2003**, *112*, 407–414, doi:10.1172/JCI17704.
215. Shirvani-Dastgerdi, E.; Amini-Bavil-Olyaei, S.; Alavian, S.M.; Trautwein, C.; Tacke, F. Comprehensive Analysis of Mutations in the Hepatitis Delta Virus Genome Based on Full-Length Sequencing in a Nationwide Cohort Study and Evolutionary Pattern during Disease Progression. *Clinical Microbiology and Infection* **2015**, *21*, 510.e11-510.e23, doi:10.1016/j.cmi.2014.12.008.
216. Karimzadeh, H.; Kiraithe, M.M.; Oberhardt, V.; Salimi Alizei, E.; Bockmann, J.; Schulze zur Wiesch, J.; Budeus, B.; Hoffmann, D.; Wedemeyer, H.; Cornberg, M.; et al. Mutations in Hepatitis D Virus Allow It to Escape Detection by CD8+ T Cells and Evolve at the Population Level. *Gastroenterology* **2019**, *156*, 1820–1833, doi:10.1053/j.gastro.2019.02.003.
217. Wranke, A.; Heidrich, B.; Hardtke, S.; Wedemeyer, H. Current Management of HBV/HDV Coinfection and Future Perspectives. *Curr Hepatol Rep* **2015**, *14*, 284–292, doi:10.1007/s11901-015-0280-8.
218. Heller, T.; Rotman, Y.; Koh, C.; Clark, S.; Haynes-Williams, V.; Chang, R.; McBurney, R.; Schmid, P.; Albrecht, J.; Kleiner, D.E.; et al. Long-Term Therapy of Chronic Delta Hepatitis with Peginterferon Alfa. *Aliment Pharmacol Ther* **2014**, *40*, 93–104, doi:10.1111/apt.12788.
219. Alavian, S.-M.; Tabatabaei, S.V.; Behnava, B.; Rizzetto, M. Standard and Pegylated Interferon Therapy of HDV Infection: A Systematic Review and Meta- Analysis. *J Res Med Sci* **2012**, *17*, 967–974.
220. Glebe, D.; Urban, S.; Knoop, E. v.; Çağ, N.; Krass, P.; Grün, S.; Bulavaite, A.; Sasnauskas, K.; Gerlich, W.H. Mapping of the Hepatitis B Virus Attachment Site by Use of Infection-Inhibiting PreS1 Lipopeptides and Tupaia Hepatocytes. *Gastroenterology* **2005**, *129*, 234–245, doi:10.1053/j.gastro.2005.03.090.
221. Loureiro, D.; Castelnau, C.; Tout, I.; Boyer, N.; Narguet, S.; Menasria Benazzouz, S.; Louis, Z.; Pons-Kerjean, N.; Giuly, N.; Marcellin, P.; et al. New Therapies for Hepatitis Delta Virus Infection. *Liver International* **2021**, *41*, 30–37, doi:10.1111/liv.14838.
222. Kang, C.; Syed, Y.Y. Bulevirtide: First Approval. *Drugs* **2020**, *80*, 1601–1605, doi:10.1007/s40265-020-01400-1.
223. Sandmann, L.; Cornberg, M. Experimental Drugs for the Treatment of Hepatitis D. *J Exp Pharmacol* **2021**, *Volume 13*, 461–468, doi:10.2147/JEP.S235550.

224. Al-Mahtab, M.; Bazinet, M.; Vaillant, A. Safety and Efficacy of Nucleic Acid Polymers in Monotherapy and Combined with Immunotherapy in Treatment-Naive Bangladeshi Patients with HBeAg+ Chronic Hepatitis B Infection. *PLoS One* **2016**, *11*, e0156667, doi:10.1371/journal.pone.0156667.
225. Yuen, M.-F.; Heo, J.; Jang, J.-W.; Yoon, J.-H.; Kweon, Y.-O.; Park, S.-J.; Tami, Y.; You, S.; Yates, P.; Tao, Y.; et al. Safety, Tolerability and Antiviral Activity of the Antisense Oligonucleotide Bepirovirsen in Patients with Chronic Hepatitis B: A Phase 2 Randomized Controlled Trial. *Nat Med* **2021**, *27*, 1725–1734, doi:10.1038/s41591-021-01513-4.
226. Krogsgaard, K.; Kryger, P.; Aldershvile, J.; Andersson, P.; Sørensen, T.I.A.; Nielsen, J.O. Δ -Infection and Suppression of Hepatitis B Virus Replication in Chronic HBsAg Carriers. *Hepatology* **1987**, *7*, 42–45, doi:10.1002/hep.1840070110.
227. Genesca, J.; Jardi, R.; Buti, M.; Vives, L.; Prat, S.; Esteban, J.I.; Esteban, R.; Guardia, J. Hepatitis b Virus Replication in Acute Hepatitis B, Acute Hepatitis B Virus-Hepatitis Delta Virus Coinfection and Acute Hepatitis Delta Superinfection. *Hepatology* **1987**, *7*, 569–572, doi:10.1002/hep.1840070325.
228. Schaper, M.; Rodriguez-Frias, F.; Jardi, R.; Tabernero, D.; Homs, M.; Ruiz, G.; Quer, J.; Esteban, R.; Buti, M. Quantitative Longitudinal Evaluations of Hepatitis Delta Virus RNA and Hepatitis B Virus DNA Shows a Dynamic, Complex Replicative Profile in Chronic Hepatitis B and D. *J Hepatol* **2010**, *52*, 658–664, doi:10.1016/j.jhep.2009.10.036.
229. Wu, J.C.; Chen, P.J.; Kuo, M.Y.; Lee, S.D.; Chen, D.S.; Ting, L.P. Production of Hepatitis Delta Virus and Suppression of Helper Hepatitis B Virus in a Human Hepatoma Cell Line. *J Virol* **1991**, *65*, 1099–1104, doi:10.1128/jvi.65.3.1099-1104.1991.
230. Negro, F.; Korba, B.E.; Forzani, B.; Baroudy, B.M.; Brown, T.L.; Gerin, J.L.; Ponzetto, A. Hepatitis Delta Virus (HDV) and Woodchuck Hepatitis Virus (WHV) Nucleic Acids in Tissues of HDV-Infected Chronic WHV Carrier Woodchucks. *J Virol* **1989**, *63*, 1612–1618, doi:10.1128/jvi.63.4.1612-1618.1989.
231. Sureau, C.; Taylor, J.; Chao, M.; Eichberg, J.W.; Lanford, R.E. Cloned Hepatitis Delta Virus CDNA Is Infectious in the Chimpanzee. *J Virol* **1989**, *63*, 4292–4297, doi:10.1128/jvi.63.10.4292-4297.1989.
232. Rizzetto, M.; Canese, M.G.; Gerin, J.L.; London, W.T.; Sly, D.L.; Purcell, R.H. Transmission of the Hepatitis B Virus-Associated Delta Antigen to Chimpanzees. *Journal of Infectious Diseases* **1980**, *141*, 590–602, doi:10.1093/infdis/141.5.590.
233. Lütgehetmann, M.; Mancke, L. V.; Volz, T.; Helbig, M.; Allweiss, L.; Bornscheuer, T.; Pollok, J.M.; Lohse, A.W.; Petersen, J.; Urban, S.; et al. Humanized Chimeric UPA Mouse Model for the Study of Hepatitis B and D Virus Interactions and

- Preclinical Drug Evaluation. *Hepatology* **2012**, *55*, 685–694, doi:10.1002/hep.24758.
234. Alfaiate, D.; Lucifora, J.; Abeywickrama-Samarakoon, N.; Michelet, M.; Testoni, B.; Cortay, J.-C.; Sureau, C.; Zoulim, F.; Dény, P.; Durantel, D. HDV RNA Replication Is Associated with HBV Repression and Interferon-Stimulated Genes Induction in Super-Infected Hepatocytes. *Antiviral Res* **2016**, *136*, 19–31, doi:10.1016/j.antiviral.2016.10.006.
235. Hepatitis B and Delta Virus: Advances on Studies about Interactions between the Two Viruses and the Infected Hepatocyte. *J Clin Transl Hepatol* **2015**, *3*, 220–229, doi:10.14218/JCTH.2015.00018.
236. Modahl, L.E.; Lai, M.M.C. The Large Delta Antigen of Hepatitis Delta Virus Potently Inhibits Genomic but Not Antigenomic RNA Synthesis: A Mechanism Enabling Initiation of Viral Replication. *J Virol* **2000**, *74*, 7375–7380, doi:10.1128/JVI.74.16.7375-7380.2000.
237. Williams, V.; Brichler, S.; Radjef, N.; Lebon, P.; Goffard, A.; Hober, D.; Fagard, R.; Kremsdorf, D.; Dény, P.; Gordien, E. Hepatitis Delta Virus Proteins Repress Hepatitis B Virus Enhancers and Activate the Alpha/Beta Interferon-Inducible MxA Gene. *Journal of General Virology* **2009**, *90*, 2759–2767, doi:10.1099/vir.0.011239-0.
238. Giersch, K.; Allweiss, L.; Volz, T.; Helbig, M.; Bierwolf, J.; Lohse, A.W.; Pollok, J.M.; Petersen, J.; Dandri, M.; Lütgehetmann, M. Hepatitis Delta Co-Infection in Humanized Mice Leads to Pronounced Induction of Innate Immune Responses in Comparison to HBV Mono-Infection. *J Hepatol* **2015**, *63*, 346–353, doi:10.1016/j.jhep.2015.03.011.
239. Zhang, Z.; Filzmayer, C.; Ni, Y.; Sültmann, H.; Mutz, P.; Hiet, M.-S.; Vondran, F.W.R.; Bartenschlager, R.; Urban, S. Hepatitis D Virus Replication Is Sensed by MDA5 and Induces IFN- β/λ Responses in Hepatocytes. *J Hepatol* **2018**, *69*, 25–35, doi:10.1016/j.jhep.2018.02.021.
240. Zhang, Z.; Urban, S. Interplay between Hepatitis D Virus and the Interferon Response. *Viruses* **2020**, *12*, 1334, doi:10.3390/v12111334.
241. Pétrilli, V.; Dostert, C.; Muruve, D.A.; Tschopp, J. The Inflammasome: A Danger Sensing Complex Triggering Innate Immunity. *Curr Opin Immunol* **2007**, *19*, 615–622, doi:10.1016/j.coi.2007.09.002.
242. Kunzi, M.S.; Pitha, P.M. Interferon Targeted Genes in Host Defense. *Autoimmunity* **2003**, *36*, 457–461, doi:10.1080/08916930310001605855.
243. Schoggins, J.W.; Rice, C.M. Interferon-Stimulated Genes and Their Antiviral Effector Functions. *Curr Opin Virol* **2011**, *1*, 519–525, doi:10.1016/j.coviro.2011.10.008.

244. Li, N.; Zhang, L.; Chen, L.; Feng, W.; Xu, Y.; Chen, F.; Liu, X.; Chen, Z.; Liu, W. MxA Inhibits Hepatitis B Virus Replication by Interaction with Hepatitis B Core Antigen. *Hepatology* **2012**, *56*, 803–811, doi:10.1002/hep.25608.
245. Domingo, E.; Gomez, J. Quasispecies and Its Impact on Viral Hepatitis. *Virus Res* **2007**, *127*, 131–150, doi:10.1016/j.virusres.2007.02.001.
246. Gregori, J.; Perales, C.; Rodriguez-Frias, F.; Esteban, J.I.; Quer, J.; Domingo, E. Viral Quasispecies Complexity Measures. *Virology* **2016**, *493*, 227–237, doi:10.1016/j.virol.2016.03.017.
247. Wang, S.-Y.; Wu, J.-C.; Chiang, T.-Y.; Huang, Y.-H.; Su, C.-W.; Sheen, I.-J. Positive Selection of Hepatitis Delta Antigen in Chronic Hepatitis D Patients. *J Virol* **2007**, *81*, 4438–4444, doi:10.1128/JVI.02847-06.
248. Barzon, L.; Lavezzo, E.; Militello, V.; Toppo, S.; Palù, G. Applications of Next-Generation Sequencing Technologies to Diagnostic Virology. *Int J Mol Sci* **2011**, *12*, 7861–7884, doi:10.3390/ijms12117861.
249. Caligiuri, P. Overview of Hepatitis B Virus Mutations and Their Implications in the Management of Infection. *World J Gastroenterol* **2016**, *22*, 145, doi:10.3748/wjg.v22.i1.145.
250. Zhou, Y.; Holmes, E.C. Bayesian Estimates of the Evolutionary Rate and Age of Hepatitis B Virus. *J Mol Evol* **2007**, *65*, 197–205, doi:10.1007/s00239-007-0054-1.
251. Wang, H.-Y.; Chien, M.-H.; Huang, H.-P.; Chang, H.-C.; Wu, C.-C.; Chen, P.-J.; Chang, M.-H.; Chen, D.-S. Distinct Hepatitis B Virus Dynamics in the Immunotolerant and Early Immunoclearance Phases. *J Virol* **2010**, *84*, 3454–3463, doi:10.1128/JVI.02164-09.
252. Osiowy, C.; Giles, E.; Tanaka, Y.; Mizokami, M.; Minuk, G.Y. Molecular Evolution of Hepatitis B Virus over 25 Years. *J Virol* **2006**, *80*, 10307–10314, doi:10.1128/JVI.00996-06.
253. Lin, Y.-Y.; Liu, C.; Chien, W.-H.; Wu, L.-L.; Tao, Y.; Wu, D.; Lu, X.; Hsieh, C.-H.; Chen, P.-J.; Wang, H.-Y.; et al. New Insights into the Evolutionary Rate of Hepatitis B Virus at Different Biological Scales. *J Virol* **2015**, *89*, 3512–3522, doi:10.1128/JVI.03131-14.
254. Whalley, S.A.; Murray, J.M.; Brown, D.; Webster, G.J.M.; Emery, V.C.; Dusheiko, G.M.; Perelson, A.S. Kinetics of Acute Hepatitis B Virus Infection in Humans. *Journal of Experimental Medicine* **2001**, *193*, 847–854, doi:10.1084/jem.193.7.847.
255. Horvat, R.T. Diagnostic and Clinical Relevance of HBV Mutations. *Lab Med* **2011**, *42*, 488–496, doi:10.1309/LM7SF4QZMEG5LVPN.

256. Bishop, K.N.; Holmes, R.K.; Sheehy, A.M.; Malim, M.H. APOBEC-Mediated Editing of Viral RNA. *Science (1979)* **2004**, *305*, 645–645, doi:10.1126/science.1100658.
257. Janahi, E.M.; McGarvey, M.J. The Inhibition of Hepatitis B Virus by APOBEC Cytidine Deaminases. *J Viral Hepat* **2013**, *20*, 821–828, doi:10.1111/jvh.12192.
258. Bouzidi, M.S.; Caval, V.; Suspène, R.; Hallez, C.; Pineau, P.; Wain-Hobson, S.; Vartanian, J.-P. APOBEC3DE Antagonizes Hepatitis B Virus Restriction Factors APOBEC3F and APOBEC3G. *J Mol Biol* **2016**, *428*, 3514–3528, doi:10.1016/j.jmb.2016.05.022.
259. Noguchi, C.; Imamura, M.; Tsuge, M.; Hiraga, N.; Mori, N.; Miki, D.; Kimura, T.; Takahashi, S.; Fujimoto, Y.; Ochi, H.; et al. G-to-A Hypermutation in Hepatitis B Virus (HBV) and Clinical Course of Patients with Chronic HBV Infection. *J Infect Dis* **2009**, *199*, 1599–1607, doi:10.1086/598951.
260. Abe, H.; Ochi, H.; Maekawa, T.; Hatakeyama, T.; Tsuge, M.; Kitamura, S.; Kimura, T.; Miki, D.; Mitsui, F.; Hiraga, N.; et al. Effects of Structural Variations of *APOBEC3A* and *APOBEC3B* Genes in Chronic Hepatitis B Virus Infection. *Hepatology Research* **2009**, *39*, 1159–1168, doi:10.1111/j.1872-034X.2009.00566.x.
261. Chen, Z.; Eggerman, T.L.; Bocharov, A. V.; Baranova, I.N.; Vishnyakova, T.G.; Patterson, A.P. APOBEC3-Induced Mutation of the Hepatitis Virus B DNA Genome Occurs during Its Viral RNA Reverse Transcription into (–)-DNA. *Journal of Biological Chemistry* **2021**, *297*, 100889, doi:10.1016/j.jbc.2021.100889.
262. Ren, F.; Li, W.; Xiang, A.; Wang, L.; Li, M.; Guo, Y. Distribution and Difference of APOBEC-induced Mutations in the TpCpW Context of HBV DNA between HCC and Non-HCC. *J Med Virol* **2020**, *92*, 53–61, doi:10.1002/jmv.25572.
263. Miyakawa, Y.; Mizokami, M. Classifying Hepatitis B Virus Genotypes. *Intervirology* **2003**, *46*, 329–338, doi:10.1159/000074988.
264. Kramvis, A. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology* **2014**, *57*, 141–150, doi:10.1159/000360947.
265. McNaughton, A.L.; D’Arienzo, V.; Ansari, M.A.; Lumley, S.F.; Littlejohn, M.; Revill, P.; McKeating, J.A.; Matthews, P.C. Insights From Deep Sequencing of the HBV Genome—Unique, Tiny, and Misunderstood. *Gastroenterology* **2019**, *156*, 384–399, doi:10.1053/j.gastro.2018.07.058.
266. Elizalde, M.M.; Tadey, L.; Mammana, L.; Quarleri, J.F.; Campos, R.H.; Flichman, D.M. Biological Characterization of Hepatitis B Virus Genotypes: Their Role in Viral Replication and Antigen Expression. *Front Microbiol* **2021**, *12*, doi:10.3389/fmicb.2021.758613.

267. McMahon, B.J. The Influence of Hepatitis B Virus Genotype and Subgenotype on the Natural History of Chronic Hepatitis B. *Hepatol Int* **2009**, *3*, 334–342, doi:10.1007/s12072-008-9112-z.
268. Tuttleman, J.S.; Pourcel, C.; Summers, J. Formation of the Pool of Covalently Closed Circular Viral DNA in Hepadnavirus-Infected Cells. *Cell* **1986**, *47*, 451–460, doi:10.1016/0092-8674(86)90602-1.
269. Liu, X.; Bushnell, D.A.; Kornberg, R.D. RNA Polymerase II Transcription: Structure and Mechanism. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2013**, *1829*, 2–8, doi:10.1016/j.bbagr.2012.09.003.
270. Homs, M.; Rodriguez-Frias, F.; Gregori, J.; Ruiz, A.; Reimundo, P.; Casillas, R.; Taberner, D.; Godoy, C.; Barakat, S.; Quer, J.; et al. Evidence of an Exponential Decay Pattern of the Hepatitis Delta Virus Evolution Rate and Fluctuations in Quasispecies Complexity in Long-Term Studies of Chronic Delta Infection. *PLoS One* **2016**, *11*, e0158557, doi:10.1371/journal.pone.0158557.
271. Chang, J. In Vivo RNA-Directed Transcription, with Template Switching, by a Mammalian RNA Polymerase. *EMBO J* **2002**, *21*, 157–164, doi:10.1093/emboj/21.1.157.
272. Wu, J.C.; Chiang, T.Y.; Shiue, W.K.; Wang, S.Y.; Sheen, I.J.; Huang, Y.H.; Syu, W.J. Recombination of Hepatitis D Virus RNA Sequences and Its Implications. *Mol Biol Evol* **1999**, *16*, 1622–1632, doi:10.1093/oxfordjournals.molbev.a026075.
273. Sopena, S.; Godoy, C.; Taberner, D.; Homs, M.; Gregori, J.; Riveiro-Barciela, M.; Ruiz, A.; Esteban, R.; Buti, M.; Rodríguez-Frías, F. Quantitative Characterization of Hepatitis Delta Virus Genome Edition by Next-Generation Sequencing. *Virus Res* **2018**, *243*, 52–59, doi:10.1016/j.virusres.2017.10.003.
274. Dény, P. Hepatitis Delta Virus Genetic Variability: From Genotypes I, II, III to Eight Major Clades? *Hepatitis Delta Virus* 151–171.
275. Miao, Z.; Zhang, S.; Ou, X.; Li, S.; Ma, Z.; Wang, W.; Peppelenbosch, M.P.; Liu, J.; Pan, Q. Estimating the Global Prevalence, Disease Progression, and Clinical Outcome of Hepatitis Delta Virus Infection. *J Infect Dis* **2020**, *221*, 1677–1687, doi:10.1093/infdis/jiz633.
276. Vogt, A.; Wohlfart, S.; Urban, S.; Mier, W. Medical Advances in Hepatitis D Therapy: Molecular Targets. *Int J Mol Sci* **2022**, *23*, 10817, doi:10.3390/ijms231810817.
277. Bahoussi, A.N.; Wang, P.-H.; Guo, Y.-Y.; Rabbani, N.; Wu, C.; Xing, L. Global Distribution and Natural Recombination of Hepatitis D Virus: Implication of Kyrgyzstan Emerging HDVs in the Clinical Outcomes. *Viruses* **2022**, *14*, 1467, doi:10.3390/v14071467.
278. Gomes-Gouvêa, M.S.; Soares, M.C.P.; Bensabath, G.; de Carvalho-Mello, I.M.V.G.; Brito, E.M.F.; Souza, O.S.C.; Queiroz, A.T.L.; Carrilho, F.J.; Pinho, J.R.R.

- Hepatitis B Virus and Hepatitis Delta Virus Genotypes in Outbreaks of Fulminant Hepatitis (Labrea Black Fever) in the Western Brazilian Amazon Region. *Journal of General Virology* **2009**, *90*, 2638–2643, doi:10.1099/vir.0.013615-0.
279. Su, C.; Huang, Y.; Huo, T.; Shih, H.H.; Sheen, I.; Chen, S.; Lee, P.; Lee, S.; Wu, J. Genotypes and Viremia of Hepatitis B and D Viruses Are Associated With Outcomes of Chronic Hepatitis D Patients. *Gastroenterology* **2006**, *130*, 1625–1635, doi:10.1053/j.gastro.2006.01.035.
280. Roulot, D.; Brichler, S.; Layese, R.; BenAbdesselam, Z.; Zoulim, F.; Thibault, V.; Scholtes, C.; Roche, B.; Castelnau, C.; Poynard, T.; et al. Origin, HDV Genotype and Persistent Viremia Determine Outcome and Treatment Response in Patients with Chronic Hepatitis Delta. *J Hepatol* **2020**, *73*, 1046–1062, doi:10.1016/j.jhep.2020.06.038.
281. Ardui, S.; Ameer, A.; Vermeesch, J.R.; Hestand, M.S. Single Molecule Real-Time (SMRT) Sequencing Comes of Age: Applications and Utilities for Medical Diagnostics. *Nucleic Acids Res* **2018**, *46*, 2159–2168, doi:10.1093/nar/gky066.
282. Illumina Support Center [Internet]. https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/translations/miseq-system-guide-1000000061014-esp.pdf.
283. Paul-Ehrlich-Institute [Internet]. © Paul-Ehrlich-Institut; [Cited 2022 Mar 13]. Available from: https://www.pei.de/SharedDocs/Downloads/EN/Regulation-En/Referencematerial/7657-12-ifu.pdf?__blob=publicationFile&v=3.
284. Kimura, M. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *J Mol Evol* **1980**, *16*, 111–120, doi:10.1007/BF01731581.
285. MagNA Pure Total Nucleic Acid Isolation. LifeScience Roche [Internet]. Roche Molecular Systems, Inc; C1999 [Cited 2022 Mar 13]. Available from: https://www.lifescience.roche.com/en_es/products/magna-pure-lc-total-nucleic-acid-isolation-kit.html#details.
286. Chudy, Michael, Hanschmann, Kay-Martin, Bozdayi, Mithat, Kreß, Julia, C., Nübling, Micha. et al. (2013). Collaborative Study to Establish a World Health Organization International Standard for Hepatitis D Virus RNA for Nucleic Acid Amplification Technique (NAT)-Based Assays. World Health Organization.
287. Taylor, J.M. THE STRUCTURE AND REPLICATION OF HEPATITIS DELTA VIRUS. *Annu Rev Microbiol* **1992**, *46*, 253–276, doi:10.1146/annurev.mi.46.100192.001345.

288. Soria, M.E.; Gregori, J.; Chen, Q.; García-Cehic, D.; Llorens, M.; de Ávila, A.I.; Beach, N.M.; Domingo, E.; Rodríguez-Frías, F.; Buti, M.; et al. Pipeline for Specific Subtype Amplification and Drug Resistance Detection in Hepatitis C Virus. *BMC Infect Dis* **2018**, *18*, 1–15, doi:10.1186/s12879-018-3356-6.
289. Magoc, T.; Salzberg, S.L. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* **2011**, *27*, 2957–2963, doi:10.1093/bioinformatics/btr507.
290. Ewing, B.; Hillier, L.; Wendl, M.C.; Green, P. Base-Calling of Automated Sequencer Traces Using *Phred*. I. Accuracy Assessment. *Genome Res* **1998**, *8*, 175–185, doi:10.1101/gr.8.3.175.
291. Schneider, T.D. Information Content of Individual Genetic Sequences. *J Theor Biol* **1997**, *189*, 427–441, doi:10.1006/jtbi.1997.0540.
292. Schneider, T.D.; Stephens, R.M. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res* **1990**, *18*, 6097–6100, doi:10.1093/nar/18.20.6097.
293. Domingo, E.; Perales, C. From Quasispecies Theory to Viral Quasispecies: How Complexity Has Permeated Virology. *Math Model Nat Phenom* **2012**, *7*, 105–122, doi:10.1051/mmnp/20127508.
294. Wolinsky, S.M.; Korber, B.T.M.; Neumann, A.U.; Daniels, M.; Kunstman, K.J.; Whetsell, A.J.; Furtado, M.R.; Cao, Y.; Ho, D.D.; Safrit, J.T.; et al. Adaptive Evolution of Human Immunodeficiency Virus-Type 1 During the Natural Course of Infection. *Science (1979)* **1996**, *272*, 537–542, doi:10.1126/science.272.5261.537.
295. Gregori, J.; Perales, C.; Rodríguez-Frías, F.; Esteban, J.I.; Quer, J.; Domingo, E. Viral Quasispecies Complexity Measures. *Virology* **2016**, *493*, 227–237, doi:10.1016/j.virol.2016.03.017.
296. Thomas, R.H. Molecular Evolution and Phylogenetics. *Heredity (Edinb)* **2001**, *86*, 385–385, doi:10.1046/j.1365-2540.2001.0923a.x.
297. Wu, T.T.; Netter, H.J.; Lazinski, D.W.; Taylor, J.M. Effects of Nucleotide Changes on the Ability of Hepatitis Delta Virus to Transcribe, Process, and Accumulate Unit-Length, Circular RNA. *J Virol* **1997**, *71*, 5408–5414, doi:10.1128/jvi.71.7.5408-5414.1997.
298. Wang, W.; Jia, Y.; Li, Y.; Jing, C.; Guo, X.; Shang, X.; Zhao, C.; Wang, T. RETRACTED ARTICLE: Impact of Different Promoters, Promoter Mutation, and an Enhancer on Recombinant Protein Expression in CHO Cells. *Sci Rep* **2017**, *7*, 10416, doi:10.1038/s41598-017-10966-y.
299. Chang, M.F.; Chen, C.H.; Lin, S.L.; Chen, C.J.; Chang, S.C. Functional Domains of Delta Antigens and Viral RNA Required for RNA Packaging of Hepatitis Delta Virus. *J Virol* **1995**, *69*, 2508–2514, doi:10.1128/jvi.69.4.2508-2514.1995.

300. Guidotti, L.G.; Matzke, B.; Schaller, H.; Chisari, F. V High-Level Hepatitis B Virus Replication in Transgenic Mice. *J Virol* **1995**, *69*, 6158–6169, doi:10.1128/jvi.69.10.6158-6169.1995.
301. Suárez-Amarán, L.; Usai, C.; di Scala, M.; Godoy, C.; Ni, Y.; Hommel, M.; Palomo, L.; Segura, V.; Olagüe, C.; Vales, A.; et al. A New HDV Mouse Model Identifies Mitochondrial Antiviral Signaling Protein (MAVS) as a Key Player in IFN- β Induction. *J Hepatol* **2017**, *67*, 669–679, doi:10.1016/j.jhep.2017.05.010.
302. Scholtès, C.; Hamilton, A.T.; Plissonnier, M.-L.; Charre, C.; Scott, B.; Wang, L.; Berby, F.; French, J.; Testoni, B.; Blair, A.; et al. Performance of the Cobas® HBV RNA Automated Investigational Assay for the Detection and Quantification of Circulating HBV RNA in Chronic HBV Patients. *Journal of Clinical Virology* **2022**, *150–151*, 105150, doi:10.1016/j.jcv.2022.105150.
303. Gregori, J.; Soria, M.E.; Gallego, I.; Guerrero-Murillo, M.; Esteban, J.I.; Quer, J.; Perales, C.; Domingo, E. Rare Haplotype Load as Marker for Lethal Mutagenesis. *PLoS One* **2018**, *13*, e0204877, doi:10.1371/journal.pone.0204877.
304. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. **2017**. URL <https://www.R-project.org/>.
305. Maurel, M.-C.; Leclerc, F.; Vergne, J.; Zaccari, G. RNA Back and Forth: Looking through Ribozyme and Viroid Motifs. *Viruses* **2019**, *11*, 283, doi:10.3390/v11030283.
306. Deterding, K.; Wedemeyer, H. Beyond Pegylated Interferon-Alpha: New Treatments for Hepatitis Delta. *AIDS Rev* **2019**, *21*, doi:10.24875/AIDSRev.19000080.
307. Heidrich, B.; Yurdaydin, C.; Kabaçam, G.; Ratsch, B.A.; Zachou, K.; Bremer, B.; Dalekos, G.N.; Erhardt, A.; Tabak, F.; Yalcin, K.; et al. Late HDV RNA Relapse after Peginterferon Alpha-Based Therapy of Chronic Hepatitis Delta. *Hepatology* **2014**, *60*, 87–97, doi:10.1002/hep.27102.
308. Loureiro, D.; Castelnau, | Corinne; Tout, I.; Boyer, N.; Narguet, S.; Menasria Benazzouz, S.; Louis, Z.; Pons-Kerjean, N.; Giuly, N.; Marcellin, P.; et al. New Therapies for Hepatitis Delta Virus Infection. *Liver International* **2021**, *41*, 30–37, doi:10.1111/liv.14838.
309. Asselah, T.; Loureiro, D.; le Gal, F.; Narguet, S.; Brichtler, S.; Bouton, V.; Abazid, M.; Boyer, N.; Giuly, N.; Gerber, A.; et al. Early Virological Response in Six Patients with Hepatitis D Virus Infection and Compensated Cirrhosis Treated with Bulevirtide in Real-Life. *Liver International* **2021**, *41*, 1509–1517, doi:10.1111/LIV.14950.
310. Anolli, M.P.; Degasperi, E.; Allweiss, L.; Sangiovanni, A.; Maggioni, M.; Scholtes, C.; Oberhardt, V.; Neumann-Haefelin, C.; Dandri, M.; Zoulim, F.; et al. A 3-Year

Course of Bulevirtide Monotherapy May Cure HDV Infection in Patients with Cirrhosis. *J Hepatol* **2023**, *78*, 876–880, doi:10.1016/j.jhep.2022.12.023.

311. Chang, J.; Taylor, J.M. Susceptibility of Human Hepatitis Delta Virus RNAs to Small Interfering RNA Action. *J Virol* **2003**, *77*, 9728–9731, doi:10.1128/JVI.77.17.9728-9731.2003.
312. Li, H.; Mao, Q.; Li, Q. [Inhibitory Effect of Replication and Expression of HDV by Antisense Oligodeoxynucleotides in H1 Delta 9 Cell]. *Zhonghua Gan Zang Bing Za Zhi* **1999**, *7*, 13–14.
313. Ye, X.; Tateno, C.; Thi, E.P.; Kakuni, M.; Snead, N.M.; Ishida, Y.; Barnard, T.R.; Sofia, M.J.; Shimada, T.; Lee, A.C.H. Hepatitis B Virus Therapeutic Agent ARB-1740 Has Inhibitory Effect on Hepatitis Delta Virus in a New Dually-Infected Humanized Mouse Model. *ACS Infect Dis* **2019**, *5*, 738–749, doi:10.1021/acsinfecdis.8b00192.
314. Chad, Y.-C.; Chang, M.-F.; Gust, I.; Lai, M.M.C. Sequence Conservation and Divergence of Hepatitis δ Virus RNA. *Virology* **1990**, *178*, 384–392, doi:10.1016/0042-6822(90)90335-O.
315. Shirvani-Dastgerdi, E.; Amini-Bavil-Olyaei, S.; Alavian, S.M.; Trautwein, C.; Tacke, F. Comprehensive Analysis of Mutations in the Hepatitis Delta Virus Genome Based on Full-Length Sequencing in a Nationwide Cohort Study and Evolutionary Pattern during Disease Progression. *Clinical Microbiology and Infection* **2015**, *21*, 510.e11–510.e23, doi:10.1016/j.cmi.2014.12.008.
316. Perrotta, A.T.; Been, M.D. A Pseudoknot-like Structure Required for Efficient Self-Cleavage of Hepatitis Delta Virus RNA. *Nature* **1991**, *350*, 434–436, doi:10.1038/350434a0.
317. Chadalavada, D.M.; Cerrone-Szkal, A.L.; Bevilacqua, P.C. Wild-Type Is the Optimal Sequence of the HDV Ribozyme under Cotranscriptional Conditions. *RNA* **2007**, *13*, 2189–2201, doi:10.1261/rna.778107.
318. Reymond, C.; Ouellet, J.; Bisailon, M.; Perreault, J.-P. Examination of the Folding Pathway of the Antigenomic Hepatitis Delta Virus Ribozyme Reveals Key Interactions of the L3 Loop. *RNA* **2007**, *13*, 44–54, doi:10.1261/rna.263407.
319. Tanner, N.K.; Schaff, S.; Thill, G.; Petit-Koskas, E.; Crain-Denoyelle, A.-M.; Westhof, E. A Three-Dimensional Model of Hepatitis Delta Virus Ribozyme Based on Biochemical and Mutational Analyses. *Current Biology* **1994**, *4*, 488–498, doi:10.1016/S0960-9822(00)00109-3.
320. Perrotta, A.T.; Been, M.D. Core Sequences and a Cleavage Site Wobble Pair Required for HDV Antigenomic Ribozyme Self-Cleavage. *Nucleic Acids Res* **1996**, *24*, 1314–1321, doi:10.1093/nar/24.7.1314.
321. Timm, J.; Lauer, G.M.; Kavanagh, D.G.; Sheridan, I.; Kim, A.Y.; Lucas, M.; Pillay, T.; Ouchi, K.; Reyor, L.L.; zur Wiesch, J.S.; et al. CD8 Epitope Escape and

- Reversion in Acute HCV Infection. *Journal of Experimental Medicine* **2004**, *200*, 1593–1604, doi:10.1084/jem.20041006.
322. Leslie, A.J.; Pfafferott, K.J.; Chetty, P.; Draenert, R.; Addo, M.M.; Feeney, M.; Tang, Y.; Holmes, E.C.; Allen, T.; Prado, J.G.; et al. HIV Evolution: CTL Escape Mutation and Reversion after Transmission. *Nat Med* **2004**, *10*, 282–289, doi:10.1038/nm992.
 323. Lee, C.-M.; Bih, F.-Y.; Chao, Y.-C.; Govindarajan, S.; Lai, M.M.C. Evolution of Hepatitis Delta Virus Rna during Chronic Infection. *Virology* **1992**, *188*, 265–273, doi:10.1016/0042-6822(92)90756-F.
 324. Berg, M.G.; Olivo, A.; Forberg, K.; Harris, B.J.; Yamaguchi, J.; Shirazi, R.; Gozlan, Y.; Sauleda, S.; Kaptue, L.; Rodgers, M.A.; et al. Advanced Molecular Surveillance Approaches for Characterization of Blood Borne Hepatitis Viruses. *PLoS One* **2020**, *15*, e0236046, doi:10.1371/journal.pone.0236046.
 325. Dziri, S.; Rodriguez, C.; Gerber, A.; Brichler, S.; Alloui, C.; Roulot, D.; Dény, P.; Pawlotsky, J.M.; Gordien, E.; Le Gal, F. Variable In Vivo Hepatitis D Virus (HDV) RNA Editing Rates According to the HDV Genotype. *Viruses* **2021**, *13*, 1572, doi:10.3390/v13081572.
 326. Kohsar, M.; Landahl, J.; Neumann-Haefelin, C.; Schulze Zur Wiesch, J. *Human Hepatitis D Virus-Specific T Cell Epitopes*; 2021;
 327. Pacin-Ruiz, B.; Cortese, M.F.; Tabernero, D.; Sopena, S.; Gregori, J.; García-García, S.; Casillas, R.; Najarro, A.; Aldama, U.; Palom, A.; et al. Inspecting the Ribozyme Region of Hepatitis Delta Virus Genotype 1: Conservation and Variability. *Viruses* **2022**, *14*, 215, doi:10.3390/V14020215.
 328. Lucifora, J.; Alfaiate, D.; Pons, C.; Michelet, M.; Ramirez, R.; Fusil, F.; Amirache, F.; Rossi, A.; Legrand, A.-F.; Charles, E.; et al. Hepatitis D Virus Interferes with Hepatitis B Virus RNA Production via Interferon-Dependent and -Independent Mechanisms. *J Hepatol* **2023**, doi:10.1016/j.jhep.2023.01.005.
 329. Yang, Z.; Sun, B.; Xiang, J.; Wu, H.; Kan, S.; Hao, M.; Chang, L.; Liu, H.; Wang, D.; Liu, W. Role of Epigenetic Modification in Interferon Treatment of Hepatitis B Virus Infection. *Front Immunol* **2022**, *13*, doi:10.3389/fimmu.2022.1018053.
 330. Cheng, J.; Zhao, Q.; Zhou, Y.; Tang, L.; Sheraz, M.; Chang, J.; Guo, J.-T. Interferon Alpha Induces Multiple Cellular Proteins That Coordinately Suppress Hepadnaviral Covalently Closed Circular DNA Transcription. *J Virol* **2020**, *94*, doi:10.1128/JVI.00442-20.
 331. Lucifora, J.; Xia, Y.; Reisinger, F.; Zhang, K.; Stadler, D.; Cheng, X.; Sprinzl, M.F.; Koppensteiner, H.; Makowska, Z.; Volz, T.; et al. Specific and Nonhepatotoxic Degradation of Nuclear Hepatitis B Virus CccDNA. *Science (1979)* **2014**, *343*, 1221–1228, doi:10.1126/science.1243462.
 332. Stadler, D.; Kächele, M.; Jones, A.N.; Hess, J.; Urban, C.; Schneider, J.; Xia, Y.; Oswald, A.; Nebioglu, F.; Bester, R.; et al. Interferon-induced Degradation of

- the Persistent Hepatitis B Virus CccDNA Form Depends on ISG20. *EMBO Rep* **2021**, *22*, doi:10.15252/embr.201949568.
333. Zoulim, F.; Lebossé, F.; Levrero, M. Current Treatments for Chronic Hepatitis B Virus Infections. *Curr Opin Virol* **2016**, *18*, 109–116, doi:10.1016/j.coviro.2016.06.004.
334. Sadler, H.A.; Stenglein, M.D.; Harris, R.S.; Mansky, L.M. APOBEC3G Contributes to HIV-1 Variation through Sublethal Mutagenesis. *J Virol* **2010**, *84*, 7396–7404, doi:10.1128/JVI.00056-10.
335. Köck, J.; Blum, H.E. Hypermutation of Hepatitis B Virus Genomes by APOBEC3G, APOBEC3C and APOBEC3H. *Journal of General Virology* **2008**, *89*, 1184–1191, doi:10.1099/vir.0.83507-0.
336. Vartanian, J.-P.; Henry, M.; Marchio, A.; Suspène, R.; Aynaud, M.-M.; Guétard, D.; Cervantes-Gonzalez, M.; Battiston, C.; Mazzaferro, V.; Pineau, P.; et al. Massive APOBEC3 Editing of Hepatitis B Viral DNA in Cirrhosis. *PLoS Pathog* **2010**, *6*, e1000928, doi:10.1371/journal.ppat.1000928.
337. Tatsukawa, M.; Takaki, A.; Shiraha, H.; Koike, K.; Iwasaki, Y.; Kobashi, H.; Fujioka, S.-I.; Sakaguchi, K.; Yamamoto, K. Hepatitis B Virus Core Promoter Mutations G1613A and C1653T Are Significantly Associated with Hepatocellular Carcinoma in Genotype C HBV-Infected Patients. *BMC Cancer* **2011**, *11*, 458, doi:10.1186/1471-2407-11-458.
338. Salpini, R.; Battisti, A.; Piermatteo, L.; Carioti, L.; Anastasiou, O.E.; Gill, U.S.; Di Carlo, D.; Colagrossi, L.; Duca, L.; Bertoli, A.; et al. Key Mutations in the C-Terminus of the HBV Surface Glycoprotein Correlate with Lower HBsAg Levels *in Vivo*, Hinder HBsAg Secretion *in Vitro* and Reduce HBsAg Structural Stability in the Setting of HBeAg-Negative Chronic HBV Genotype-D Infection. *Emerg Microbes Infect* **2020**, *9*, 928–939, doi:10.1080/22221751.2020.1757998.
339. Chong, C.K.; Cheng, C.Y.S.; Tsoi, S.Y.J.; Huang, F.-Y.; Liu, F.; Fung, J.; Seto, W.-K.; Lai, K.K.-Y.; Lai, C.-L.; Yuen, M.-F.; et al. HBV X Protein Mutations Affect HBV Transcription and Association of Histone-Modifying Enzymes with Covalently Closed Circular DNA. *Sci Rep* **2020**, *10*, 802, doi:10.1038/s41598-020-57637-z.
340. Karimova, M.; Beschorner, N.; Dammermann, W.; Chemnitz, J.; Indenbirken, D.; Bockmann, J.-H.; Grundhoff, A.; Lüth, S.; Buchholz, F.; Wiesch, J.S. zur; et al. CRISPR/Cas9 Nickase-Mediated Disruption of Hepatitis B Virus Open Reading Frame S and X. *Sci Rep* **2015**, *5*, 13734, doi:10.1038/srep13734.
341. Datta, S.; Banerjee, A.; Chandra, P.K.; Biswas, A.; Panigrahi, R.; Mahapatra, P.K.; Panda, C.K.; Chakrabarti, S.; Bhattacharya, S.K.; Chakravarty, R. Analysis of Hepatitis B Virus X Gene Phylogeny, Genetic Variability and Its Impact on Pathogenesis: Implications in Eastern Indian HBV Carriers. *Virology* **2008**, *382*, 190–198, doi:10.1016/j.virol.2008.09.007.

342. Chen, J.; MacCarthy, T. The Preferred Nucleotide Contexts of the AID/APOBEC Cytidine Deaminases Have Differential Effects When Mutating Retrotransposon and Virus Sequences Compared to Host Genes. *PLoS Comput Biol* **2017**, *13*, e1005471, doi:10.1371/journal.pcbi.1005471.
343. Caval, V.; Jiao, W.; Berry, N.; Khalfi, P.; Pitré, E.; Thiers, V.; Vartanian, J.-P.; Wain-Hobson, S.; Suspène, R. Mouse APOBEC1 Cytidine Deaminase Can Induce Somatic Mutations in Chromosomal DNA. *BMC Genomics* **2019**, *20*, 858, doi:10.1186/s12864-019-6216-x.
344. Yu, K. AID Function in Somatic Hypermutation and Class Switch Recombination. *Acta Biochim Biophys Sin (Shanghai)* **2022**, *54*, 759–766, doi:10.3724/abbs.2022070.
345. Sharma, S.; Patnaik, S.K.; Thomas Taggart, R.; Kannisto, E.D.; Enriquez, S.M.; Gollnick, P.; Baysal, B.E. APOBEC3A Cytidine Deaminase Induces RNA Editing in Monocytes and Macrophages. *Nat Commun* **2015**, *6*, 6881, doi:10.1038/ncomms7881.
346. Kim, K.; Calabrese, P.; Wang, S.; Qin, C.; Rao, Y.; Feng, P.; Chen, X.S. The Roles of APOBEC-Mediated RNA Editing in SARS-CoV-2 Mutations, Replication and Fitness. *Sci Rep* **2022**, *12*, 14972, doi:10.1038/s41598-022-19067-x.
347. Di Giorgio, S.; Martignano, F.; Torcia, M.G.; Mattiuz, G.; Conticello, S.G. Evidence for Host-Dependent RNA Editing in the Transcriptome of SARS-CoV-2. *Sci Adv* **2020**, *6*, doi:10.1126/sciadv.abb5813.
348. Ratcliff, J.; Simmonds, P. Potential APOBEC-Mediated RNA Editing of the Genomes of SARS-CoV-2 and Other Coronaviruses and Its Impact on Their Longer Term Evolution. *Virology* **2021**, *556*, 62–72, doi:10.1016/j.virol.2020.12.018.
349. Liang, G.; Kitamura, K.; Wang, Z.; Liu, G.; Chowdhury, S.; Fu, W.; Koura, M.; Wakae, K.; Honjo, T.; Muramatsu, M. RNA Editing of Hepatitis B Virus Transcripts by Activation-Induced Cytidine Deaminase. *Proceedings of the National Academy of Sciences* **2013**, *110*, 2246–2251, doi:10.1073/pnas.1221921110.
350. Meier, M.-A.; Suslov, A.; Ketterer, S.; Heim, M.H.; Wieland, S.F. Hepatitis B Virus Covalently Closed Circular DNA Homeostasis Is Independent of the Lymphotoxin Pathway during Chronic HBV Infection. *J Viral Hepat* **2017**, *24*, 662–671, doi:10.1111/jvh.12689.
351. Chen, Q.; Lu, X.; Zhang, X. Noncanonical NF- κ B Signaling Pathway in Liver Diseases. *J Clin Transl Hepatol* **2020**, *000*, 1–9, doi:10.14218/JCTH.2020.00063.
352. Kim, H.R.; Lee, S.H.; Jung, G. The Hepatitis B Viral X Protein Activates NF- κ B Signaling Pathway through the up-Regulation of TBK1. *FEBS Lett* **2010**, *584*, 525–530, doi:10.1016/j.febslet.2009.11.091.

353. You, L.-R.; Chen, C.-M.; Lee, Y.-H.W. Hepatitis C Virus Core Protein Enhances NF- κ B Signal Pathway Triggering by Lymphotoxin- β Receptor Ligand and Tumor Necrosis Factor Alpha. *J Virol* **1999**, *73*, 1672–1681, doi:10.1128/JVI.73.2.1672-1681.1999.
354. Faure-Dupuy, S.; Riedl, T.; Rolland, M.; Hizir, Z.; Reisinger, F.; Neuhaus, K.; Schuehle, S.; Remouchamps, C.; Gillet, N.; Schönung, M.; et al. Control of APOBEC3B Induction and CccDNA Decay by NF- κ B and MiR-138-5p. *JHEP Reports* **2021**, *3*, 100354, doi:10.1016/j.jhepr.2021.100354.
355. Real, C.I.; Lu, M.; Liu, J.; Huang, X.; Trippler, M.; Hossbach, M.; Deckert, J.; Jahn-Hofmann, K.; Ickenstein, L.M.; John, M.J.; et al. Hepatitis B Virus Genome Replication Triggers Toll-like Receptor 3-Dependent Interferon Responses in the Absence of Hepatitis B Surface Antigen. *Sci Rep* **2016**, *6*, 24865, doi:10.1038/srep24865.
356. Bai, L.; Zhang, X.; Kozlowski, M.; Li, W.; Wu, M.; Liu, J.; Chen, L.; Zhang, J.; Huang, Y.; Yuan, Z. Extracellular Hepatitis B Virus RNAs Are Heterogeneous in Length and Circulate as Capsid-Antibody Complexes in Addition to Virions in Chronic Hepatitis B Patients. *J Virol* **2018**, *92*, doi:10.1128/JVI.00798-18.
357. Nedialkov, Y.A.; Gong, X.Q.; Hovde, S.L.; Yamaguchi, Y.; Handa, H.; Geiger, J.H.; Yan, H.; Burton, Z.F. NTP-Driven Translocation by Human RNA Polymerase II. *Journal of Biological Chemistry* **2003**, *278*, 18303–18312, doi:10.1074/jbc.M301103200.
358. Samuel, C.E. Antiviral Actions of Interferons. *Clin Microbiol Rev* **2001**, *14*, 778–809, doi:10.1128/CMR.14.4.778-809.2001.
359. Patterson, J.B.; Thomis, D.C.; Hans, S.L.; Samuel, C.E. Mechanism of Interferon Action: Double-Stranded RNA-Specific Adenosine Deaminase from Human Cells Is Inducible by Alpha and Gamma Interferons. *Virology* **1995**, *210*, 508–511, doi:10.1006/viro.1995.1370.
360. George, C.X.; Samuel, C.E. Human RNA-Specific Adenosine Deaminase *ADAR1* Transcripts Possess Alternative Exon 1 Structures That Initiate from Different Promoters, One Constitutively Active and the Other Interferon Inducible. *Proceedings of the National Academy of Sciences* **1999**, *96*, 4621–4626, doi:10.1073/pnas.96.8.4621.
361. Kawakubo, K.; Samuel, C.E. Human RNA-Specific Adenosine Deaminase (*ADAR1*) Gene Specifies Transcripts That Initiate from a Constitutively Active Alternative Promoter. *Gene* **2000**, *258*, 165–172, doi:10.1016/S0378-1119(00)00368-1.
362. Chen, R.; Linnstaedt, S.; Casey, J. RNA Editing and Its Control in Hepatitis Delta Virus Replication. *Viruses* **2010**, *2*, 131–146, doi:10.3390/v2010131.
363. Polson, A.G.; Ley, H.L.; Bass, B.L.; Casey, J.L. Hepatitis Delta Virus RNA Editing Is Highly Specific for the Amber/W Site and Is Suppressed by Hepatitis Delta Antigen. *Mol Cell Biol* **1998**, *18*, 1919–1926, doi:10.1128/MCB.18.4.1919.

364. Casey, J.L.; Tennant, B.C.; Gerin, J.L. Genetic Changes in Hepatitis Delta Virus from Acutely and Chronically Infected Woodchucks. *J Virol* **2006**, *80*, 6469–6477, doi:10.1128/JVI.00245-06.
365. Jayan, G.C.; Casey, J.L. Increased RNA Editing and Inhibition of Hepatitis Delta Virus Replication by High-Level Expression of ADAR1 and ADAR2. *J Virol* **2002**, *76*, 3819–3827, doi:10.1128/JVI.76.8.3819-3827.2002.
366. Cotrina, M.; Buti, M.; Jardí, R.; Quer, J.; Rodríguez, F.; Pascual, C.; Esteban, R.; Guardia, J. Hepatitis Delta Genotypes in Chronic Delta Infection in the Northeast of Spain (Catalonia). *J Hepatol* **1998**, *28*, 971–977, doi:10.1016/S0168-8278(98)80345-0.
367. Palom, A.; Sopena, S.; Riveiro-Barciela, M.; Carvalho-Gomes, A.; Madejón, A.; Rodríguez-Tajes, S.; Roade, L.; García-Eliz, M.; García-Samaniego, J.; Lens, S.; et al. One-quarter of Chronic Hepatitis D Patients Reach HDV-RNA Decline or Undetectability during the Natural Course of the Disease. *Aliment Pharmacol Ther* **2021**, *54*, 462–469, doi:10.1111/apt.16485.
368. Yll, M.; Cortese, M.F.; Guerrero-Murillo, M.; Orriols, G.; Gregori, J.; Casillas, R.; González, C.; Sopena, S.; Godoy, C.; Vila, M.; et al. Conservation and Variability of Hepatitis B Core at Different Chronic Hepatitis Stages. *World J Gastroenterol* **2020**, *26*, 2584–2598, doi:10.3748/wjg.v26.i20.2584.

Chapter 11

Manuscript

Article

Inspecting the Ribozyme Region of Hepatitis Delta Virus Genotype 1: Conservation and Variability

Beatriz Pacin-Ruiz ^{1,2}, María Francesca Cortese ^{1,2,*}, David Tabernero ^{1,2,*}, Sara Sopena ¹, Josep Gregori ³, Selene García-García ^{1,2}, Rosario Casillas ¹, Adrián Najarro ¹, Unai Aldama ¹, Adriana Palom ⁴, Ariadna Rando-Segura ⁵, Anna Galán ¹, Marta Vila ¹, Mar Riveiro-Barciela ^{2,4}, Josep Quer ³, Gloria González-Aseguinolaza ⁶, María Buti ^{2,4,†} and Francisco Rodríguez-Frías ^{1,2,7,†}

¹ Liver Pathology Unit, Departments of Biochemistry and Microbiology, Vall d'Hebron University Hospital, 08035 Barcelona, Spain; beatriz.pacin@vhir.org (B.P.-R.); sarasopena91@gmail.com (S.S.); selene.garcia@vhir.org (S.G.-G.); chr.casillas.gmz@gmail.com (R.C.); a.najarro@vhebron.net (A.N.); ualdama@vhebron.net (U.A.); annagalangonzalez@gmail.com (A.G.); marta.vila.salvador@vhir.org (M.V.); frarodri@vhebron.net (F.R.-F.)

² Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Instituto de Salud Carlos III, 28029 Madrid, Spain; mmriveir@vhebron.net (M.R.-B.); mbuti@vhebron.net (M.B.)

³ Liver Unit, Liver Disease, Laboratory-Viral Hepatitis, Vall d'Hebron Institut Recerca-Vall d'Hebron University Hospital, 08035 Barcelona, Spain; Josep.gregori@gmail.com (J.G.); josep.quer@vhir.org (J.Q.)

⁴ Liver Unit, Department of Internal Medicine, Vall d'Hebron University Hospital, 08035 Barcelona, Spain; adrianapalom@gmail.com

⁵ Department of Microbiology, Vall d'Hebron University Hospital, 08035 Barcelona, Spain; a.rando@vhebron.net

⁶ Center for Applied Medical Research (CIMA), University of Navarra, 31008 Pamplona, Spain; ggasegui@unav.es

⁷ Biochemistry and Molecular Biology Department, Universitat Autònoma de Barcelona (UAB), Plaça Cívica, 08193 Bellaterra, Spain

* Correspondence: maria.cortese@vhir.org (M.F.C.); david.tabernero@ciberehd.org (D.T.)

† These authors contributed equally to this work.



Citation: Pacin-Ruiz, B.; Cortese, M.F.; Tabernero, D.; Sopena, S.; Gregori, J.; García-García, S.; Casillas, R.; Najarro, A.; Aldama, U.; Palom, A.; et al. Inspecting the Ribozyme Region of Hepatitis Delta Virus Genotype 1: Conservation and Variability. *Viruses* **2022**, *14*, 215. <https://doi.org/10.3390/v14020215>

Academic Editor: Stefan Urban

Received: 23 December 2021

Accepted: 19 January 2022

Published: 22 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The hepatitis delta virus (HDV) genome has an autocatalytic region called the ribozyme, which is essential for viral replication. The aim of this study was to use next-generation sequencing (NGS) to analyze the ribozyme quasispecies (QS) in order to study its evolution and identify highly conserved regions potentially suitable for a gene-silencing strategy. HDV RNA was extracted from 2 longitudinal samples of chronic HDV patients and the ribozyme (nucleotide, nt 688–771) was analyzed using NGS. QS conservation, variability and genetic distance were analyzed. Mutations were identified by aligning sequences with their specific genotype consensus. The main relevant mutations were tested *in vitro*. The ribozyme was conserved overall, with a hyper-conserved region between nt 715–745. No difference in QS was observed over time. The most variable region was between nt 739–769. Thirteen mutations were observed, with three showing a higher frequency: T23C, T69C and C64 deletion. This last strongly reduced HDV replication by more than 1 log *in vitro*. HDV Ribozyme QS was generally highly conserved and was maintained during follow-up. The most conserved portion may be a valuable target for a gene-silencing strategy. The presence of the C64 deletion may strongly impair viral replication, as it is a potential mechanism of viral persistence.

Keywords: hepatitis delta virus; ribozyme; next-generation sequencing; quasispecies; conservation; variability; viral fitness; persistence; target; gene silencing

1. Introduction

More than 250 million people worldwide are living with the hepatitis B virus (HBV), and between 15 and 20 million of them are chronically co-infected with hepatitis delta virus (HDV). HDV coinfection is associated with a higher risk of cirrhosis, hepatocellular carcinoma (HCC), and liver decompensation, causing the most severe form of viral hepatitis [1].

Like HBV, HDV is differently distributed in the world, with regions with a high prevalence such as Central and West Africa, Central Asia, the Pacific Islands, the Middle East, Eastern Europe, and South America (Amazon basin) [2,3].

HDV is composed of an RNA molecule with high intermolecular self-complementarity, giving rise to a rod-like structure. This 1.2 kb genome presents only one reading frame and codes for a protein existing in two isoforms of different length: the short (S-HDAg) isoform consisting of 195 amino acids (aa) and the long (L-HDAg) delta antigen with 214 aa (27 kDa) [4]. HDV genomic RNA replicates through a rolling circle process mediated by cellular RNA polymerase [5], by producing a concatemer of antigenomic monomers. The individual antigenomic molecules are obtained through a self-cleavage process led by the viral ribozyme [6,7]. Some of these antigenomic monomers are later circularized by a still unclear mechanism [8,9] to be used as templates for the genomic RNA synthesis through another process of rolling circle amplification. The remaining antigenomic molecules enter the transduction process by producing HDAg.

Therefore, an essential step in the viral life cycle is the co-transcriptional self-cleavage activity of the viral ribozyme, which generates both genome and coding viral RNAs. The ribozyme sequence is mainly observed in plant viroids and can catalyze biochemical reactions involved in RNA splicing, gene regulation and other processes [10]. The hepatitis delta virus is the only mammalian virus possessing this auto-catalytic 85-nucleotide-long sequence [10,11]. Notably, its secondary structure is a determining feature for its catalytic activity [12]. It is characterized by four double-stranded domains (P1, P1.1, P2, P3, and P4), three single-stranded regions (J1/2, J1.1/4, and J4/2) and 2 loop regions (L3 and L4) [13]. The ribozyme cleavage site resides in the P1 domain (positions 689/688 and 901/900 in genomic and antigenomic RNA, respectively). The P2 and P4 regions are fundamental for structure stabilization; P3 and J4/2 are involved in catalytic activity, and L3, part of the catalytic site, is essential for the correct RNA split [14–16].

Similarly to HBV, HDV shows a high variability and circulates as a population of closely related genetic variants called quasispecies (QS) [17,18]. A mutation rate of $1.4\text{--}3.2 \times 10^{-5}$ base substitutions/site/year has been reported for HBV [19], mainly caused by the viral reverse transcriptase which lacks proof-reading capacity. Consequently, a higher QS diversity has been reported in HBV-DNA related to HBV-RNA [20].

The origin of HDV genome variability, on the other hand, is still unclear [21]. Although the cellular RNA polymerase has proofreading activity with a low transcription error rate, HDV shows a high rate of evolution (1.2×10^{-3} to 9.5×10^{-3} nt substitutions/site/year) [22,23]. As proof of this variability, eight different genotypes have been described worldwide [24], with a divergence of up to 16% within the same genotype and between 20% and 40% between different genotypes [25].

The next-generation sequencing technique (NGS) is a very sensitive technique that makes it possible to analyze the less frequent polymorphisms within a variant's population, providing valuable information on the viral QS and its evolution [26–28].

To date, the therapeutic strategy mainly used against HDV infection is based on interferon α , which has a limited administration time due to its adverse effects, without providing long-term suppression of viral replication [29]. Moreover, it has been reported that treatment with nucleotide analogues (NA) targeting HBV reverse transcriptase did not provide an improvement of HBV/HDV patients' clinical outcome [30]. Notably, new and specific therapeutic options are currently under study or have been recently approved in Europe, such as the Bulevirtide (BLV) (Hepcludex®), an acetylated fragment that inhibits viral entry [31].

Gene silencing is another promising antiviral strategy [32]. Silencing HDV and HBV expression may inhibit viral expression and limit liver disease progression [21]. However, due to the high viral genome variability, identifying highly conserved regions is essential to designing a strategy that may be effective in the presence of different viral genotypes and quasispecies. Notably, the HDAg region is characterized by high variability, which does not make it an optimal target [33]. On the other hand, due to its essential

function in HDV replication, the ribozyme may be a valuable candidate for designing gene-silencing molecules.

We used NGS to analyze the ribozyme region in two longitudinal samples in order to study QS evolution in this region and identify highly conserved regions that may be valuable targets of a gene-silencing strategy.

2. Materials and Methods

2.1. Patients and Samples

Patients were selected from those attending the outpatient clinics at Vall d'Hebron University Hospital, Barcelona, Spain. Enrolled patients presented at least 3log IU/ mL of HDV RNA and presented chronic hepatitis (CHD). Patients with other co-infections or autoimmune disease were excluded. Two plasma samples were collected per patient: one sample at the start of the study and the other at the end, with a mean follow-up of 2.25 years.

2.2. Serologic and Molecular Assays

Quantification of HDV-RNA was performed by means of an in-house one-step quantitative RT-PCR technique using the WHO international standard (1st World Health Organization International Standard for Hepatitis D Virus RNA for Nucleic Acid Amplification Techniques-based assays) [34] with linearity ranging from 5.75×10^2 to 5.75×10^5 IU/ mL and a detection limit of 5.75×10^1 IU/ mL. HBV DNA was quantified using real-time PCR with a detection limit of 10 IU/ mL (COBAS 6800, Roche Diagnostics, Rotkreuz, Switzerland). HBV serological markers such as surface antigen (HBsAg) were tested using commercial chemiluminescent assays on a COBAS 8000 analyzer (Roche Diagnostics, Rotkreuz, Switzerland).

The virus was genotyped by analyzing the HDAg region (between nt 910–1270) using next-generation sequencing (Table S1). Genotypes were determined using the Kimura-80 model and a dendrogram was constructed using the unweighted pair group method with arithmetic mean (UPGMA).

2.3. Next-Generation Sequencing of the Ribozyme Region

HDV RNA was extracted depending on the sample volume using the automated MagNA Pure LC system (Roche Applied Science, Indianapolis, IN, USA) or by manual extraction following the QIAamp Viral RNA mini kit protocol (QIAGEN®, Hilden, Germany). Extracted HDV RNA was denatured at 98 °C for 5 min and immediately transferred to 80 °C and successively retro-transcribed using Accuscript HiFi enzyme (Agilent Technologies, Santa Clara, CA, USA). To amplify the ribozyme (positions 688–771 on the antigenome), three nested PCRs were performed, as shown in Table 1. The first PCR step amplified a larger region (between nt 1454–308) that included the ribozyme portion. In the second amplification, the ribozyme was amplified with the addition of M13 sequences that were used to include the multiplex identifier (MID) specific to each sample in the third and last step. PCR products were charged on 1.5% agarose gel diluted in 1× tris-acetate-EDTA (TAE) buffer (Corning Mediatech Inc., Manassas, VA, USA), analyzed by electrophoresis and purified using the QIAquick Gel Extraction Kit protocol (QIAGEN®, Hilden, Germany). Purified samples were fluorometrically quantified with the automated system Freedom EVO® (Tecan, Mannedorf, Switzerland) coupled to the Infinite 200 Pro (Tecan, Mannedorf, Switzerland) fluorimeter using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, Camarillo, CA, USA).

Table 1. Protocols for ribozyme amplification. The table shows the different steps for ribozyme amplification, including the retro-transcription and the three nested-PCR (1st PCR; M13 PCR, MID PCR). The primer sequence and amplification region for each step is shown. The M13 sequence tail is underlined. Abbreviations: Fw indicates forward; Rv, reverse; MID, multiplex identifier.

Amplification Step	Primer	Amplified Region	Primer Sequence (5' to 3')	Protocol
RT	RT rv	1435–1454	TGGCTGGGAAACATCAAAGG	RT 42 °C 60 min; inactivation 70 °C 10 min; cooling 20 °C 10 min
1st PCR	1a fw	1435–1454	TGGCTGGGAAACATCAAAGG	95 °C 1 min; (94 °C 20 s, 54 °C 20 s, 72 °C 45 s) × 40 cycles; 72 °C 3 min
	1a rv	308–326	CCTCCAGAGGACCCCTTCA	
M13 PCR	M13-fw	883–900	<u>CACAGGAAACAGCTATGACCTCGGCATGGCATCTCCAC</u>	95 °C 2 min; (94 °C 20 s, 60 °C 20 s, 72 °C 30 s) × 35 cycles; 72 °C 3 min
	M13-rv	663–683	<u>GTGTGTTAAACGACGGCCAGTCGCGTTCCATCCTTTCTTACC</u>	
MID PCR	MID fw	-	MID-GTTGTAAAACGACGGCCAGT	95 °C 2 min; (94 °C 20 s, 60 °C 20 s, 72 °C 45 s) × 25 cycles; 72 °C 3 min
	MID rv	-	MID-CACAGGAAACAGCTATGACC	

Samples were then normalized to the concentration of 1.00×10^{10} molecules/ μ L. Samples were pooled and then processed using NGS according to the Miseq illumina platform protocol (Illumina, San Diego, CA, USA).

2.4. Quasispecies Analysis and Statistics

The sequences (reads) obtained were bioinformatically filtered (R software [35]). Only those reads with a complete sequence, a good overlapping between forward and reverse strands (less than 10% of mismatches), and optimal quality were maintained and demultiplexed to obtain the unique QS sequences (haplotypes).

QS conservation was analyzed by calculating the information content (IC) of each position in a multiple alignment of all haplotypes obtained by means of NGS, followed by a sliding window analysis, as previously described by our group [36].

Inter-patient and intra-patient sequence distance was studied to evaluate QS variability. Moreover, different complexity indices were also considered, as previously reported by our group [37]: number of reads per sample, number of master reads (Mstr), master percentage (Mpct), number of haplotypes, polymorphic sites, number of mutations, Shannon index, Gini–Simpson coefficient, functional attribute diversity (FAD), mutation frequency (Mf), nucleotide diversity (Pi) and Pi to Mf ratio. The mutations in ribozyme sequences were identified by aligning the QS sequence for each sample and each patient with its specific genotype consensus. Mutation frequency was obtained by summing the relative frequencies of each haplotype carrying the specific mutation.

2.5. In Vitro Test of Mutations

To evaluate the effect of the observed mutations in HDV replication capacity, the most relevant changes were introduced using site-directed mutagenesis (QuickChange Lightning site-directed mutagenesis kit -Agilent Technologies, Waldbronn, Germany), in a plasmid (pCMV-HDV-1.2X) containing 1.2 copies of genomic cDNA [38], following the manufacturer's protocol. The correct introduction of the mutation was ensured by analyzing the plasmid using Sanger sequencing. Mutated plasmids were extracted using the Endotoxin free NucleoBond Xtra Midi Plus kit (Machery-Nagel, Düren, Germany). The concentration (ng/ μ L) of the extracted plasmids was determined by fluorometric quantification using Qubit fluorometers (Thermo Fisher Scientific-Life Technologies, Waltham, MA, USA).

Hepatocarcinoma Huh7 cells were cultured with Dulbecco Eagle's minimal essential medium (DMEM) supplemented by 10% of fetal bovine serum (FBS) and penicillin (100 U/ mL), streptomycin (100 μ g/ mL), and Glutamax (2 mM). Cells were plated at 160,000 cells/ mL and transfected with wild-type (wt) and mutated plasmids using the Magnetofectamine O2 kit (OZbiosciences, Marseille, France) according to the manufacturer's protocol. To guarantee production of HDV viral particles, a plasmid containing 1.3-length HBV genome (pTriEx-HBV) [39] was included in each condition. The supernatant was collected 72 h after transfection.

To quantify HDV RNA release in vitro, viral RNA was extracted from the cell supernatant using the automated MagNA Pure LC system (Roche Applied Science, Indianapolis, IN, USA). To ensure plasmid removal, a DNase step of the extracted RNA was performed (DNase I, Amplification grade, Thermo Fisher Scientific, Camarillo, CA, USA). HDV RNA was then quantified as reported above for plasma samples. To ensure that transfection did not affect cell viability, control negative cells were transfected with the empty pCMV backbone.

3. Results

3.1. Patients and Sequencing

Twenty-five patients were included in the study with a total of 50 samples. Most patients were infected by genotype 1, but 1 patient presented genotype 8 HDV (P04). However, after applying the quality filters, only 19 patients (38 samples) were later considered. Notably, all of them were infected by genotype 1 virus. HDV viremia did not change between the two timepoints (median [IQR] log₁₀ HDV RNA of 5.76 (5.02–5.83) and 5.76 (3.68–5.76), respectively (Table 2).

Table 2. Main clinical and viral characteristics of chronic hepatitis delta (CHD) patients included in the study. The table shows the clinical and viral characteristics of the patients between the two timepoints (Sample A and Sample B). P values were obtained by applying the Kruskal–Wallis test. Abbreviations (normal quantification values): HDV, Hepatitis delta virus; AST, Aspartate aminotransferase (normal value 12–50 IU/mL); ALT, alanine aminotransferase (normal value 8–50 IU/mL); platelets (140–400 UI/mL); IQR, interquartile range; HBsAg, hepatitis B virus surface antigen; IND, undetectable.

Markers	Sample A	Sample B	p
HDV RNA Median (IQR) Log ₁₀ (IU/mL)	5.76 (5.02–5.83)	5.76 (3.68–5.76)	0.383
AST (UI/L) Median (IQR)	88 (45–133)	90 (36.25–126)	0.551
ALT (UI/L) Median (IQR)	100 (58–158.5)	89.25 (44.5–133.75)	0.439
PLATELETS (UI/L) Median (IQR)	89 (123–212)	79 (118–197)	0.966
HBV-DNA (IU/mL)	Low/ IND	Low/ IND	
HBsAg Median (IQR) Log ₁₀ (IU/mL)	3.96 (3.57–4.11)	3.92 (3.39–4.03)	0.827

Among the 38 samples that passed this filter, we obtained a median (IQR) of reads of 4550.5 (1342.24–5892.74) per patient.

3.2. Ribozyme Conservation

QS conservation was analyzed by aligning all the haplotypes, applying a sliding window analysis, and calculating the information content considering their relative frequency or not. The ribozyme (between positions 688 and 771) was overall highly conserved and 85% of the nucleotide positions presented 2 bits of information content (100% conservation) (Figure 1A). No difference was observed between considering or not considering haplotype frequency. Of the 85 nucleotides included in the ribozyme sequence, just 3 nt positions (around 3.5% of the total) presented a conservation of less than 1.5 bits (Figure 1B).

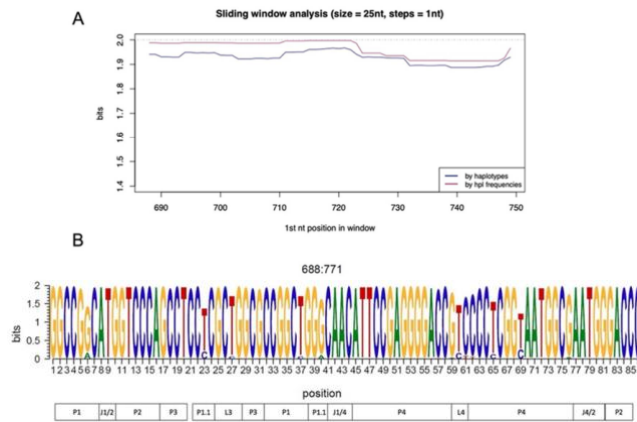


Figure 1. Conservation of ribozyme region by sliding window analysis and represented by logo. (A) The sliding window analysis is the result of the mean information content (bits) of the 25-nt windows with a displacement between them of 1-nt obtained by multiple alignments of all the quasispecies (QS) haplotypes. The analysis was implemented by considering (red line) or not (blue line) haplotypes frequency. The dashed line represents the maximum level of conservation (2 bits). (B) Logo representation of the nucleotide sequence corresponding to the entire ribozyme region from the genome positions 688 (corresponding to position 1 in the ribozyme) to 771 (corresponding to position 86 in the logo representation). The height of each letter represents the grade of conservation from a maximum of 2 bits to a minimum of 0. The sequence is shown in the genome sense. The different structural and functional domains of the ribozyme are reported at the bottom [40].

The most conserved region was between nt 715–745 (Figure 2A), where 100% of the nucleotide presented a conservation of more than 1.5 bits. The most variable region encompassed positions 739 and 769, as described previously [41]. In this region, 3 nt positions presented a conservation of less than 1.5 bits (3/ 31 nts), although 80% (25/ 31) of the nts showed a high level of conservation (Figure 2B).

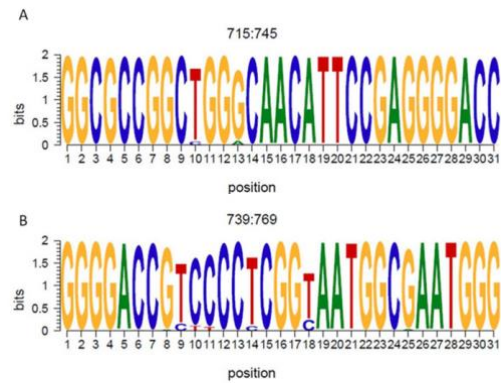


Figure 2. Logo representation of the most conserved and most variable portion of the ribozyme. The logo represents the most conserved (A) (nt 715–745) and most variable (B) (nt 739–769) regions of the ribozyme. The height of each nucleotide represents its information content in bits (2 bits indicates 100% of conservation). The sequence is shown in the genome sense.

3.3. Ribozyme Quasispecies (QS) Evolution and Variability during Follow-Up

The genetic distance between samples was determined to evaluate ribozyme QS evolution during the patient's follow-up (Figure 3).

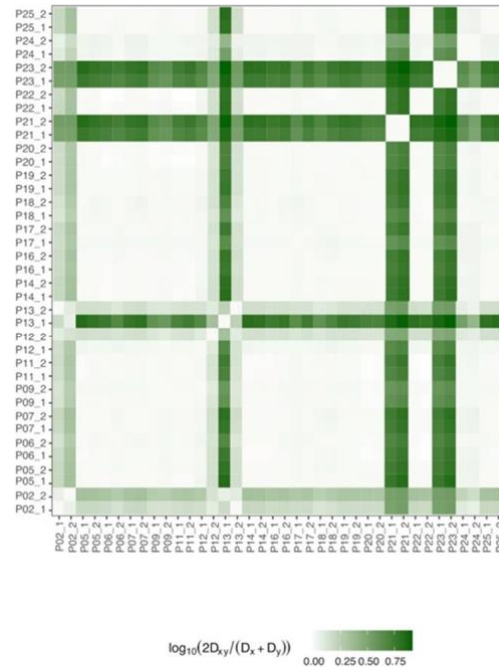


Figure 3. Sequences distance between samples. The heatmap shows the distance of the ribozyme sequences between the different samples included in the study as the logarithm of the ratio between QS nucleotide distance (D_{xy}) and the nucleotide diversity average of each quasispecies (D_x and D_y). The tone of the green indicates the degree of distance, from the lighter green, which corresponds to a shorter distance, to the darker green, which indicates greater sample-sample distances. Distance was calculated by considering the QS consensus of each sample.

As expected, considering the high level of conservation, no distance (<0.25) was observed between the 2 samples for each patient (Figure 3). Although patients 2, 12, 13, 21, and 23 presented a greater distance (>0.50) compared to the others, when their sequences were examined in greater depth, this greater distance was determined by just a few nucleotides.

To inspect even more ribozyme QS variability, different complexity indices were analyzed and compared between sample A and sample B. Notably, no difference was detected for any tested index (Table S2).

3.4. Analysis of Mutations

Mutations were identified by aligning haplotypes sequences with the corresponding genotype consensus. Although the ribozyme was conserved overall, when studying each sample individually, a total of 48 mutations were observed. Of these, 12 were in at least six patients (Table 3). When looking at mutation type, we observed that the most observed changes were transitions: C \rightarrow T (16 mutated positions), G \rightarrow A (10 mutated positions) and T \rightarrow C (9 mutated positions) (Figure S1). Moreover, four of the 12 mutations observed involved the P4 domain, as reported in previous studies [42].

Table 3. Most prevalent mutations observed in the ribozyme. Table shows the 12 most prevalent mutations observed in the ribozyme. The relative frequency of each mutation between sample A and sample B is reported as mean \pm standard deviation. Mutations are numbered starting from the first nt in the ribozyme (nt 688).

Mutation	Sample A (%freq)	Sample B (%freq)
G6A	0.84 \pm 0.43%	0.83 \pm 0.31%
T23C	3.04 \pm 2.27%	5.55 \pm 8.91%
T27C	0.12 \pm 0.3%	0.88 \pm 3.5%
G40A	0.25 \pm 0.22%	0.41 \pm 0.24%
G59A	0.07 \pm 0.03%	0.08 \pm 0.02%
T60C	6.15 \pm 30.64%	6.42 \pm 35.01%
C61T	0.15 \pm 0.24%	0.14 \pm 0.27%
C62T	0.12 \pm 0.17%	0.08 \pm 0.12%
C64d	47.17 \pm 1.66%	48.88 \pm 2.47%
T65C	6.02 \pm 39.66%	6.29 \pm 44.04%
T69C	12.02 \pm 21.99%	12.75 \pm 44.04%
G76A	0.53 \pm 2.49%	0.44 \pm 17.68%

Of the observed mutations, only three involved at least eight patients and were maintained or selected (mutation frequency that changes between the two-time point) in the follow-up samples: T710C (position 23 if considering the first nt in ribozyme starting from 688), T756C (T69C) and the deletion in position 751 (C64d). The T23C mutation involved 17 of the 19 patients in sample A with a mean frequency of 2.99%, and 18/19 of the patients with a mean frequency of 4.96% in sample B (Figure 4A). The T69C mutation is found in both samples in 100% of patients, with a relative frequency that was maintained between the two samples in the follow-up (10.6% and 10.8% in sample A and B, respectively) (Figure 4C). The deletion in position 64 was observed in eight of the 19 patients in both samples with a frequency of 98.7% in sample A and 98.3% in sample B (Figure 4B).

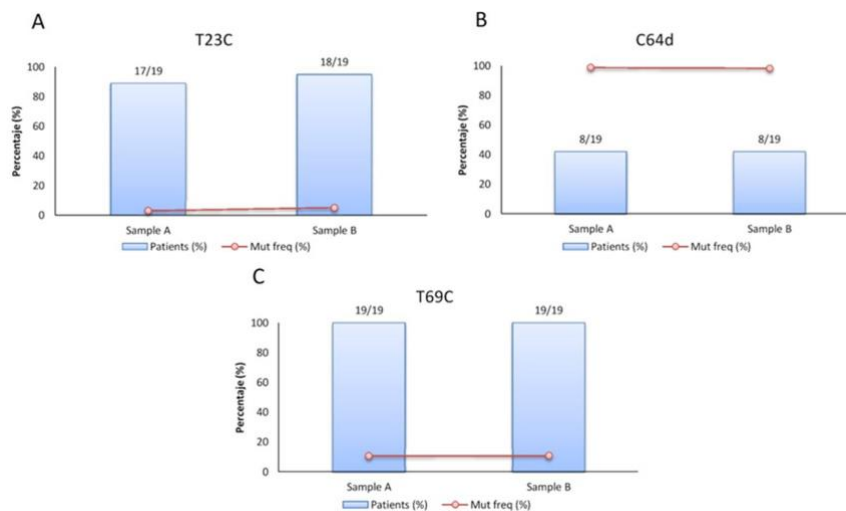


Figure 4. Incidence of the principal mutations observed in the ribozyme. Graphs show the relative frequency (red line), the percentage (%) and number of patients (blue bar) presenting the mutation T23C (A), C64d (B) and T69C (C) and how they change between the 2 time points (between sample A and sample B).

Although the relative frequency of mutations is generally maintained, in some patients some mutations were positively or negatively selected, such as the T23C mutation in patient 2 (from 5.29% to 17.07%), and in patient 12 (from 9.52% to 36.97) (Figure S2A) and the T69C mutation in patients 2 (from 43.88% to 71.85%) and patient 12 (from 2.79% to 24.11%) retracted in patient 13 (from 91.15% to 39.09%) (Figure S2C).

3.5. In Vitro Test of Mutations

Considering the high degree of ribozyme conservation, the observed mutations may potentially affect HDV fitness, thus promoting or inhibiting viral expression. To test the effect of the three detected principal mutations (T23C, C64d, T69C) of HDV expression, the plasmids carrying the desired mutations will be transfected in the presence of HBV and the HDV RNA titer was quantified in transfected cell supernatants 72 h after transfection. We efficiently expressed HDV in vitro, obtaining a mean \pm SD titer of 2.86 ± 0.61 IU/mL 72h after transfection. The T69C mutant showed a similar replication rate than wt (2.78 ± 1.04 IU/mL, $p = 0.662$). Of note, in the presence of the mutation T23C (2.7 ± 0.38 IU/mL, $p = 1$), the HDV titer was reduced around 1.05-fold. The interference with viral expression was even more strong in presence of the deletion in position 64 (C64d) that caused a reduction in viral expression of around 1.22log (HDV RNA = 1.63 ± 0.71 IU/mL, $p = 0.08$) (Figure 5).

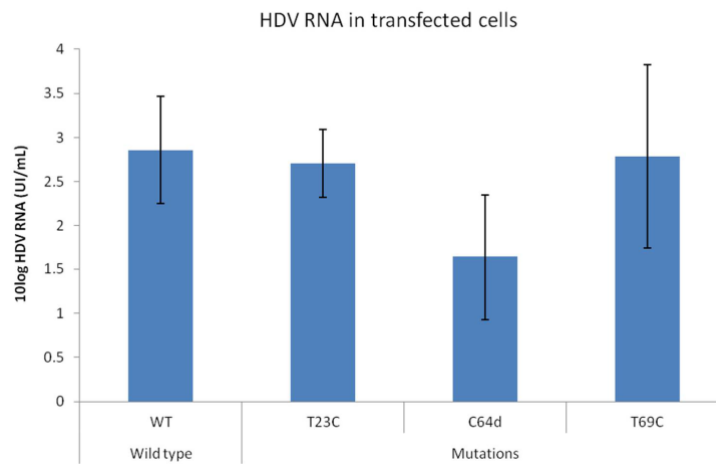


Figure 5. Effect of the most relevant observed mutations in HDV replication in vitro. The graph shows HDV titer in cell supernatant after transfection with HBV and HDV wt or mutated p values were obtained by applying the Mann-Whitney test.

4. Discussion

Due to the lack of a viral polymerase, no specific antiviral therapy is available against HDV infection. New treatment options against HDV are currently under study, such as HBsAg release, prenylation and viral entry inhibitors [43]. Among them, the entry inhibitor BLV has been recently approved for use, alone or in association with peg-IFN α , in CHD patients with compensated liver disease [44,45].

Notably, a gene therapy approach may be a valuable strategy to promote HDV RNA elimination and block disease progression. Due to their intranuclear location, the genome and antigenome seem to be resistant to interfering RNA (siRNA) activity [46], whose silencing activity is developed in the cell cytoplasm. HDV genomes, however, may be targeted by antisense oligonucleotides (ASOs), which are active within the nucleus [47]. Although inhibition of the expression of HBsAg through a siRNA may interfere with HDV

infection [48], a combination of silencing molecules targeting both viruses may represent a highly valuable therapeutic strategy. To date, however, no silencing HDV-specific molecules have been reported. The extreme variability of HDV RNA may make it very difficult to design an effective antisense oligonucleotide. To this end, identifying highly conserved regions in the HDV genome to use as target may be essential.

Ribozyme activity is crucial to producing the excision of unitary RNA monomers during viral replication. Given its key role, this region may be a valuable target for antisense oligonucleotides. Although the HDV genome has been extensively studied using Sanger sequencing [42,49], this is the first report to focus on studying ribozyme QS conservation and variability using next-generation sequencing. As expected, the ribozyme was overall highly conserved, with a hyper-conserved region encompassing positions 715–745. This region involves the P1 domain, which is essential for ribozyme activity since it is the region where the self-cleavage site resides [50].

To evaluate ribozyme QS evolution during follow-up, two samples were included per patient. Notably, when considering sequence distance or complexity indices, no differences were observed between the two longitudinal samples, suggesting that the QS in this genome portion did not vary over time.

However, a variable region between nt 739–769 was also identified, although changes were observed in just a few nucleotide positions (specifically in six positions). This variable region falls into the P4 structural domain as observed in a previous study that reported a higher mutations level for both genomic and antigenomic RNA [42]. Although this domain is not directly involved in the self-cleavage process, it has been reported that the conservation of the nucleotide sequence is essential for a stable base pairing, which is a necessary condition for efficient self-cleavage activity [51].

A recently reported study shows that a wt ribozyme sequence guarantees the optimal ribozyme activity in co-transcriptional conditions [52]. Different mutations were experimentally introduced in ribozyme sequence to study its functional domains, demonstrating that some positions cover important roles in cleavage activity such as the U23 and C24 in the L3 domain [53] or the T20, C21 and C75 which are involved in coordinating Mg²⁺ ion and cleavage catalysis [54]. By analyzing the HDV full genome in CHD patients, different mutations were found in the ribozyme sequence [42]. Notably, many of these identified mutations were also observed in this study. We identified a total of 48 mutations, 12 of which were found in at least six patients. The identification of mutations in such a conserved and active region of the HDV genome may indicate that they do not affect viral fitness. As previously reported [42], the observed mutations mainly involved the P4 domain (positions 60, 61, 62, 64, 65 and 69). Of these 12 mutations, three presented a relative frequency of 1% or more and were tested *in vitro*: T23C (in L3 domain), C64d, and T69C (both in P4 domain).

In vitro testing of the ribozyme self-cleavage activity in the presence of the T23C mutations showed a decrease of 10³-fold of the cleavage rate [53], however when we tested the viral replication *in vitro* in the presence of this mutation, we observed that the change in position 23 had a limited impact on viral replication. This mutation falls into the L3 domain, which is an important element in the cleavage activity [16,48]. Of note, as previously reported, this change is less effective in interfering with ribozyme activity than others (in positions 20, 21, and 25) that may entail the loss of ribozyme activity [42].

Changes in positions 64 and 69, however, involved the P4 domain. Higher mutation levels in this domain have been previously reported [42]. As previously mentioned, this domain shows a stabilizing function of the ribozyme structure, and it is not directly involved in RNA catalysis. Indeed, it has been observed that its elimination does not compromise virus replication [15,16]. Although the presence of the T69C change did not impact viral replication, the deletion in position 64 strongly reduced HDV replication by more than 1log IU/mL. Considering the non-essentiality of the P4 domain, the presence of this 1nt-long deletion in this portion could probably alter the ribozyme secondary structure, thus affecting viral replication. Moreover, this change was observed at a relative frequency

of around 50%. The fact that it had been observed and maintained at such a high frequency in viral QS suggests that this mutation may promote viral persistence, as has been reported for other viruses such as hepatitis C [55] and human immunodeficiency viruses [56].

This is the first study to focus on studying ribozyme QS using NGS, but it does have some limitations. Although 25 patients were selected, only 19 of them were correctly amplified and passed the quality filters. In addition, all patients were infected with HDV genotype 1, which is the main genotype in infected patients in Spain [57]. More patients, infected with more viral genotypes should be included to confirm these results. Moreover, in silico modeling should be implemented to highlight the effects of the C64 deletion on the secondary structure and, consequently, on the function of the ribozyme.

In conclusion, as expected considering its essential role in the viral life cycle, the ribozyme is overall highly conserved in viral QS and did not change over time. The most conserved portion involved a domain that plays a direct role in the auto-catalysis process and may be a valuable target for designing a new gene-silencing strategy against HDV. The most variable region, however, involves a domain that is not essential from a functional point of view, but a deletion in this region may strongly impair viral replication. Considering its relatively high frequency, it may be a potential mechanism of viral persistence.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v14020215/s1>, Table S1: Protocols for delta antigen (HDAG) encoding HDV genome region amplification, Table S2: complexity indexes calculated in samples A and B in the ribozyme region, Figure S1: Number and type of mutations found in the ribozyme region, Figure S2: Evolution of mutation frequency between the two follow-up samples (A and B).

Author Contributions: Conceptualization, F.R.-F. and M.B.; methodology R.C., A.N., U.A., A.G. and M.V. software, J.G.; validation, M.F.C., D.T., G.G.-A., M.B. and F.R.-F.; formal analysis, M.F.C. and D.T.; investigation, B.P.-R., S.G.-G. and S.S.; resources, A.R.-S., A.P., M.R.-B. and J.Q.; writing—original draft preparation, B.P.-R.; writing—review and editing, M.F.C., D.T., J.Q., M.B. and F.R.-F.; supervision, M.F.C.; project administration, A.P. and M.R.-B.; funding acquisition, M.B. and F.R.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Institute of Health Carlos III, grant number PI20/01692 and co-financed by the European Regional Development Fund (ERDF), and by the European Regional Development Fund (ERDF)-Ministry of Economy, Industry and Competitiveness, grantRTI2018-101936-B-I00.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Vall d'Hebron Research Institute (protocol code PR(AG)247/2018 and date of approval 20 July 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: NGS data were submitted to the GenBank SRA database (BioProject accession number PRJNA798886). See Biosample numbers list in Supplementary Material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sureau, C.; Negro, F. The Hepatitis Delta Virus: Replication and Pathogenesis. *J. Hepatol.* **2016**, *64*, S102–S116. [CrossRef]
2. Chen, H.-Y.; Shen, D.-T.; Ji, D.-Z.; Han, P.-C.; Zhang, W.-M.; Ma, J.-F.; Chen, W.-S.; Goyal, H.; Pan, S.; Xu, H.-G. Prevalence and Burden of Hepatitis D Virus Infection in the Global Population: A Systematic Review and Meta-Analysis. *Gut* **2019**, *68*, 512–521. [CrossRef] [PubMed]
3. Wedemeyer, H.; Negro, F. Devil Hepatitis D: An Orphan Disease or Largely Underdiagnosed? *Gut* **2019**, *68*, 381–382. [CrossRef] [PubMed]
4. Wang, T.-C.; Chao, M. Molecular Cloning and Expression of the Hepatitis Delta Virus Genotype IIb Genome. *Biochem. Biophys. Res. Commun.* **2003**, *303*, 357–363. [CrossRef]
5. Farci, P.; Anna Niro, G. Current and Future Management of Chronic Hepatitis D. *Gastroenterol. Hepatol.* **2018**, *14*, 342–351.
6. Sharmeen, L.; Kuo, M.Y.; Dinter-Gottlieb, G.; Taylor, J. Antigenomic RNA of Human Hepatitis Delta Virus Can Undergo Self-Cleavage. *J. Virol.* **1988**, *62*, 2674–2679. [CrossRef]
7. Reid, C.E.; Lazinski, D.W. A Host-Specific Function Is Required for Ligation of a Wide Variety of Ribozyme-Processed RNAs. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 424–429. [CrossRef]

8. Wu, H.N.; Lin, Y.J.; Lin, F.P.; Makino, S.; Chang, M.F.; Lai, M.M. Human Hepatitis Delta Virus RNA Subfragments Contain an Autocleavage Activity. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 1831–1835. [CrossRef] [PubMed]
9. Sharmeen, L.; Kuo, M.Y.; Taylor, J. Self-Ligating RNA Sequences on the Antigenome of Human Hepatitis Delta Virus. *J. Virol.* **1989**, *63*, 1428–1430. [CrossRef] [PubMed]
10. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**, *31*, 147–157. [CrossRef]
11. Golden, B.L. Two Distinct Catalytic Strategies in the Hepatitis Delta Virus Ribozyme Cleavage Reaction. *Biochemistry* **2011**, *50*, 9424–9433. [CrossRef] [PubMed]
12. Tang, J.; Breaker, R.R. Structural Diversity of Self-Cleaving Ribozymes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5784–5789. [CrossRef]
13. Ferré-D'Amaré, A.R.; Zhou, K.; Doudna, J.A. Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature* **1998**, *395*, 567–574. [CrossRef] [PubMed]
14. Riccitelli, N.; Lupták, A. HDV Family of Self-Cleaving Ribozymes. *Prog. Mol. Biol. Transl. Sci.* **2013**, *120*, 123–171. [CrossRef] [PubMed]
15. Thill, G.; Vasseur, M.; Tanner, N.K. Structural and Sequence Elements Required for the Self-Cleaving Activity of the Hepatitis Delta Virus Ribozyme. *Biochemistry* **1993**, *32*, 4254–4262. [CrossRef] [PubMed]
16. Been, M.D.; Wickham, G.S. Self-Cleaving Ribozymes of Hepatitis Delta Virus RNA. *Eur. J. Biochem.* **1997**, *247*, 741–753. [CrossRef] [PubMed]
17. Quer, J.; Rodríguez-Frias, F.; Gregori, J.; Tabernero, D.; Soria, M.E.; García-Cehic, D.; Homs, M.; Bosch, A.; Pintó, R.M.; Esteban, J.I.; et al. Deep Sequencing in the Management of Hepatitis Virus Infections. *Virus Res.* **2017**, *239*, 115–125. [CrossRef] [PubMed]
18. Domingo, E.; Perales, C. Viral Quasispecies. *PLoS Genet.* **2019**, *15*, e1008271. [CrossRef]
19. Orito, E.; Mizokami, M.; Ina, Y.; Moriyama, E.N.; Kameshima, N.; Yamamoto, M.; Gojobori, T. Host-Independent Evolution and a Genetic Classification of the Hepadnavirus Family Based on Nucleotide Sequences. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7059–7062. [CrossRef]
20. Revill, P.A.; Tu, T.; Netter, H.J.; Yuen, L.K.W.; Locarnini, S.A.; Littlejohn, M. The Evolution and Clinical Impact of Hepatitis B Virus Genome Diversity. *Nat. Rev. Gastroenterol. Hepatol.* **2020**, *17*, 618–634. [CrossRef]
21. Tabernero, D.; Cortese, M.F.; Buti, M.; Rodríguez-Frias, F. HDV Evolution—Will Viral Resistance Be an Issue in HDV Infection? *Curr. Opin. Virol.* **2018**, *32*, 100–107. [CrossRef] [PubMed]
22. Homs, M.; Rodríguez-Frias, F.; Gregori, J.; Ruiz, A.; Reimundo, P.; Casillas, R.; Tabernero, D.; Godoy, C.; Barakat, S.; Quer, J.; et al. Evidence of an Exponential Decay Pattern of the Hepatitis Delta Virus Evolution Rate and Fluctuations in Quasispecies Complexity in Long-Term Studies of Chronic Delta Infection. *PLoS ONE* **2016**, *11*, e0158557. [CrossRef] [PubMed]
23. Lee, C.-M.; Bih, F.-Y.; Chao, Y.-C.; Govindarajan, S.; Lai, M.M.C. Evolution of Hepatitis Delta Virus Rna during Chronic Infection. *Virology* **1992**, *188*, 265–273. [CrossRef]
24. le Gal, F.; Gault, E.; Ripault, M.-P.; Serpaggi, J.; Trinchet, J.-C.; Gordien, E.; Deny, P. Eighth Major Clade for Hepatitis Delta Virus. *Emerg. Infect. Dis.* **2006**, *12*, 1447–1450. [CrossRef] [PubMed]
25. Dény, P. Hepatitis Delta Virus Genetic Variability: From Genotypes I, II, III to Eight Major Clades? In *Hepatitis Delta Virus*; Springer: Berlin/ Heidelberg, Germany, 2006; pp. 151–171.
26. Homs, M.; Caballero, A.; Gregori, J.; Tabernero, D.; Quer, J.; Nieto, L.; Esteban, R.; Buti, M.; Rodríguez-Frias, F. Clinical Application of Estimating Hepatitis B Virus Quasispecies Complexity by Massive Sequencing: Correlation between Natural Evolution and On-Treatment Evolution. *PLoS ONE* **2014**, *9*, e112306. [CrossRef] [PubMed]
27. Sopena, S.; Godoy, C.; Tabernero, D.; Homs, M.; Gregori, J.; Riveiro-Barciela, M.; Ruiz, A.; Esteban, R.; Buti, M.; Rodríguez-Frias, F. Quantitative Characterization of Hepatitis Delta Virus Genome Edition by Next-Generation Sequencing. *Virus Res.* **2018**, *243*, 52–59. [CrossRef] [PubMed]
28. García-García, S.; Cortese, M.F.; Rodríguez-Algarra, F.; Tabernero, D.; Rando-Segura, A.; Quer, J.; Buti, M.; Rodríguez-Frias, F. Next-Generation Sequencing for the Diagnosis of Hepatitis B: Current Status and Future Prospects. *Expert Rev. Mol. Diagn.* **2021**, *21*, 381–396. [CrossRef]
29. Lempp, F.A.; Ni, Y.; Urban, S. Hepatitis Delta Virus: Insights into a Peculiar Pathogen and Novel Treatment Options. *Nat. Rev. Gastroenterol. Hepatol.* **2016**, *13*, 580–589. [CrossRef] [PubMed]
30. Scheller, L.; Hilgard, G.; Anastasiou, O.; Dittmer, U.; Kahraman, A.; Wedemeyer, H.; Deterding, K. Poor Clinical and Virological Outcome of Nucleos(t)ide Analogue Monotherapy in HBV/ HDV Co-Infected Patients. *Medicine* **2021**, *100*, e26571. [CrossRef] [PubMed]
31. Kang, C.; Syed, Y.Y. Bulevirtide: First Approval. *Drugs* **2020**, *80*, 1601–1605. [CrossRef] [PubMed]
32. Abbas, Z. Management of Hepatitis Delta: Need for Novel Therapeutic Options. *World J. Gastroenterol.* **2015**, *21*, 9461. [CrossRef] [PubMed]
33. Wang, S.-Y.; Wu, J.-C.; Chiang, T.-Y.; Huang, Y.-H.; Su, C.-W.; Sheen, I.-J. Positive Selection of Hepatitis Delta Antigen in Chronic Hepatitis D Patients. *J. Virol.* **2007**, *81*, 4438–4444. [CrossRef] [PubMed]
34. Paul-Ehrlich-Institut A WHO Collaborating Centre Bundesinstitut Für Impfstoffe Und Biomedizinische Arzneimittel for Quality Assurance of Blood Products and Federal Institute for Vaccines and Biomedicines in Vitro Diagnostic Devices. Available online: https://www.pei.de/SharedDocs/Downloads/EN/regulation-en/referencematerial/7657-12-ifu.pdf?__blob=publicationFile&v=2 (accessed on 22 January 2016).

35. R Core Team (2020)—European Environment Agency. Available online: <https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006> (accessed on 9 January 2022).
36. Cortese, M.F.; González, C.; Gregori, J.; Casillas, R.; Carioti, L.; Guerrero-Murillo, M.; Riveiro-Barciela, M.; Godoy, C.; Sopena, S.; Yll, M.; et al. Sophisticated Viral Quasispecies with a Genotype-Related Pattern of Mutations in the Hepatitis B X Gene of HBeAg-ve Chronically Infected Patients. *Sci. Rep.* **2021**, *11*, 4215. [CrossRef]
37. Godoy, C.; Taberner, D.; Sopena, S.; Gregori, J.; Francesca Cortese, M.; González, C.; Casillas, R.; Yll, M.; Rando, A.; López-Martínez, R.; et al. Characterization of Hepatitis B Virus X Gene Quasispecies Complexity in Mono-Infection and Hepatitis Delta Virus Superinfection. *World J. Gastroenterol.* **2019**, *25*, 1566–1579. [CrossRef] [PubMed]
38. Wu, T.T.; Netter, H.J.; Lazinski, D.W.; Taylor, J.M. Effects of Nucleotide Changes on the Ability of Hepatitis Delta Virus to Transcribe, Process, and Accumulate Unit-Length, Circular RNA. *J. Virol.* **1997**, *71*, 5408–5414. [CrossRef]
39. Homs, M.; Buti, M.; Quer, J.; Jardi, R.; Schaper, M.; Taberner, D.; Ortega, I.; Sanchez, A.; Esteban, R.; Rodríguez-Frias, F. Ultra-Deep Pyrosequencing Analysis of the Hepatitis B Virus PreCore Region and Main Catalytic Motif of the Viral Polymerase in the Same Viral Genome. *Nucleic Acids Res.* **2011**, *39*, 8457–8471. [CrossRef] [PubMed]
40. Yll, M.; Cortese, M.F.; Guerrero-Murillo, M.; Orriols, G.; Gregori, J.; Casillas, R.; González, C.; Sopena, S.; Godoy, C.; Vila, M.; et al. Conservation and Variability of Hepatitis B Core at Different Chronic Hepatitis Stages. *World J. Gastroenterol.* **2020**, *26*, 2584–2598. [CrossRef]
41. Chang, M.F.; Chen, C.H.; Lin, S.L.; Chen, C.J.; Chang, S.C. Functional Domains of Delta Antigens and Viral RNA Required for RNA Packaging of Hepatitis Delta Virus. *J. Virol.* **1995**, *69*, 2508–2514. [CrossRef] [PubMed]
42. Shirvani-Dastgerdi, E.; Amini-Bavil-Olyae, S.; Alavian, S.M.; Trautwein, C.; Tacke, F. Comprehensive Analysis of Mutations in the Hepatitis Delta Virus Genome Based on Full-Length Sequencing in a Nationwide Cohort Study and Evolutionary Pattern during Disease Progression. *Clin. Microbiol. Infect.* **2015**, *21*, 510.e11–510.e23. [CrossRef]
43. Mentha, N.; Clément, S.; Negro, F.; Alfaia, D. A Review on Hepatitis D: From Virology to New Therapies. *J. Adv. Res.* **2019**, *17*, 3–15. [CrossRef] [PubMed]
44. Loureiro, D.; Castelnau, C.; Tout, I.; Boyer, N.; Narguet, S.; Menasria Benazzouz, S.; Louis, Z.; Pons-Kerjean, N.; Giuly, N.; Marcellin, P.; et al. New Therapies for Hepatitis Delta Virus Infection. *Liver Int.* **2021**, *41*, 30–37. [CrossRef]
45. Asselah, T.; Loureiro, D.; le Gal, F.; Narguet, S.; Brichtler, S.; Bouton, V.; Abazid, M.; Boyer, N.; Giuly, N.; Gerber, A.; et al. Early Virological Response in Six Patients with Hepatitis D Virus Infection and Compensated Cirrhosis Treated with Bulevirtide in Real-Life. *Liver Int.* **2021**, *41*, 1509–1517. [CrossRef]
46. Chang, J.; Taylor, J.M. Susceptibility of Human Hepatitis Delta Virus RNAs to Small Interfering RNA Action. *J. Virol.* **2003**, *77*, 9728–9731. [CrossRef]
47. Li, H.; Mao, Q.; Li, Q. Inhibitory Effect of Replication and Expression of HDV by Antisense Oligodeoxynucleotides in H1 Delta 9 Cell. *Chin. J. Hepatol.* **1999**, *7*, 13–14.
48. Ye, X.; Tatenno, C.; Thi, E.P.; Kakuni, M.; Snead, N.M.; Ishida, Y.; Barnard, T.R.; Sof, A. M.J.; Shimada, T.; Lee, A.C.H. Hepatitis B Virus Therapeutic Agent ARB-1740 Has Inhibitory Effect on Hepatitis Delta Virus in a New Dually-Infected Humanized Mouse Model. *ACS Infect. Dis.* **2019**, *5*, 738–749. [CrossRef] [PubMed]
49. Chad, Y.-C.; Chang, M.-F.; Gust, I.; Lai, M.M.C. Sequence Conservation and Divergence of Hepatitis δ Virus RNA. *Virology* **1990**, *178*, 384–392. [CrossRef]
50. Rosenstein, S.P.; Been, M.D. Hepatitis Delta Virus Ribozymes Fold to Generate a Solvent-Inaccessible Core with Essential Nucleotides Near the Cleavage Site Phosphate. *Biochemistry* **1996**, *35*, 11403–11413. [CrossRef] [PubMed]
51. Perrotta, A.T.; Been, M.D. A Pseudoknot-like Structure Required for Efficient Self-Cleavage of Hepatitis Delta Virus RNA. *Nature* **1991**, *350*, 434–436. [CrossRef] [PubMed]
52. Chadalavada, D.M.; Cerrone-Szakal, A.L.; Bevilacqua, P.C. Wild-Type Is the Optimal Sequence of the HDV Ribozyme under Cotranscriptional Conditions. *RNA* **2007**, *13*, 2189–2201. [CrossRef]
53. Perrotta, A.T.; Been, M.D. Core Sequences and a Cleavage Site Wobble Pair Required for HDV Antigenomic Ribozyme Self-Cleavage. *Nucleic Acids Res.* **1996**, *24*, 1314–1321. [CrossRef] [PubMed]
54. Tanner, N.K.; Schaff, S.; Thill, G.; Petit-Koskas, E.; Crain-Denoyelle, A.-M.; Westhof, E. A Three-Dimensional Model of Hepatitis Delta Virus Ribozyme Based on Biochemical and Mutational Analyses. *Curr. Biol.* **1994**, *4*, 488–498. [CrossRef]
55. Timm, J.; Lauer, G.M.; Kavanagh, D.G.; Sheridan, I.; Kim, A.Y.; Lucas, M.; Pillay, T.; Ouchi, K.; Rey, L.L.; zur Wiesch, J.S.; et al. CD8 Epitope Escape and Reversion in Acute HCV Infection. *J. Exp. Med.* **2004**, *200*, 1593–1604. [CrossRef] [PubMed]
56. Leslie, A.J.; Pfafferoth, K.J.; Chetty, P.; Draenert, R.; Addo, M.M.; Feeney, M.; Tang, Y.; Holmes, E.C.; Allen, T.; Prado, J.G.; et al. HIV Evolution: CTL Escape Mutation and Reversion after Transmission. *Nat. Med.* **2004**, *10*, 282–289. [CrossRef]
57. Cotrina, M.; Buti, M.; Jardi, R.; Quer, J.; Rodríguez, F.; Pascual, C.; Esteban, R.; Guardia, J. Hepatitis Delta Genotypes in Chronic Delta Infection in the Northeast of Spain (Catalonia). *J. Hepatol.* **1998**, *28*, 971–977. [CrossRef]

Appendix. Publications and Congresses

Publications during this doctoral tesis

Cortese MF, González C, Gregori J, Casillas R, Carioti L, Guerrero-Murillo M, Riveiro-Barciela M, Godoy C, Sopena S, Yll M, Quer J, Rando A, Lopez-Martinez R, **Pacín Ruiz B**, García-García S, Esteban-Mur R, Tabernero D, Buti M, Rodríguez-Frías F. Sophisticated viral quasispecies with a genotype-related pattern of mutations in the hepatitis B X gene of HBeAg-ve chronically infected patients. *Sci Rep.* 2021 Feb 18;11(1):4215. doi: 10.1038/s41598-021-83762-4. PMID: 33603102; PMCID: PMC7892877.

Garcia-Garcia S, Cortese MF, Tabernero D, Gregori J, Vila M, **Pacín B**, Quer J, Casillas R, Castillo-Ribelles L, Ferrer-Costa R, Rando-Segura A, Trejo-Zahínos J, Pumarola T, Casis E, Esteban R, Riveiro-Barciela M, Buti M, Rodríguez-Frías F. Cross-sectional evaluation of circulating hepatitis B virus RNA and DNA: Different quasispecies? *World J Gastroenterol.* 2021 Nov 7;27(41):7144-7158. doi: 10.3748/wjg.v27.i41.7144. PMID: 34887634; PMCID: PMC8613647.

Cortese MF, Riveiro-Barciela M, Tabernero D, Rodriguez-Algarra F, Palom A, Sopena S, Rando-Segura A, Roade L, Kuchta A, Ferrer-Costa R, Quer J, **Pacín B**, Vila M, Casillas R, Garcia-Garcia S, Esteban R, Pumarola T, Buti M, Rodriguez-Frias F. Standardized Hepatitis B Virus RNA Quantification in Untreated and Treated Chronic Patients: a Promising Marker of Infection Follow-Up. *Microbiol Spectr.* 2022 Apr 27;10(2):e0214921. doi: 10.1128/spectrum.02149-21. Epub 2022 Apr 4. PMID: 35377229; PMCID: PMC9045303.

García-García S, Caballero-Garralda A, Tabernero D, Cortese MF, Gregori J, Rodriguez-Algarra F, Quer J, Riveiro-Barciela M, Homs M, Rando-Segura A, **Pacín-Ruiz B**, Vila M, Ferrer-Costa R, Pumarola T, Buti M, Rodriguez-Frias F. Hepatitis B Virus Variants with Multiple Insertions and/or Deletions in the X Open Reading Frame 3' End: Common Members of Viral Quasispecies in Chronic Hepatitis B Patients. *Biomedicines.* 2022 May 21;10(5):1194. doi: 10.3390/biomedicines10051194. PMID: 35625929; PMCID: PMC9139148.

Palom A, Rando-Segura A, Vico J, **Pacín B**, Vargas E, Barreira-Díaz A, Rodríguez-Frías F, Riveiro-Barciela M, Esteban R, Buti M. Implementation of anti-HDV reflex testing among HBsAg-positive individuals increases testing for hepatitis D. *JHEP Rep.* 2022 Jul 21;4(10):100547. doi: 10.1016/j.jhepr.2022.100547. PMID: 36052219; PMCID: PMC9425021.

Participation in congresses during this doctoral thesis

Pacin, B; Camps, G; Tabernero, D; Cortese, M.F; Gregori, J; Vila, M; Casillas, R; Garcia, S; Rando, A; Quer, J; Esteban, R; Riveiro-Barciela, M; Buti, M; González-Aseguinolaza, G; Rodriguez-Frias, F. Estudio del efecto del virus de la hepatitis D sobre la quasispecies del virus de la hepatitis B en modelos de ratones mediante secuenciación masiva. 46o Congreso Anual de la Asociación Española para el Estudio del Hígado (AEEH). Asociación Española para el Estudio del Hígado (AEEH). **2021**. España. **Poster Blast**.

Pacin-Ruiz, B; Cortese, M.F; Gregori, J; Vila, M; Garcia-Garcia, S; Tabernero, D; Casillas, R; Rando, A; Palom, A; Riveiro-Barciela, M; Rodriguez-Frias, F; Buti, M. Conservación y variabilidad del antígeno delta (HDAg) en pacientes con infección crónica. 46o Congreso Anual de la Asociación Española para el estudio del Hígado (AEEH). Asociación Española para el Estudio del Hígado (AEEH). **2021**. España. **Poster**.

Pacin, B; Camps, G; Tabernero, D; Cortese, M.F; Gregori, J; Vila, M; Casillas, R; Garcia, S; Rando, A; Quer, J; Esteban, R; Riveiro-Barciela, M; Buti, M; González-Aseguinolaza, G; Rodriguez-Frias, F. Study of the effect of the hepatitis D virus on the hepatitis B virus quasispecies in mice models by a next-generation sequencing approach. The International Liver Congress (ILC). European Association for the Study of the Liver (EASL). **2021**. United Kingdom. **Poster**.

Pacin-Ruiz, B; Cortese, M.F; Gregori, J; Vila, M; Garcia-Garcia, S; Tabernero, D; Casillas, R; Rando, A; Palom, A; Riveiro-Barciela, M; Rodríguez-Frias, F; Buti, M. Conservation, variability and evolution of Hepatitis Delta virus in antigen coding region. The International Liver Congress (ILC). European Association for the Study of the Liver (EASL). **2021**. United Kingdom. **Poster**.

Pacin-Ruiz, B; Camps, G; Cortese, M.F; Gregori, J; Garcia, S; Tabernero, D; Vales-Aranguren, A; Najarro, A; Olague-Micheltoarena, C; Rando, A; Quer, J; Esteban, R; Riveiro-Barciela, M; Buti, M; González-Aseguinolaza, G; Rodriguez-Frias, F. Inhibición del VHB mediada por el VHD en un modelo de super-infección en ratones: papel del INF tipo I. 47o Congreso Anual de la Asociación Española para el Estudio del Hígado (AEEH). Asociación Española para el Estudio del Hígado (AEEH). **2022**. España. **Oral communication**.

Pacin-Ruiz, B; Cortese, M.F; Sopena, S; Gregori, J; Garcia-Garcia, S; Tabernero, D; Casillas, R; Vila, M; Rando-Segura, A; Palom, A; Riveiro-Barciela, M; Rodríguez-Frias, F; Buti, M. Conservación y variabilidad de la quasispecies del VHD: región de la ribozima. 47o Congreso Anual de la Asociación Española para el estudio del Hígado

(AEEH). Asociacion Española para el estudio del Hígado (AEEH). **2022**. España. **Poster**.

Pacin- Ruiz, B; Cortese, M.F; Sopena, S; Gregori, J; Garcia-Garcia, S; Tabernero, D; Casillas, R; Vila, M; Rando-Segura, A; Palom, A; Riveiro-Barciela, M; Rodriguez-Frias, F; Buti, M. Hepatitis Delta virus quasispecies conservation and variability in ribozyme region. The International Liver Congress (ILC). European Association for the Study of the Liver (EASL). **2022**. United Kingdom. **Poster**.

Pacin-Ruiz, B; Camps, G; Cortese, M.F; Gregori, J; Garcia, S; Tabernero, D; Vales-Aranguren, A; Najarro, A; Olague-Micheltoarena, C; Rando, A; Quer, J; Esteban, R; Riveiro-Barciela, M; Buti, M; González-Aseguinolaza, G; Rodriguez-Frias, F. HDV-mediated inhibition of HBV in superinfection mouse model: the role of type I Interferon. The International Liver Congress (ILC). European Association for the Study of the Liver (EASL). **2022**. United Kingdom. **Poster**.

Cortese, M F; Garcia-Garcia, S; Pacin Ruiz, B; Rando-Segura, A; Tabernero, D; Riveiro Barciela, M; Buti, M; Rodriguez-Frias, F. The combination of Gapmers and/or siRNA as a potential hepatitis B virus gene therapy strategy against hepatitis B virus: preliminary in vitro results. XVI Congreso Nacional de Virología. Sociedad Española De Virología. **2022**. Spain. **Poster**.

Pacin-Ruiz, B; Camps, G; Cortese, M F; Gregori, J; Garcia, S; Tabernero, D; Vales-Aranguren, A; Najarro, A; Olague-Micheltoarena, C; Rando, A; Quer, J; Esteban, R; Riveiro-Barciela, M; Buti, M; González-Aseguinolaza, G; Rodriguez-Frias, F. HDV-mediated inhibition of HBV in a superinfection mouse model: the role of type I Interferon. XVI Congreso Nacional de Virología. Sociedad Española De Virología. **2022**. Spain. **Oral communication**.

Pacin-Ruiz, B; Camps, G; Cortese, M F; Gregori, J; Garcia, S; Tabernero, D; Vales-Aranguren, A; Najarro, A; Olague-Micheltoarena, C; Rando, A; Quer, J; Esteban, R; Riveiro-Barciela, M; Buti, M; González-Aseguinolaza, G; Rodriguez-Frias, F. HDV-mediate inhibition of HBV in a superinfection mouse model: the role of type I interferon.. 2022 International HBV meeting. Hepatitis B Foundation. **2022**. France. **Oral communication**.

Pacin- Ruiz, B; Cortese, M F; Palom, A; Garcia-Garcia, S; Tabernero, D; Rando-Segura, A; Najaro, A; Vila, M; Gregori, J; Quer, J; Riveiro-Barciela, M; Buti, M; Rodriguez-Frias, F. Studying by next-generation sequencing the delta antigen quasispecies: complexity, variability and epitopes binding prediction. Delta Cure: First International Congress. **2022**. Italy. **Poster**.

Pacin-Ruiz, B; Rando- Segura, A, Gregori, J, Palom, A, Garcia-Garcia, Tabernero, D, Quer, J ; Riveiro-Barciela, M; Buti, M. Rodríguez-Frias, F, Cortese, M.F Diferencias en la evolución de las quasiespecies del virus de la hepatitis delta: estudio comparativo entre el antígeno delta y la ribozima. 48o Congreso Anual de la Asociacion Española para el estudio del Hígado (AEEH). Asociacion Española para el estudio del Hígado (AEEH). **2023**. España. **Poster**.

Pacin-Ruiz, B; Cortese, M.F, Palom, A, Gregori, J, Garcia-Garcia, Tabernero, D, Rando-Segura, A, Najarro, A, Vila, V, Riveiro-Barciela, M; Buti, M. Rodríguez-Frias, F. Análisis por secuenciación masiva de la quasiespecies del antígeno delta del virus de la hepatitis delta: complejidad, variabilidad y estudios de la capacidad de unión epitópica. 48o Congreso Anual de la Asociacion Española para el estudio del Hígado (AEEH). Asociacion Española para el estudio del Hígado (AEEH). **2023**. España. **Poster**.