




Universitat Autònoma de Barcelona

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:  http://cat.creativecommons.org/?page_id=184

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:  <http://es.creativecommons.org/blog/licencias/>

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

Departamento de Bioquímica y Biología Molecular

Programa de doctorado en Bioquímica, Biología Molecular y Biomedicina

Tesis doctoral

Complejidad y conservación de la quasispecies del virus de la hepatitis B analizada mediante secuenciación masiva en el gen X. Asociación con la actividad replicativa e identificación de regiones híper-conservadas en el genoma viral

Autora de la tesis

Carolina González Fernández

Directores de la tesis

Dr. Francisco Rodríguez-Frías Dra. María Buti Ferret Dr. David Tabernero Caellas

Tutor de la tesis

Dr. Francisco Rodríguez-Frías

Universitat Autònoma de Barcelona, Bellaterra, 2019

“La mente es como una paracaídas...solo funciona si la tenemos abierta”

Alberth Einstein

Índice

ÍNDICE	5
AGRADECIMIENTOS	9
ABREVIATURAS	13
RESUMEN	17
CAPITULO 1: INTRODUCCIÓN	18
1. Virología molecular del virus de la hepatitis B	21
1.1 Historia	21
1.2 Taxonomía y origen	21
1.3 Estructura de las partículas virales y subvirales	22
1.3.1 Partículas Dane:	22
1.3.2 Partículas subvirales y partículas virales incompletas:	23
1.4 Estructura del genoma del VHB	25
1.4.1 Marcos de lectura abiertos (ORF) y proteínas virales	26
1.5 Regulación de la transcripción	48
1.5.1 Trancritos, promotores y enhancers	48
1.5.2 Otros elementos reguladores	52
1.6 Ciclo replicativo	53
1.6.1 Entrada del virus	54
1.6.2 Transporte de las cápsides a través del citoplasma hacia el núcleo y liberación del ADNrc en el núcleo	55
1.6.3 Formación del ADNccc	56
1.6.4 Transcripción Viral, Formación de Partículas Core	57
1.6.5 Maduración y secreción de viriones y partículas subvirales	61
2. Variabilidad genética	62
2.1 Complejidad de quasiespecies	62
2.1.1 Definición y significado clínico	62
2.1.2 Índices para el estudio de la complejidad de la quasiespecie	65
2.1.3 Estudio de la conservación de la quasiespecie	67
2.2 Herramientas para el estudio de la quasiespecies	69
2.2.1 Métodos indirectos:	69
2.2.2 Técnicas de secuenciación de primera generación o secuenciación directa:	69
2.2.3 Secuenciación masiva o next generation sequencing (NGS)	70

2.2.4	Secuenciación de cuarta generación.....	82
2.2.5	Aplicaciones y limitaciones de NGS.....	82
2.3	Genotipos y subgenotipos: impacto en la biología y la patogénesis.....	83
2.3.1	Distribución geográfica.....	85
2.3.2	Recombinación intergenotípica.....	87
2.3.3	Diferencias genéticas, clínicas y evolutivas entre genotipos del VHB.....	88
2.4	Principales variantes genéticas de interés clínico.....	91
2.4.1	Variabilidad en el ORF preS/S.....	92
2.4.2	Variabilidad en el ORF P.....	93
2.4.3	Variabilidad en el ORF PC/Core.....	94
2.4.4	Variabilidad en el ORF X.....	95
3.	Epidemiología.....	97
3.1	Mecanismos de transmisión.....	99
3.2	Historia natural de la infección.....	100
3.2.1	Infección aguda.....	101
3.2.2	Infección crónica.....	102
3.2.3	Inmunopatogénesis del VHB.....	105
4.	Tratamiento anti-VHB.....	106
4.1	¿Qué se entiende por curación de la infección del VHB?.....	106
4.2	Objetivo actual del tratamiento antiviral.....	107
4.3	Estrategias de tratamiento.....	108
4.4	Nuevas terapias antivirales.....	109
4.4.1	Antivirales de acción directa (DAA) (Figura 39).....	110
4.4.2	La respuesta inmune al VHB y las implicaciones para terapias inmunomoduladoras..	116
4.5	Perspectivas de futuro.....	119
	CAPÍTULO 2: HIPÓTESIS Y OBJETIVOS.....	121
1.	Hipótesis.....	123
2.	Objetivos.....	124
	CAPÍTULO 3: PRIMER ESTUDIO.....	125
1.	Introducción.....	127
2.	Hipótesis y objetivo.....	128
2.1	Hipótesis:.....	128
2.2	Objetivo:.....	129
3.	Resumen del estudio.....	129
3.1	Diseño del estudio:.....	129

3.2 Análisis bioinformático de los datos de secuenciación	131
3.2.1.Filtraje de los datos de secuenciación masiva:	131
3.2.2.Genotipado de los haplotipos obtenidos:.....	133
3.2.3.Análisis de la conservación en la región analizada:	134
3.3 Análisis de las secuencias de NGS obtenidas y resultados de genotipado.....	135
3.4 Conservación de la secuencia de nucleótidos del <i>HBX</i> en la región de interés.....	135
3.5 Conservación de la secuencia de aminoácidos HBx	138
4. Conclusiones	140
5. Manuscrito completo.....	141
6. Material suplementario.....	155
CAPÍTULO 4: SEGUNDO ESTUDIO.....	157
1. Introducción	159
2. Hipótesis y objetivo.....	159
2.1.Hipótesis.....	159
2.2.Objetivo.....	160
3. Resumen de estudio.....	160
3.1.Diseño del estudio	160
3.2.Análisis bioinformático de los datos de secuenciación	162
3.3.Análisis de conservación de las quasiespecies	164
3.4.Análisis de la diversidad y complejidad de la QS	164
3.5.Análisis de las mutaciones	165
3.6.....Análisis de la complejidad de la quasiespecies. Comparación entre pacientes IC y los otros grupos.....	166
3.7.Estudio de la conservación de la secuencia.....	168
3.8.Identificación de mutaciones específicas de los pacientes IC	171
4. Conclusiones	173
5. Manuscrito completo.....	174
6. Material suplementario.....	207
CAPÍTULO 5: DISCUSIÓN.....	213
1. Búsqueda de las regiones híper-conservadas en la región 5' del <i>HBX</i>	216
2. Estudio de la QS de la región 5' de la proteína HBx: el caso de la Infección crónica por el VHB HBeAg (-).	220
3. Estudios de las mutaciones genotipo-específicas.	222
4. Resumen.....	224
5. Perspectivas de futuro:	224

CAPÍTULO 6: CONSCLUSION	227
1. Conclusiones:	229
ANEXO	231
1. Anexo 1	233
2. Anexo 2	235
BIBLIOGRAFÍA	237
1. Bibliografía	239

Agradecimientos:

Aquí estoy, después de casi 6 años del comienzo de este proyecto. Pensaba que esta sección sería la más fácil de todas, que saldría sola, pero la verdad es que no tengo ni idea de por dónde empezar. No me había planteado lo difícil que resulta transmitir lo mucho que han significado para mí a todas esas personas que han ayudado a conseguir que este trabajo fuera una realidad. Toda la gente que me conoce bien sabe que las letras nunca han sido lo mío y creo que esto me va a costar más incluso que secuenciar al grupo de la infección crónica HBeAg negativo. Aun así, es algo que realmente quiero intentar hacer bien.

En primer lugar, quisiera agradecer el tiempo dedicado y las enseñanzas que me han regalado mis tres directores de tesis el Dr. Francisco Rodríguez-Frías, la Dr. María Buti y el Dr. David Tabernero. En especial al Dr. Francisco Rodríguez-Frías (para muchos, Paco), por haber confiado en mí desde el principio, por animarme siempre, por elogiarme cuando lo necesitaba y por pararme los pies en los momentos necesarios. Pero, sobre todo, por transmitirme que creía en mí, por enseñarme que con esfuerzo todo se puede, que no hay que desanimarse ante los inconvenientes, que hay que afrontarlos y no detenerse. Por todo esto, estoy profundamente agradecida, Paco. Sin ti esta tesis no podría haber tenido lugar.

También me gustaría resaltar y agradecer el papel del Dr. Tabernero, que pese a que no ha sido mi director desde el principio, siempre se ha comportado como si lo fuera, mostrando desde el inicio su apoyo incondicional. Nunca olvidaré que fue el primero que me enseñó la metódica del laboratorio, que fue él quien me explicó con esmero cómo hacer mi primera PCR, base fundamental de este trabajo, y sin la que jamás hubiese logrado llevar a cabo esta tesis que por fin cierro.

Además, quisiera dar las gracias al Dr. Josep Quer, a la Dra. Anna Messeguer y a la Dra. María Cubero por dedicarme su tiempo y ofrecerme sus reflexiones, su perspectiva más amplia, que en muchas ocasiones ha sido de gran ayuda para tener una visión más clara del proyecto durante las sesiones de la comisión de seguimiento. En especial, me gustaría reconocer al Dr. Quer su cercanía y su cariño y agradecerle que me hubiese marcado una fecha para terminar la tesis, ya que esa determinación resultó ser el empujón final que me hacía falta.

AGRADECIMIENTOS

Tampoco puedo dejar pasar la oportunidad de dar las gracias a mis compañeros de laboratorio: Chari, Marçal, Marta, Irene, Cris, Sara y Gerard que siempre me han ayudado en todo lo que he necesitado, que me han hecho reír y en los que he encontrado un motivo para ir con ánimo y ganas al laboratorio, incluso cuando estaba rendida después de un turno de noche. Y, por supuesto, gracias a Francesca, por ayudarme, por ilustrarme, por explicarme tanto, con su estilo fresco y conciso, y enseñarme a vender mis ideas como si fueran las mejores de mundo. Por todo ello muchas gracias. La mitad de esta tesis es tuya.

Como no podría ser de otra manera, quiero agradecer de corazón a mi familia, en especial a mis abuelos y a mi hermano Fon, por apoyarme siempre en todas mis ideas, por enseñarme que el esfuerzo es necesario, que al final merecería la pena. Pero por encima de todo eso, GRACIAS -así, mayúsculas- por creer en mí, por demostrarme el orgullo que sienten y por aguantarme, sostenerme y tirar de mí siempre que he dicho que ya no podía más y que quería dejarlo.

También me gustaría agradecer esta tesis, a mis amigas; Pilar, Silvia, Andrea y Alicia, porque siempre han estado, ahí de todas las maneras posibles: se han preocupado por mí, me han animado cuando ya estaba saturada, me han comprendido cuando más aislada y “menos sociable” estaba por los picos de trabajo y, sobre todo, me han enseñado que no hay nada que no se solucione ni que más despeje que una birra entre amigas. La mejor cura para todo.

Además, en este texto de agradecimiento, no podría olvidarme de “nosotras”, que más que mis amigas de siempre son mi familia, las personas que siempre han estado, están y seguirán estando ahí, tanto para lo bueno como para lo malo. Las mejores aliadas para cualquier plan, animadas para una escapada cuando estoy más que saturada, que siempre se preocupan por mí e intentan animarme y apoyarme cuando lo necesito. Gracias por soportarme y escucharme cuando tanto cuando lo necesitaba y como cuando yo no era consciente de esa necesidad, tan liberadora. Y en especial a Antía, mi amiga de siempre, mi compañera de piso durante muchos años y familia, que pese a tener nuestros altibajos, siempre que la he necesitado ha estado ahí para mí y se ha prestado a echarme una mano con esta parte, ya que conoce perfectamente mis limitaciones. Anti, muchas gracias.

También me gustaría agradecerle de una manera especial a Sergi, mi pareja, por estar ahí siempre para mí, por su paciencia infinita cuando me he pasado días enteros trabajando delante del ordenador sin apenas descanso. Ha soportado todas las “neuras” que me han

AGRADECIMIENTOS

surgido durante este largo proceso, pero, lo más importante es que ha hecho que sienta que está orgulloso de mí y que estoy haciendo algo grande.

A todos vosotros muchas gracias

Abreviaturas

A

aa	amono-ácido
ADN	Ácido desoxirribonucleico
ADNccc	ADN covalentemente cerrado
ADNdsl	ADN lineal de doble cadena
ADNmc	ADN monocatenario
ADNpol/RT	Actividad de la ADN polimerasa y retrotranscriptasa de la proteína P del VHB
ADNrc	ADN relajado circular
ADN-VHB	Genoma del VHB
ADV	Adefovir
AG	Aparato de golgi
Ag	Antígeno
ALT	Alanina aminotransferasa
AN	Análogos nucleós(t)idos
anti-HBc	Anticuerpos contra el antígeno core del VHB
anti-HBe	Anticuerpos contra el antígeno e del VHB
anti-HBs	Anticuerpos contra el antígeno de superficie del VHB
ARN	Ácido ribonucleico
ARNm	ARN mensajero
	ARN
ARNpc	precore
ARNpg	ARN pregenómico

C

C	Región core del genoma del VHB
C/EBP α	<i>CAAT/enhancer-binding protein α</i>
CCD	<i>Charge-coupled device</i>
CHB	Hepatitis B crónicas
CI	Contenido de información
CREB	<i>Cyclic-AMP-response element binding protein</i>
CTL	Células T citotóxicas
CURS	<i>Core upstream regulatory sequence</i>
Cys	Cisteína

D

DDB1	DNA damage-binding protein 1
DHVB	<i>Duck hepatitis B virus</i>
dNTP	Desoxirribonucleótido trifosfato
DR1/DR2	Repeticiones directas 1 y 2

E

emPCR	Amplificación clonal en la PCR en emulsión
EMT	Epitelio-mesénquima
EN1/EN2	Enhancer 1 y 2
ERH	Elemento de respuesta a hormonas
ESCRT	<i>Endosomal sorting complexes required for transport</i>
ETV	Entecavir
eVHB	Secuencias virales endógenas

F

FT	Factor de transcripción
----	-------------------------

G

Gen/A	Genotipo A del VHB
Gen/B	Genotipo B del VHB
Gen/C	Genotipo C del VHB
Gen/D	Genotipo D del VHB
Gen/E	Genotipo E del VHB
Gen/H	Genotipo H del VHB
Gen/I	Genotipo I del VHB
Gen/J	Genotipo J del VHB

H

HBc	Monómero core
HBcAg	Antígeno core
HBeAg	Antígeno e del VHB
HBsAg	Antígeno de superficie
HBx	Proteína X
HBX	Gen de la región X del genoma del VHB
HCC	Hepatocarcinoma
HFN	Factores hepáticos nucleares
H _{GS}	Índice Gini-Simpson
HSC	Células estrelladas hepáticas
H _{SN}	Entropía de Shannon
Hsp	Heat-shock proteínas
HSPG	Heparán sulfato

I

IL	Interleucina
INF	Interferón
iNOS	Sintetasa inducible del óxido nítrico

K

KDA	Quilodaltons
-----	--------------

L

LAG	Gen de activación de linfocitos
LAM	Lamivudina
LHBs	Proteína de superficie grande

M

Mf	Frecuencia de mutación
MHBs	Proteína de superficie mediana
MHR	<i>Major hydrophilic region</i>
MT	Microtúbulos
MTA1/HDAC	Proteína 1 asociada a metastasis/histona deacetilasa

N

NGS	Next generation sequencing
NPC	Poro nuclear
NSL	Señal de localización nuclear
nt	Nucleótido
NTCP	Péptido cotrasportador de sodio-taurocolato

O

OBI	Infección oculta del VHB
ORF	Marcos de lectura abiertos

P

P	Proteína codificada por el ORF P (polimerasa)
pb	Pares de bases
PC/C	Región preCore/core del genoma del VHB
PD	Célula programada
Peg-INF	Interferón pegilado
π	Diversidad nucleótida
Ppi	Pirofosfato inorgánico
PRE	Elemento regulador post-transcripcional
preS/S	Región preS/superficie del genoma del VHB
PT	Dominio que codifica la proteína terminal del genoma del VHB

Q

QS	Quasiespecie
----	--------------

R

RE	Retículo endoplasmático
RENAVE	Red Nacional de Vigilancia Epidemiológica
RFLP	Análisis de polimorfismos de longitud de fragmento de restricción
RFMP	Polimorfismo de masa de fragmento de restricción
RH	Dominio ribonucleasa H de la polimerasa del VHB
RN	Receptores nucleares
RT	Transcriptasa reversa

S

SBL	Secuenciación por ligación
SBS	Secuenciación por síntesis
SHBs	Proteína de superficie pequeña
SLC10A1	<i>Solute carrier family 10A1</i>
SNA	Secuenciación por síntesis que adiciona nucleótidos individuales
SP	Región espaciadora de la polimerasa del VHB
SVP	Partículas subvirales

T

TAF	Tenofovir Alafenamide
TALENs	Transcription activator-like effector nucleases
TDF	Tenofovir
TDP	Tirosil-ADN fosfodiesterasas
TGF- β	Transforming growth factor β
Th	Linfocitos T helper
Thr	Treonina
TLR	<i>Toll-like receptor</i>
TM	Dominio transmembrana
TNF	Factor necrosis
TVD	Telbivudina

U

UDPS	<i>Ultra-deep pyrosequencing</i>
UI/ml	Unidades internacionales por mililitro
URR	<i>Upper regulatory region</i>

V

VEGF	Factor endotelial del crecimiento vascular
VHB	Virus de la hepatitis B
VHC	Virus de la hepatitis C
VHD	Virus de la hepatitis D

X

X	Región X del genoma del VHB
---	-----------------------------

Z

ZFN	Nucleasas con dedos de zinc
-----	-----------------------------

Resumen

La proteína X del virus de la hepatitis B (HBx), codificada por el gen X de este virus (*HBX*), es crucial para la replicación del VHB y regula la expresión de múltiples genes del huésped. Por este motivo deben existir zonas altamente conservadas a nivel nucleotídico (*HBX*) y/o aminoacídico (HBx), esenciales para su actividad reguladora. Sin embargo, las inserciones y deleciones descritas en el extremo C-terminal de esta proteína, parecen limitar la posibilidad de que estas zonas estén en la región 3' de *HBX*, por eso se ha escogido estudiar mediante la tecnología de secuenciación masiva la conservación del extremo 5' de *HBX*.

En esta región se han identificado 2 regiones híper-conservadas, que además podrían ser utilizadas como potenciales dianas terapéuticas para una nueva estrategia de tratamiento antiviral basada en la terapia génica contra la HBx mediante el silenciamiento génico. Este tratamiento podría ser útil para conseguir la “cura funcional” de la infección en todos los estadios clínicos y en presencia de todos los genotipos virales. Además de la conservación del extremo 5' de *HBX* en esta tesis también se ha estudiado en profundidad la complejidad y variantes específicas de la quasiespecies (QS) de esta región en pacientes en diferentes etapas clínicas de la infección crónica por el VHB HBeAg (-), para investigar su relación con los mecanismos que permiten controlar la replicación viral. El estudio de la complejidad y la variabilidad de la QS de un grupo de pacientes con infección crónica HBeAg (-) (IC) evidenció, en estos pacientes, una QS más conservada y compleja que en los pacientes con hepatitis crónica (en presencia o no de lesiones hepáticas como la cirrosis y el HCC), caracterizada por la presencia de un alto número de haplotipos altamente mutados probablemente a muy bajas frecuencias, por lo que no afectan a la conservación. Asimismo, se ha evidenciado un grupo de mutaciones genotipo-específicas. En concreto se ha descrito un patrón mutacional en los haplotipos de genotipo D de los IC asociado a una probable reducción de la expresión viral. Este patrón mutacional genotipo-específico evidencia la necesidad de genotipar adecuadamente el virus en el seguimiento de los pacientes infectados por VHB.

En resumen, el extremo 5' de *HBX* contiene regiones híper-conservadas que podrían ser esenciales para la función de la HBx y explotadas como diana para el tratamiento antiviral basado en terapia génica. Además, la QS de los pacientes IC presenta características diferenciales de complejidad y conservación. Estas características junto con patrones de mutaciones genotipo-específicas, podrían estar relacionados con la baja replicación del VHB en estos pacientes.

Abstract

The hepatitis B virus X protein (HBx), encoded by the X gene of this virus (*HBX*), is crucial for HBV replication and regulates the expression of multiple host genes. For this reason, there must be highly conserved areas at nucleotide (*HBX*) and/or amino acid (HBx) level, essential for its regulatory activity. However, the insertions and deletions described in the C-terminal end of this protein seem to limit the possibility that these areas are located in the 3' region of *HBX*. By this reason, the 5'-end of *HBX* has been chosen to study its conservation by next-generation sequencing.

In this region, 2 hyper-conserved regions have been identified. These regions could be used as potential therapeutic targets for a new antiviral treatment strategy, based on gene therapy against HBx through gene silencing. This treatment could be useful to achieve “functional cure” of the HBV infection in all clinical stages and in the presence of all viral genotypes. In addition to the conservation of the 5' end of *HBX*, in this thesis the complexity and specific variants of the quasispecies (QS) of this region have also been studied in depth in patients at different clinical stages of HBeAg (-) chronic HBV infection, to investigate its relationship with the mechanisms that allow to control viral replication. The study of the complexity and variability of the QS of a group of patients with HBeAg (-) chronic infection (IC) showed a more conserved and complex QS in these patients than in patients with chronic hepatitis (with or without liver lesion such liver cirrhosis and HCC), characterized by the presence of a high number of highly mutated haplotypes, probably at very low frequencies, so they do not affect conservation. Likewise, a group of genotype-specific mutations has been evidenced. In particular, a mutational pattern has been described in genotype D haplotypes of IC patients, probably associated with a reduction of viral expression. This genotype-specific mutational pattern demonstrates the need to adequately genotype the virus in the HBV infected patients follow-up.

In summary, the 5'-end of *HBX* contains hyper-conserved regions that could be essential for the HBx function and could be exploited as a target for antiviral treatment based on gene therapy. In addition, the IC patients' QS presents differential characteristics of complexity and conservation. These characteristics, together with patterns of genotype-specific mutations, could be related to the low replication of HBV in these patients.

CAPÍTULO 1: INTRODUCCIÓN

1. Virología molecular del virus de la hepatitis B

El virus de la hepatitis B (VHB) es un patógeno muy común que infecta de forma crónica a más de 257 millones de personas en el mundo y que tienen un alto riesgo de desarrollar cirrosis y/o carcinoma hepatocelular (HCC), siendo así una causa importante de morbilidad, y llegando a alcanzar las cifras de 887000 muertes en 2015^{1,2}. Muchas de estas personas no saben que son portadoras del virus y por lo tanto son una fuente importante de transmisión. A pesar de que se dispone de una vacuna efectiva, y de las campañas de prevención, esta infección, continúa siendo un grave problema de salud mundial debido a su alta prevalencia.

1.1 Historia

El primer registro que se tiene de una epidemia causada por el VHB fue por Lurman en 1885, después de un brote de viruela en el que se vacunaron a unos obreros en Alemania, con la linfa de otras personas. Después de varios meses, muchos de ellos desarrollaron ictericia y fueron diagnosticados de hepatitis³. En los años 30 y 40 se investigaron varios casos de hepatitis transmitida por la sangre, después de brotes de hepatitis como consecuencia de la vacunación contra el sarampión, paperas y fiebre amarilla, ya que estas vacunas contenían suero o plasma agregado como un estabilizador, infectado con el virus. Posteriormente, se identificaron las transfusiones sanguíneas y el uso repetido de agujas no esterilizadas como causa de epidemias de hepatitis. El virus fue finalmente descrito por Blumberg, después de descubrir el antígeno Australia^{4,5} (posteriormente llamado antígeno de superficie de la hepatitis B), en aborígenes australianos, lo cual lo hizo merecedor del Premio Nobel de Medicina en 1976⁶. En 1970, Dane et al. visualizaron por primera vez las partículas virales mediante microscopía electrónica⁷ y en 1979 el genoma viral fue secuenciado en su totalidad⁸.

1.2 Taxonomía y origen

El VHB humano es el miembro prototipo de la familia *Hepadnaviridae* (*Hepa*; infectan al hígado y *dna* virus de ácido desoxirribonucleico). La familia hepadnavirus se divide en dos géneros: *Orthohepadnavirus* y *avihepadnavirus*, aunque existe un grupo de miembros de esta familia que no se han asignados a ninguno de estos dos géneros. El género *orthohepadnavirus* incluye ocho especies: el VHB cuyo hospedador natural es el ser

CAPÍTULO 1. INTRODUCCIÓN

humano y este es el único reservorio para infectar a otros individuos de su especie⁹, el virus de la hepatitis de las marmotas (woodchuck hepatitis virus [WHV]), el virus de las ardillas terrestres (ground squirrel hepatitis virus [GSHV]), y el virus de la hepatitis del mono lanudo (woolly monkey hepatitis virus, (WHMV) y varios virus que afectan a murciélagos (long-fingered bat hepatitis B virus (LFBHBV), Pamaona Bat hepatitis B virus (PBVHB), roundleaf bat hepatitis B virus (RLBHBV) y tenmaking hepatitis B virus (TMBHBV). El género avihepadnavirus comprende el virus de la hepatitis B de la garza real (heron hepatitis B virus, HHBV), el virus de la hepatitis B del pato doméstico (duck hepatitis B virus, DHVB), y el virus del loro (parrot hepatitis B virus, PBVHB) aunque también puede afectar a otras especies de aves como las ocas, las grullas, y a los anades reales salvajes¹⁰. Entre los hepadnavirus sin un género asignado está el virus de la hepatitis B del pez lechón blanco (WSHBV), este virus diverge con más del 30% de todos los hepadnavirus conocidos previamente¹¹. Los hepadnavirus son los virus más pequeños que infectan animales, comparten una organización genómica similar, tienen tropismo por los mismos órganos y utilizan una estrategia de replicación de su genoma única que, a pesar de ser virus de ADN, incluye un paso de transcripción inversa de un ARN mensajero viral¹².

Se estima que los hepadnavirus divergieron hace unos 400 millones de años de un grupo de virus que se han descubierto hace poco en peces óseos denominado *nackednavirus*, con el que comparten algunas características clave de su biología¹³. Con el descubrimiento de las secuencias virales endógenas relacionadas con el VHB (eVHB) se ha calculado que las aves y los hepadnavirus han coexistido durante 70 millones de años ya que se sabe que las eVHB datan de 82 a 12.1 millones de años¹³. Incluso se han reportado eVHBs en genomas de cocodrilos, serpientes y tortugas, cuyas endogenizaciones ocurrieron antes de las aves; como por ejemplo, la eVHB de la tortuga que calcula que es de hace 200 millones de años¹³.

1.3 Estructura de las partículas virales y subvirales

1.3.1 Partículas Dane:

Son las partículas virales circulantes en sangre, de entre 40-42 nm de diámetro, que están compuestas por una envuelta externa lipídica y una cápside interna proteica de entre 30-32 nm, dentro de la cual se encuentra el genoma viral, la polimerasa viral y otras proteínas del hospedador que participan en el ciclo replicativo (figura 1). La envuelta, derivada de la célula a la que infecta, está formada por una bicapa de

fosfolípidos y contiene el HBsAg, que engloba tres tipos diferentes de glicoproteínas de superficie; grande (LHBs), mediana (MHBs) y pequeña (SHBs)¹⁴. La nucleocápside está compuesta por 180-240 monómeros de una única cadena polipeptídica de 20 kDa, con conformación icosaédrica, conocida como antígeno core (HBcAg)^{15,16}. En el interior de la cápside se encuentra el genoma viral, formado por una molécula de ADN parcialmente bicatenaria, que interacciona con la polimerasa viral mediante una unión estabilizada por un complejo de proteínas de la familia de las chaperonas^{17,18}. Además dentro de las cápsides del VHB también se han identificado otras proteínas del hospedador, como las isoformas específicas de la proteína quinasa C (PKC) $\alpha / \beta / \gamma$ ^{19,20}.

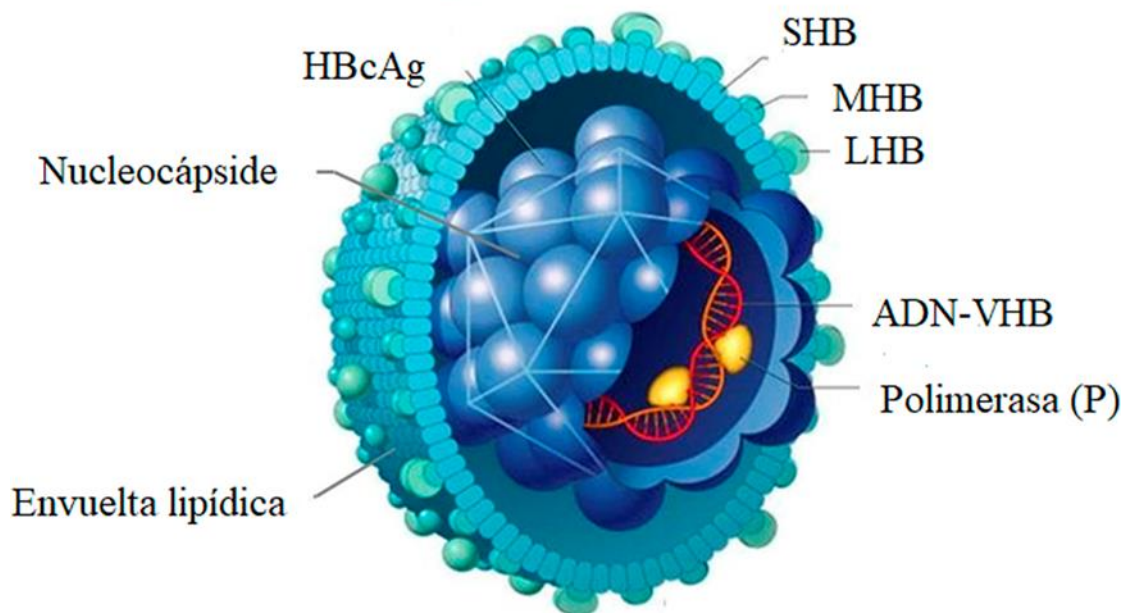


Figura 1 Partículas virales (partícula de Dane) de VHB circulantes en sangre²¹. Imagen modificada de © 2002 James A. Perkins, Medical and Scientific illustrations, (<http://people.rit.edu/japfaa/infectious.html>).

1.3.2 Partículas subvirales y partículas virales incompletas:

Además de los viriones (partículas Dane), las células infectadas secretan 3 clases de partículas que se caracterizan por no contener el ADN viral.

Partículas subvirales (les falta tanto el genoma como la cápside): El HBsAg puede ensamblarse espontáneamente dando lugar a unas partículas lipoproteicas, llamadas partículas subvirales. Estas partículas no incluyen nucleocápside ni proteínas del

huésped por lo que no son infecciosas, y se secretan al suero de los portadores del VHB en un gran exceso respecto a los viriones completos (partículas Dane). Las partículas subvirales pueden presentar dos morfologías diferentes; quasiesféricas o filamentosas (figura 2). Las formas quasiesféricas de unos 20 nm de diámetro (más pequeño que el de las partículas virales) y de simetría octaédrica están formadas por 48 dímeros de SHBs²², en el suero de una persona infectada pueden ser de hasta 10^4 veces más abundantes que las propias partículas virales (hasta 10^{14} copias/ml)²³. Estas partículas están compuestas esencialmente por SHBs y MHBs y se ensamblan en las membranas entre el compartimento post-Retículo endoplasmático (RE) y el pre-Aparato de Golgi (AG)²⁴. Las formas filamentosas tienen un diámetro de 20 nm y se caracterizan por tener una longitud variable (figura 1), las cuales pese a que se excretan en una cantidad inferior a las quasiesféricas, siguen estando presentes en el suero en una concentración 10^2 veces más alta que las partículas virales (hasta 10^{11} copias / mL)^{16,21}. Las formas filamentosas contienen un porcentaje más elevado de LHBs que las quasiesféricas y se ensamblan en los cuerpos multivesiculares (MVB), igual que las viriones completos y las partículas virales incompletas²⁵. Se cree que las partículas subvirales pueden actuar como señuelos para modular la respuesta inmune del huésped y que son esenciales para la entrada del virus en los hepatocitos¹³.

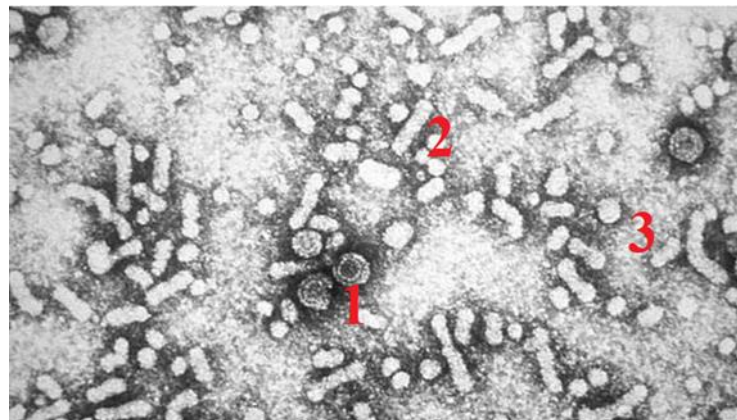


Figura 2 Partículas subvirales y partículas virales completas del VHB vistas con microscopía electrónica. 1. Partícula viral. 2. Partícula subviral filamentosas. 3. Partícula quasiesférica. Imagen modificada <http://www.abc.es/salud/enfermedades/abci-virus-hepatitis-dificil-erradicar-organismo-201603161651-noticia.html>.

Partículas vacías (les falta solo el genoma): recientemente se han descubierto los viriones vacíos, que contienen las proteínas de la superficie que encierran una cápside viral vacía, sin el genoma, y se encuentran típicamente en niveles 10^2 veces superiores a los viriones completos (partículas Dane) en la sangre de individuos infectados (10^{11} copias / mL)²³.

Partículas de ARN: También se ha informado que se pueden secretar partículas que contienen ARN del VHB aunque a un nivel más bajo, de 10^2 a 10^3 veces más bajo que las partículas Dane (10^6 copias / ml)²³.

El papel de las partículas vacías y de ARN en el ciclo del VHB todavía no se conoce.

1.4 Estructura del genoma del VHB

El genoma del VHB está constituido por una molécula de ADN circular de pequeño tamaño, parcialmente de doble cadena, que consta de una cadena larga (negativa) de unas 3.200 bases y una cadena corta (positiva) de 1.700 a 2.800 bases^{26,27}. El genoma se mantiene en conformación circular a pesar de no estar cerrado covalentemente (ADNrc) gracias a una región cohesiva en los extremos 5' en ambas cadenas, las repeticiones directas 1 (DR1, nt 1824-UUCACCUCUG-1834, en el extremo 5' de la hebra(-)) y 2 (DR2, nt 1590-UUCACCUCUGC-1600, en el extremo 3' de la hebra(+)) dos regiones cortas muy conservadas de 11 pb²⁸. Además, la polimerasa viral permanece unida covalentemente al extremo 5' de la hebra (-), la cual tiene en sus extremos 5' y 3' dos secuencias de 8-9 nts denominadas redundancias terminales (r), críticas para la síntesis de ADNrc. La hebra (+) tiene un pequeño ribonucleótido, el cual es un residuo de la transcripción inversa del genoma viral, en su extremo 5' (Figura 3). En el núcleo del hepatocito, el ADN viral pasa de la conformación ADNrc a ser circular y covalentemente cerrado (ADNccc) por acción de la maquinaria de reparación del genoma del huésped¹³. El ADNccc tiene una estructura superenrollada y se organiza en una estructura muy similar a la cromatina, interaccionando con histonas del huésped y el HBcAg del propio VHB^{29,30}. Esta forma del genoma viral permanece en el hepatocito durante toda la vida de éste y sirve como molde para la síntesis de los diferentes transcritos del VHB, mediante la acción de la ARN polimerasa II del huésped, por lo que lo podemos considerar el reservorio natural del propio virus^{18,31}.

En ocasiones, el ADN del VHB puede integrarse en el genoma del hospedador, pudiendo alterar los genes del hospedador, cambiar sus niveles de transcripción, e incluso puede

producir proteínas de fusión oncogénicas³². La fuente principal de ADN del VHB integrado es el ADN del VHB lineal de doble cadena (ADNds1), que surge como un producto de replicación aberrante debido a un proceso alternativo de síntesis de la cadena (+) conocido como *priming in situ*^{13,32}. Aunque el ADN del VHB integrado está asociado con el HCC, su función en el mantenimiento de la infección por VHB es desconocida (ver en el Capítulo 1 el apartado de 1.6.4 *transcripción viral. Formación de partículas core*).



Figura 3 Representación del ADNrc y el ADNccc. Imagen modificada de³³

1.4.1 Marcos de lectura abiertos (ORF) y proteínas virales.

El genoma de este virus tiene cuatro ORF en la cadena (-) del genoma del VHB altamente solapados; preS/Superficie (preS/S), Polimerasa (P), preCore/Core (PC/C) y X, cuya transcripción por la ARN polimerasa II del hospedador da lugar a cinco ARNm; preS1, preS2/S, pregenómico (ARNpg), preCore (ARNpc) y X, bajo el control de cuatro promotores virales (preS1p, Sp, Cp y Xp), que se traducen en siete proteínas virales; tres tipos de proteínas de superficie (SHBs, MHBs y LHBs), la polimerasa viral y los antígenos/proteínas e (HBeAg), core (HBcAg), y X (HBx) (Tabla 1)^{31,34-36}. Debido al tamaño relativamente pequeño del genoma del VHB (3,2 Kb) hay un alto grado de solapamiento entre los cuatro ORFs de forma que la mayoría de nt codifican 2 o más proteínas. Por ejemplo, en el caso de los nt que codifican la SHBs, también forman parte de las otras dos proteínas de superficie (MHBs y LHBs) y de la transcriptasa reversa (RT), por lo tanto, cuatro proteínas diferentes (figura 4). Además, el elevado grado de solapamiento del genoma del VHB no se limita a los ORFs ya que en la misma secuencia

de nucleótidos también contiene varias secuencias reguladoras y señales en cis, necesarias para la replicación viral.

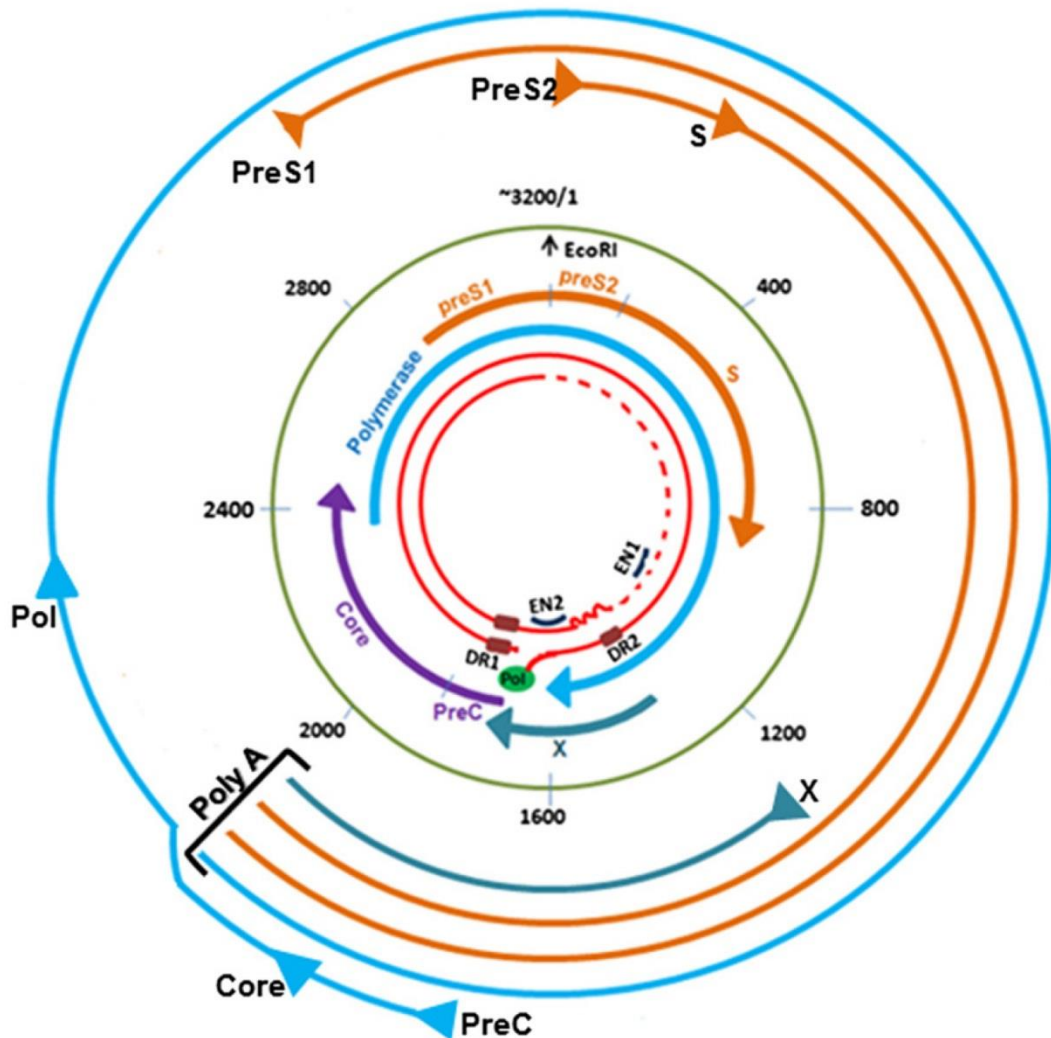


Figura 4 Diagrama esquemático del genoma y mapa de los transcritos del VHB. Los círculos rojos más internos representan el ADNrc. Comenzado por el interior en rojo se representa la hebra (+) con el ribonucleótido (línea roja ondulante) unido en su extremo 5'. La sigue la hebra (-) también representada en rojo y unida a la polimerasa en su extremo 5'. Las posiciones de DR1, DR2 y potenciadores (enhancer 1 (ENHI) y enhancer 2 (ENHII)) también están indicadas. La línea verde indica las posiciones del genoma viral en nucleótidos (aproximada). Las cuatro regiones de codificación de proteínas se muestran entre los círculos verde y rojo mediante flechas de colores semicirculares. Incluyen los genes precore y core (violeta), el gen de la polimerasa (azul), el gen X (verde-azulado) y los genes de la envuelta preS1, preS2 y S (naranja). Las líneas semicirculares más externas con puntas de flecha representan los 4 ARN (genómicos y subgenómicos) correspondientes a los ORF. Las puntas de flecha indican las posiciones de diferentes codones de iniciación dentro de cada ORF. Extraído de ³⁷.

ORF	Posición en el genoma del VHB*	Promotor	Posición	ARNm	Longitud (kb)	Proteína	PM (kDa)
preS/S	nt 2854-835	preS1p	nt 2712-2824	preS1	2,4	LHBs	39
		Sp	nt 3152-11	preS2/S	2,1	MHBs	33
						SHBs	24
P	nt 2357-1623	preCore/ pregenómico	nt 1613-1849	ARNpg	3,5	Polimerasa	90
preCore/Core	nt 1814-2458					HBeAg	21
				ARNpc	3,5	HBeAg	16
X	nt 1374-1838	X	nt 1250-1350	X	0,7 y 0,9	HBx	16,5

Tabla 1 ORFs, promotores, transcritos y proteínas del VHB. nt: nucleótido en el genoma del VHB. *Las posiciones del genoma hacen referencia a una secuencia de patrón genotipo D (número de acceso AJ344117.1). Estas posiciones pueden variar en genotipos diferentes. PM: Peso molecular

1.4.1.1 ORF PC/Core y las proteínas preCore y core:

El gen preCore/Core (nt 1814-2452) (hace referencia a la secuencia AJ344117.1) tiene dos codones de inicio de traducción (ATG) en el mismo ORF, el primero se encuentra en la posición 1814 del genoma del VHB y el segundo en la posición 1901, la secuencia entre ambos codones se denomina región preCore y entre las posiciones 1901 y 2452 región Core (figura 5). Este ORF es el que tiene un grado de solapamiento con los otros ORF del genoma del VHB más bajo que el resto, de modo que sólo los primeros nt de su extremo 3' están solapados con el extremo 3' del ORF X y los últimos del extremo 5' con el ORF P (figura 4)³¹.

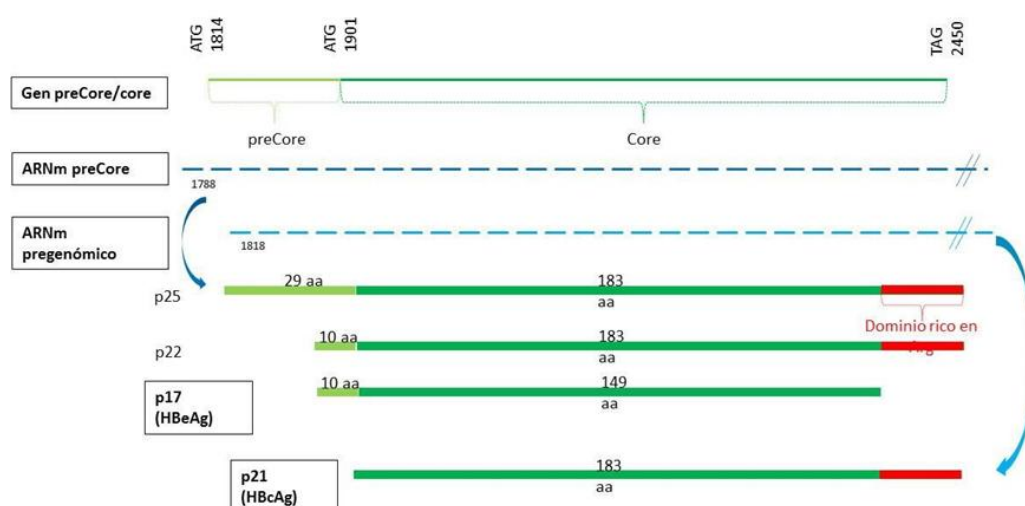


Figura 5 El ORF preCore/Core tiene dos codones de inicio por lo que sintetiza dos proteínas co-terminales, la proteína preCore/Core que es la precursora del HBeAg, usando el primer codón, y la proteína core desde el segundo. Las dos proteínas son sintetizadas desde diferentes transcritos el ARNpc y el ARNpg, respectivamente. Se indica el dominio rico en argininas³⁸.

CAPÍTULO 1. INTRODUCCIÓN

Proteína preCore (PC): Esta proteína es el precursor del antígeno e del VHB (HBeAg), HBeAg es secretado como una proteína soluble no estructural, y aunque no se requiere para el ensamblaje viral o la replicación, está conservado en todos los ortohepadnavirus y es importante para la infección natural *in vivo*^{39,40}.

Esta proteína se empieza a traducir por el codón de inicio que está el nt 1814 y termina en el nt 2452 (hace referencia a la secuencia AJ344117.1), con lo cual da lugar a la proteína PC o proteína p25, que contiene las regiones preCore y Core y tiene un peso molecular de 25 quilodaltons (KDa) y una longitud de entre 212 y 214 aminoácidos (aa) (figura 5).

La función principal del HBeAg es atenuar las respuestas inmunitarias innatas y específicas contra el HBcAg⁴¹⁻⁴³, por lo que es una molécula inmunomoduladora. En cuanto a la inmunidad adquirida, el HBeAg produce una respuesta principalmente humoral dependiente de células T que da lugar al anticuerpo anti-HBeAg ya que induce preferentemente una respuesta tipo Th2. También es capaz de inducir respuesta tipo Th1 (que puede provocar la activación de macrófagos) según las citoquinas del entorno⁴⁴. Sin embargo, el HBeAg disminuye la presentación antigénica *down*-regulando moléculas co-estimuladoras de las células de Kupffer y macrófagos, con lo cual bloquea la maduración de Th⁴⁵. Puede atravesar la placenta y establecer tolerancia en el útero Th específica para HBeAg y HBcAg, este mecanismo podría estar relacionada con de la alta tasa de cronicidad observada en recién nacidos de madres HBsAg(+) infectados perinatalmente⁴⁴.

El HBeAg también es capaz de inducir un estado de inmunosupresión de la respuesta innata que evita la eliminación inmunitaria de la infección⁴⁵. Para ello, puede interactuar con TLRs y moléculas de la cascada de señalización, con lo cual suprime la activación de factores inflamatorios, por ejemplo reduciendo la expresión de TLR2 en los hepatocitos, células de Kupffer y monocitos periféricos en pacientes con hepatitis B crónica (CHB)⁴⁵.

El HBeAg también es capaz de regular la expresión de diferentes genes de la célula relacionados con regulación de la transcripción, receptores de la superficie de la membrana celular, regulación del ciclo celular, biosíntesis de ácidos nucleicos y señalización y transporte intracelulares⁴⁵. Por lo tanto, en las CHB, la positividad de este antígeno es un predictor fuerte de desarrollo de hepatocarcinoma (HCC)⁴⁶.

Proteína Core(C): El HBcAg es el producto que tiene lugar al comenzar la transcripción de la ORF PC / Core por su 2º codón de inicio (posición 1901, figura 5). Este transcrito da lugar a una proteína de entre 183 y 185 aa y un peso molecular de 21 KDa, por eso el HBcAg también se le llama p21²¹.

CAPÍTULO 1. INTRODUCCIÓN

La estructura de los monómeros core (HBc) está dominada por una larga horquilla α -helicoidal (Figura 6A), la asociación de dos horquillas anfipáticas α -helicoidales resulta en la formación de un dímero, unido a través de puentes disulfuro intermoleculares entre cisteínas (Cys 61) (Figura 6B). Estos dímeros se pueden ensamblar en dos tipos de partículas, grandes y pequeñas, con un diámetro de 34 y 30 nm y una simetría con número de triangulación de empaquetamiento $T=4$ y $T=3$, conteniendo 240 y 180 moléculas de HBc respectivamente. Se forma entonces un haz de cuatro hélices, dos de cada monómero, con una protusión que forma una espículas de aproximadamente 25 Å de largo y 20 Å de ancho (Figura 6C y 6D). El extremo Ct de cada HBc, rico en argininas, queda orientado hacia el interior de la cápside e interacciona con las cargas negativas del ARNpg, lo cual es esencial para su encapsidación. En este dominio de 34 aa se encuentran los residuos serina que son los principales lugares de fosforilación del HBcAg, esenciales para la encapsidación y retrotranscripción del pgARN⁴⁷ así como para regular la localización nuclear de las nucleocápsidas a través de la activación de su señal de localización nuclear⁴⁸. Por otro lado, el extremo Nt (de entre 149 y 151 aa) queda en el exterior y posibilita la dimerización de los monómeros y el ensamblaje de la cápside. La secuencia entre los aa 76-81 localizado en la punta de las espículas constituye la mayor parte de la región inmunodominante principal MIR (*Major Immunodominant Region*) de la partícula core. La cápside está fenestrada con poros de diámetro 12-15 Å, que podrían ser esenciales para el acceso de los nt al interior durante la síntesis de ADN^{42,43}.

Aunque la función principal es formar las cápsides del VHB, la proteína HBc también puede interactuar con el ADNccc, formando parte de la estructura de los minicromosomas virales³⁰. Además, puede encontrarse tanto en el citoplasma como en el núcleo celular o en ambos, en función de la fase de la infección: se encuentra preferentemente en el citoplasma en la inflamación hepática durante la fase de actividad inmune, mientras que incrementa su concentración nuclear en la fase de tolerancia inmune⁴⁹. La dinámica estructural que permite a HBcAg la transición entre estos estados parece ser una clave importante para su función⁵⁰.

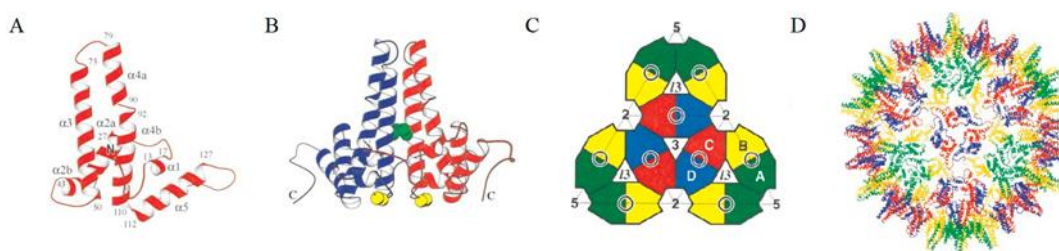


Figura 6. Representación de la estructura de la cápside y sus componentes. A: Plegamiento de los monómeros de la cápside. B: Formación de los dímeros. En la interfaz del dímero se forma un puente disulfuro entre las Cys61 (en verde). C: Representación esquemática de 4 subunidades independientes en cápsides T=4, desde una cara icosaédrica. La posición de un haz de 4 hélices aparece en círculos negros. Uno de los dímeros en azul-tojo, otro en verde-amarillo. D: Visión espacial del empaquetamiento de los dímeros en cápsides T=4)¹⁵.

1.4.1.2 ORF P y la polimerasa:

La ORF P es el más grande del genoma del VHB, ocupa el 80% de su secuencia, tiene una capacidad de codificación de alrededor de 830 aa y se solapa con la totalidad del ORF S, con una porción del extremo 3' del pC/C y con la porción 5' del X (figura 4). Este ORF codifica una única poliproteína denominada polimerasa viral de 90 KDa, proteína multifuncional implicada en la encapsidación, retrotranscripción y posterior degradación del ARNpg y en la síntesis de ADN (la hebra (+) del genoma viral). Estas actividades que es capaz de realizar la polimerasa del VHB evidencian la similitud del mecanismo de replicación del VHB con el de los retrovirus.

Estas funciones las realiza gracias a sus cuatro dominios repartidos a lo largo de su secuencia: proteína terminal (PT, nt 2307 hasta 2862), espaciador (SP, nt 2863-129), transcriptasa reversa (RT, nt 130-1161) y ARNasa H (RH, nt 1162 a 1623) (hace referencia a la secuencia AJ344117.1) (figura 7)⁵¹. No se dispone de datos estructurales directos de estos dominios porque todavía no se ha podido cristalizar esta proteína debido a su alta insolubilidad¹⁸, no obstante mediante análisis genéticos y bioinformáticos se ha demostrado que el dominio RT de Hepadnavirus es homólogo a los correspondientes dominios RT de retrovirus.

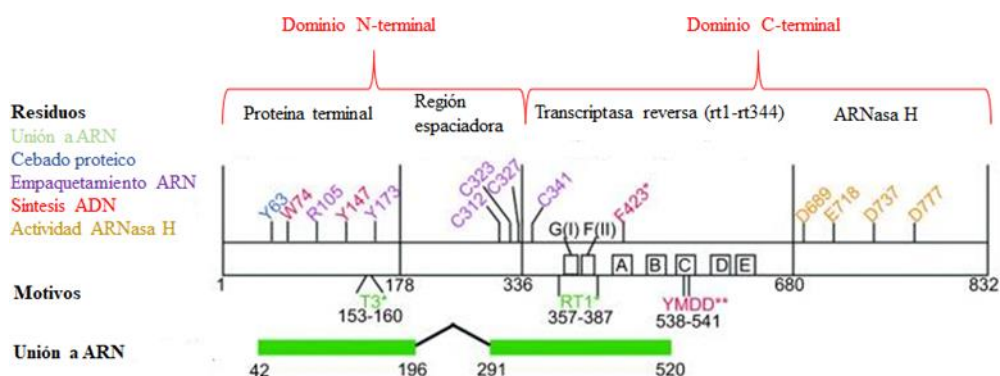


Figura 7 El ORF P se representa esquemáticamente con sus dominios, motivos importantes y residuos críticos indicados. Los motivos y los residuos están codificados por colores para indicar los pasos conocidos de la replicación del VHB para los que son importantes. Las pequeñas cajas A-G denotan las regiones conservadas a través de transcriptasas inversas. Las regiones requeridas para la unión de ARN se representan como cuadros verdes. *función verificada en DHBV, pero no en HBV. ** sitio activo YMDD es necesario para el cebado de proteínas y la síntesis de ADN Imagen modificada de⁵².

Dominio N-terminal: Esta región de unos 349 aa contiene los dominios TP y SP. Al no disponer de datos estructurales directos de la proteína P no se sabe prácticamente nada sobre la estructura tridimensional de estos dos dominios, porque no comparten homología con otras proteínas de estructura conocida. Sin embargo, se han podido identificar motivos y residuos críticos para varias de las actividades de la polimerasa mediante estudios genéticos y bioquímicos⁵³.

El dominio proteína terminal: es esencial para la replicación del VHB ya que es imprescindible para la unión de la polimerasa viral con el ARNpg. Esta unión tiene lugar a través de la interacción con una estructura secundaria que se forma en el extremo 5' del ARNpg con la secuencia de encapsidación (ϵ), la cual está localizada en la secuencia de la región PC (entre los nt 1851 y 1904)^{54,55}, con la colaboración de chaperonas como las *heat-shock proteins* Hsp90, HSC70 y sus respectivas co-chaperonas Hop y Hsp40^{17,18}.

Región espaciadora: A continuación del dominio PT se encuentra una región interna que es el dominio SP, que parece tener como única función unir la PT con los dominios del extremo C-terminal mediante una conexión flexible como una bisagra²¹(figura 7). Este dominio contiene tres cisteínas en el extremo C-terminal, que junto con otra cisteína en el N-terminal del dominio RT son necesarias para la encapsidación del ARNpg⁵⁶ (figura 7)

Dominios de la región C-terminal: En esta región de unos 500 aa se encuentran los dos dominios catalíticos de la polimerasa, RT y RH (figura 7).

1. **Dominio RT:** actúa como transcriptasa reversa (utiliza el ARNpg como molde para la síntesis de la hebra (-) del ADNrc) y ADN polimerasa (utiliza la hebra (-) como molde para la síntesis de la hebra (+)).

Gracias a los datos estructurales directos obtenidos de la transcriptasa reversa del Virus de la Inmunodeficiencia humana-1 (VIH-1) se ha podido generar modelos estructurales plausibles del dominio RT del VHB⁵⁷⁻⁵⁹. Según estos modelos, el dominio RT presenta una configuración similar a una mano derecha, con una palma (*palm*), unos dedos (*finger*) y un pulgar (*thumb*)^{52,60} (figura 8). La posición de los subdominios *finger* y *thumb* crea una ranura en el subdominio *palm* a través del cual pasa la cadena 5' en extensión, este subdominio además contiene el motivo catalítico YMDD, común a todas las RT y que es el sitio activo que cataliza la adición de nt al ADN en el proceso de síntesis. Los nt se unen adyacentes al extremo 3' de la hebra naciente. Los subdominios *finger* facilitan la interacción con los nt entrantes y el molde al que están unidos. El subdominio *thumb* podría tener una función en el posicionamiento, procesado y translocación del dúplex de ADN⁶¹. Existe una nomenclatura consenso de los aa del dominio, el inicio está en el motivo muy conservado EDWGPCDEHG y contiene 344 aa en total, así el dominio catalítico está en las posiciones rt203-rt206 (Y203M204D205D206)⁶¹. Actualmente este dominio de la polimerasa viral es la única diana del VHB que se ha logrado aprovechar con éxito para la terapia antiviral mediante antivirales orales análogos de nucleós(t)idos (ANs). De hecho varios inhibidores de la transcriptasa reversa del VIH-1 también tienen una potente actividad contra el VHB y las mutaciones que confieren resistencia a estos tratamientos tienen lugar en posiciones equivalentes según las predicciones estructurales⁵⁹.

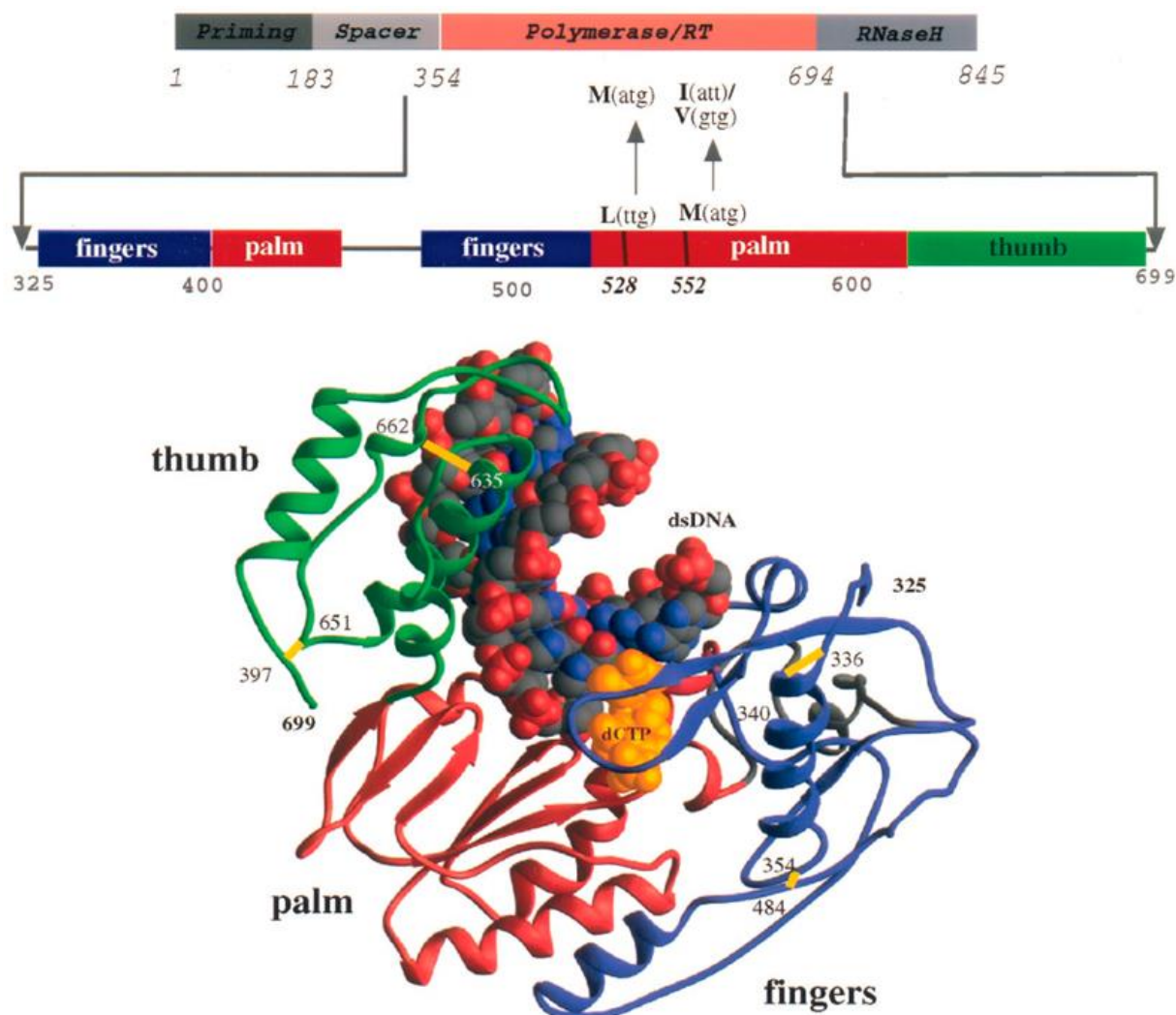


Figura 8 Representación esquemática de la transcriptasa inversa del VHB, basándose en la homología a la RT del VIH-1. La polimerasa del VHB se muestra como un diagrama de cinta con los dedos (325 a 403 y 469 a 519), la palma (404 a 440 y 520 a 613), y el pulgar (614 a 699) subdominios en azul, rojo y verde, respectivamente. El límite del cebador del ADN bicatenario se muestra en gris (con átomos de N y O en azul y rojo, respectivamente), y dCTP en dorado. Los cuatro enlaces disulfuro propuestos están representados por líneas amarillas⁵⁹.

2. Dominio ARNasa H: Este dominio actúa como una ARNasa H, unas endonucleasas no específicas de secuencia que degradan las cadenas de ARN en híbridos ARN/ADN. En el caso del VHB, el dominio RH degrada el ARN_pg durante la síntesis de la hebra (-) del ADN_rc⁶². Se ha observado en diferentes miembros de la familia de las ARNasa H que la degradación de la cadena de ARN se catalizada por dos cationes divalentes, coordinados gracias a una región de unos 100 aa de su extremo N-terminal que contiene los aa DEDD⁶³ (figura 7). Estos aa cargados negativamente también se encuentran en el dominio RH de la polimerasa del VHB y son esenciales para la

CAPÍTULO 1. INTRODUCCIÓN

replicación viral^{62,64}. Además de la actividad ARNasa H, el dominio RH también colabora en la encapsidación del complejo ribonucleoproteico formado por ARNpg, chaperonas y la propia polimerasa viral⁶⁵. A diferencia del dominio RT la homología de secuencia del dominio RH con ARNasa H de transcriptasas reversas de retrovirus y otros retroelementos es baja, pero sigue siendo significativa^{5,66,67}.

1.4.1.3 ORF S y las proteínas de superficie:

El ORF S está totalmente solapado al P, comparte la secuencia de antes del dominio SP y una gran parte del RT (figura 4). El marco de lectura preS/S codifica tres proteínas de superficie diferentes, aunque estructuralmente relacionadas porque son sintetizadas desde tres codones de inicio alternativos en la misma pauta de lectura; LHBs desde la posición nt 2848, MHBs desde la posición nt 3205 y SHBs desde la posición nt 155 (hace referencia a la secuencia AJ344117.1). Así pues, las tres comparten el mismo extremo C-terminal, pero tienen diferente longitud en el extremo N-terminal. Estas proteínas son traducidas desde dos ARNm subgenómicos, el más largo (ARNm preS1) codifica la LHBs, mientras que el ARNm preS2/S codifica las proteínas MHBs y SHBs desde codones de inicio separados (figura 4). La síntesis de MHBs y SHBs está regulada por un potente promotor, no específico del hígado, Sp, mientras que la transcripción de la LHBs está controlada por un promotor más débil, hepato-específico, preS1p⁶⁸. Todas tienen la región principal hidrofílica MHR (*Major hydrophilic region*) entre los aa 99-169 de la secuencia de la SHBs, donde se localiza el determinante “a” (aa 124-147), principal dominio antigénico diana para la respuesta neutralizante de anticuerpos de células B tras una infección o la vacunación⁵³. Estas proteínas se encuentran insertadas en la envuelta viral (son proteínas integrales de membrana) y en conjunto forman el antígeno de superficie del VHB o HBsAg, el principal marcador diagnóstico de la infección por el VHB.

Proteína SHBs: Esta proteína consta de 226 aa y un peso molecular entre 24 y 27 KDa. Estos aa son los que tienen en común en el extremo C-terminal todas las proteínas de superficie, y es el principal componente de la envuelta del VHB (figura 9). Su secuencia es rica en aa hidrofóbicos, lo que le permite crear dominios transmembrana a través de los que las proteínas de superficie se insertan en la envuelta de las partículas virales y subvirales⁶⁹.

CAPÍTULO 1. INTRODUCCIÓN

La inserción de la proteína S a la membrana se inicia por una secuencia señal en la región N-terminal (señal I) que no se escinde y que constituye el primer dominio transmembrana (TM1, aa 8-22), lo que hace que los aa 1 a 7 se trasloquen al lumen del RE donde se sintetizan estas proteínas⁶⁹. una segunda señal (señal II), que forma el segundo dominio transmembrana (TM2, aa 80-90), dirige la translocación de la cadena creciente a través de la membrana del RE hacia el lumen, hasta que comienza, alrededor del aa 170, el extremo C-terminal hidrofóbico del dominio S, localizado dentro de la membrana del RE. La estructura detallada de la región C-terminal (aa 170-226) no se entiende completamente y se considera como una o dos regiones transmembrana (TM3 / TM4)^{20,70}(figura 10). Por lo tanto, en una visión global de la proteína S completa, el extremo N-terminal (aa 1-7) y el bucle entre aa 99 y aa 169 (que da lugar al MHR) se enfrenta al lumen del RE, y el dominio (aa 23-79) se sitúa hacia el citoplasma. La orientación luminal del bucle entre los aa 99-169 permite la N-glicosilación de la aspargina (Asn) en la posición 146 (dentro del determinante a) por las N-glicosiltransferasas. Esta glicosilación se da solo de manera parcial (de ahí la variabilidad en el peso molecular de esta proteína) lo que da lugar a al menos una forma glicosilada y otra sin glicosilar de todas la proteínas de superficie¹⁴. Estos dominios luminales están expuestos a la superficie externa de la partícula viral (figura10)^{20,71}.

Proteína MHBs: Esta proteína se traduce a partir de las secuencias preS2 y S, por lo tanto, contiene la secuencia de la SHBs con la misma topología y 55 aa extras, que dan lugar a una proteína de 281 aa (figura 9). Los 55 aa del preS2 son traslocados al lumen del RE por la señal I del TM-1⁷², dando lugar al acceso de N-glicosiltransferasas para que se produzca la N-glicosidación sobre la Asn-4. Teniendo en cuenta la glicosilación parcial de la Asn-146 de la secuencia de la SHBs, esto conduce a las tres posibles formas de MHBs: la no glicosilada (p30), la monoglicosilada (gp33) y la biglicosilada (gp36)²⁰. Además, en algunos casos las MHBs y en menor medida las LHBs pueden tener una O-glicosilación en el residuo treonina (Thr) 37 (T37) de la región preS2^{50,73} (figura 9).

Proteína LHBs: Esta proteína tiene entre 389 y 400 aa porque contiene la secuencia de todo el ORF S, es decir, los 226 aa codificados por la región S, los 55 aa por la preS2 y los 108-119 aa (según el genotipo) para la preS1 (figura 9). La LHBs muestra una biosíntesis inusual, el dominio PreS1-PreS2 y una parte del dominio S, hasta el aa 79, no se traslocan co-traduccionalmente por lo que permanecen en la cara citosólica del RE. TM2 determina la inserción co-traduccional de esta proteína, por lo tanto el dominio TM1 no se inserta en la membrana del RE, y dianas potenciales de N-glicosilación como

la Asn-4 del preS2 y Asn-15 de preS1 no se glicosila porque quedan inaccesibles a las N-glicosiltransferasas del lumen del RE²⁰ (figura 9). Así pues, la LHBs puede tener una forma no glicosilada de 39 KDa (p39) y otra de monoglicosilada (en el residuo Asn-146 de la secuencia de la SHBs) de 42 KDa (gp42)⁵⁰. Otra modificación co-traducciona es que el residuo de glicina en el aa 2 del dominio PreS1 está miristilado (se le une una molécula de ácido místico covalentemente)^{50,74,75}. La miristilación de las LHBs es esencial para la infectividad de las partículas virales y es una señal de retención en RE (figuras 9 y 10)^{20,76,77}.

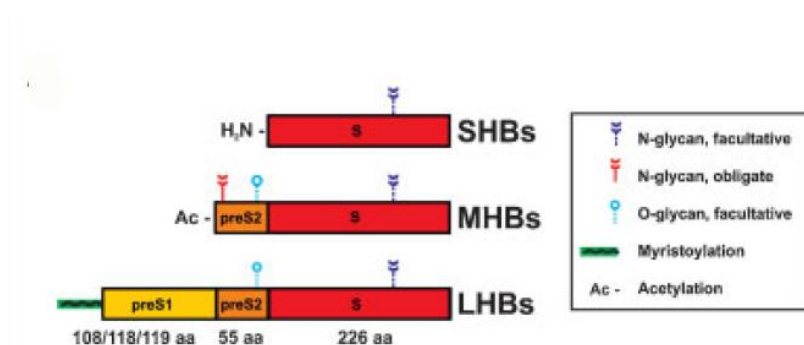


Figura 9 Esquema de los dominios (S,preS2 y preS1) que forman parte de las HBs del VHB⁷⁸.

Después de la traducción, aproximadamente el 50% de las LHBs sufren una reorientación topológica post-traducciona por la que su extremo N-terminal se transloca al lumen del RE⁷⁹. En estas proteínas LHBs el dominio TM1 es integrado en la membrana del RE y el dominio preS1-preS2 mira al lumen del RE, y por lo tanto hacia el exterior de las partículas virales (figura 10). Dada la orientación de su dominio preS, esta forma de la LHBs es responsable del reconocimiento de la célula huésped y de la infectividad del virus, a través de su receptor específico, el péptido co-transportador de sodio-taurocolato (NTCP)⁸⁰. Cuando el dominio preS1-preS2 permanece en el citoplasma, y por lo tanto mira hacia el interior de las partículas virales, LHBs juega un papel esencial en la morfogénesis del virus, interaccionando con la nucleocápside, y está asociado con la función adicional de preS2 como proteína reguladora. En esta conformación citosólica, el dominio preS2 se une y activa la protein kinasa C, resultando en la activación de la cascada de señalización c-Raf/MEK que controla la expresión de una variedad de promotores celulares y virales²⁰

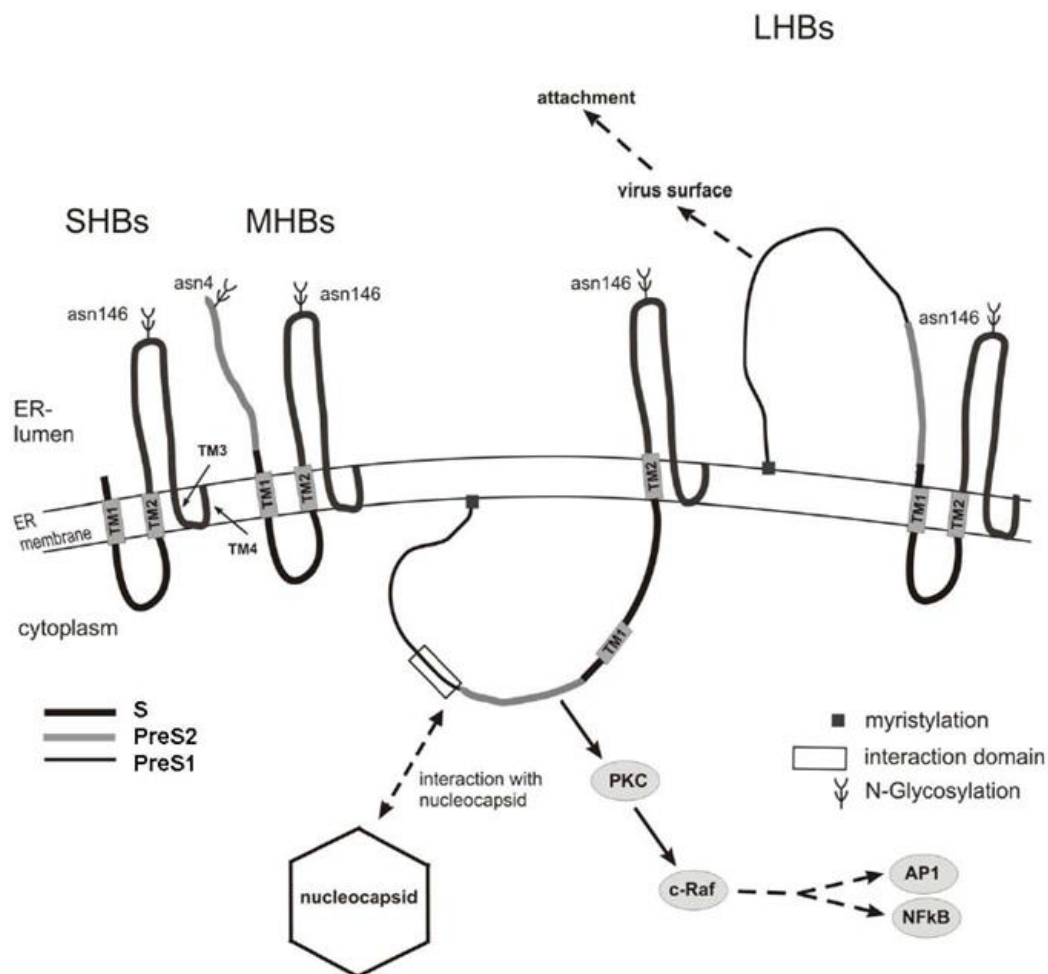


Figura 10. Esquema de la estructura de las proteínas de superficie del VHB. Se indican los residuos que pueden ser glicosilados en cada una de ellas²⁰.

1.4.1.4 ORF X y la proteína X:

El ORF X es el más pequeño, con una longitud de 495 pb (nt 1374-1838) y se solapa en buena parte con el dominio RH de la ORF P, con el promotor y con el extremo 5' del ORF PC / C (figura 4). El ORF X codifica para la HBx que es un polipéptido de entre 145 a 154 aa (según el genotipo del VHB) y un peso molecular de unos 17 KDa. La HBx es una proteína reguladora codificada por el VHB y está conservada a lo largo de todos los Hepadnavirus de mamíferos, pero ausente en los aviáres, aunque en algunos de ellos expresan proteínas que engloban la secuencia del HBcAg y de la HBx⁸¹. El hecho de que los avihepadnavirus no den lugar a HCC pero sí los orthohepadnavirus sugirió para esta proteína un papel oncogénico en la infección por el VHB que se ha ido confirmando en estudios posteriores⁸².

La secuencia de nucleótidos del ORF X también contiene la mayor superposición entre la región de codificación y los elementos de regulación⁸³, incluyendo el Cp, con las cuatro *TATA boxes-like*, el ENHII, el elemento regulador negativo (NRE, nt 1611-1634) y las DR1 y DR2. Además contiene una región de unión a miARN (nt 1362-1383), un motivo tipo BH3 (nt 1719- 1769) de interacción con proteínas anti-apoptóticas Bcl-2 y Bcl-XL y regiones epitópicas para células B (nt 1458-1517) y células T (nt 1719-1754)^{49,83-86} (figura 11).

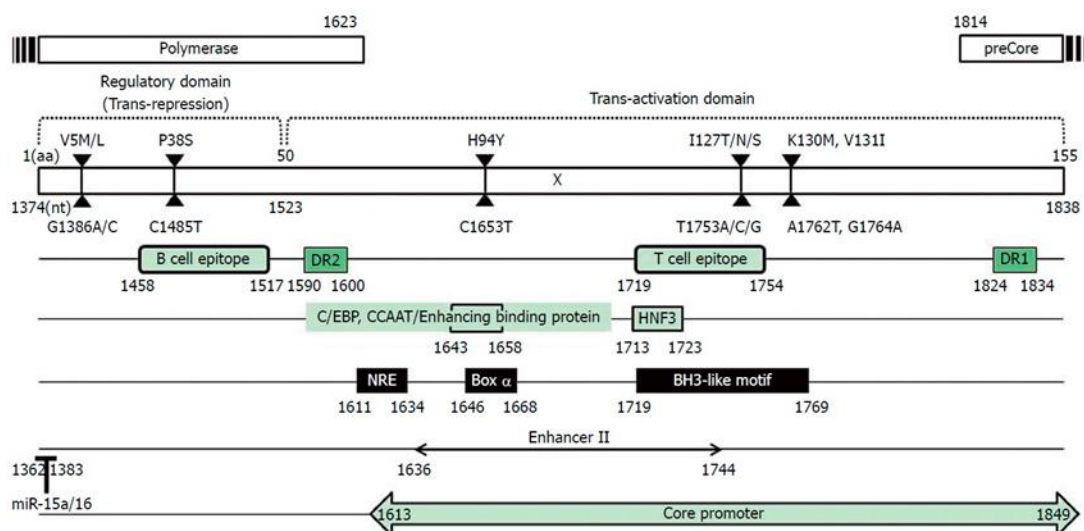


Figura 11 Estructura de la región codificante de la HBx del VHB. La región X contiene 495 pb (nt 1374 a 1838) que codifica la HBx de 17 kDa compuesta por 154 aa en este esquema, que también contiene múltiples elementos reguladores⁸⁶.

1.4.1.4.1 Estructura

La HBx no ha sido cristalizada y sus características bioquímicas no han sido completamente esclarecidas dadas las dificultades de producir cantidades significativas de la HBx soluble⁸⁷, con la excepción del dominio de interacción a DDB1 (*DNA damage-binding protein 1*) que si ha podido ser co-cristalizado⁸⁸, con lo cual, la estructura ha sido caracterizada por ensayos menos precisos como difracción circular, fluorescencia, resonancia magnética nuclear o predicción *in silico* con aproximaciones bioinformáticas (figura 12).

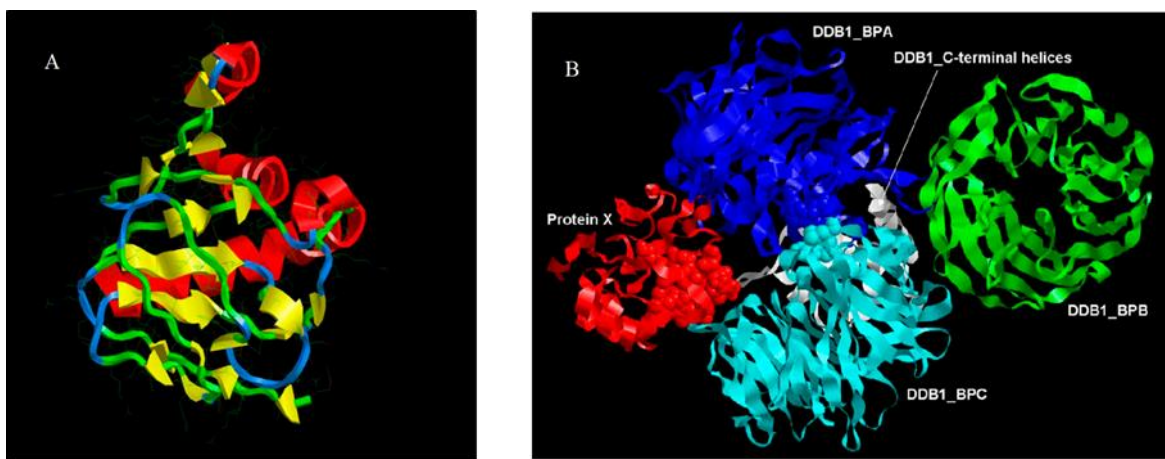


Figura 12: Modelo de la estructura 3D de la HBx. A: Modelo de la secuencia consenso del genotipo D. Rojo, amarillo, azul y verde para las hélices, láminas, giros y otros, respectivamente. B: Modelo de interacción HBx-DDB1. La HBx (aa 88-100) interactúa con DDB1 en un bolsillo creado por dos dominios hélices- β en DDB1 (BPC y BPA) con el extremo C-terminal de la HBx saliendo del complejo. La HBx en rojo, los dominios A, B y C de DDB1 en azul turquesa, verde y azul oscuro respectivamente y el dominio α -helicoidal C-terminal de DDB1 en blanco⁸⁸.

La secuencia de la HBx abarca 154 aa y está organizada en dos dominios funcionales⁸⁹. El N-terminal es un dominio regulador negativo, mientras que el C-terminal actúa como un dominio de transactivación. El dominio regulador negativo abarca los primeros 50 aa, incluyendo una región rica en Ser y Pro (residuos 21-50) una región que se cree que es esencial para la dimerización de la HBx⁸⁹. El dominio de transactivación (residuos 53 - 142), posee tres regiones altamente conservadas a lo largo de los Hepadnavirus (aa 58-84, 61-69 y 132-138), que son esenciales para la transactivación. Dos de esas secuencias (aa 61-69 y 132-138) son homólogas al dominio Kunitz, los dominios Kunitz son relativamente pequeños, y son los dominios activos de las proteínas que inhiben la función de las enzimas degradantes de proteínas, más específicamente, son inhibidores de la proteasa⁹⁰. El extremo C-terminal (aa 143-154) no parece ser esencial para la transactivación, ya que las HBx que pierden una pequeña región C-terminal mantienen esta actividad⁹¹⁻⁹³. Los principales dominios de esta proteína y sus aminoácidos (aa) más conservados se indican en la figura 13.

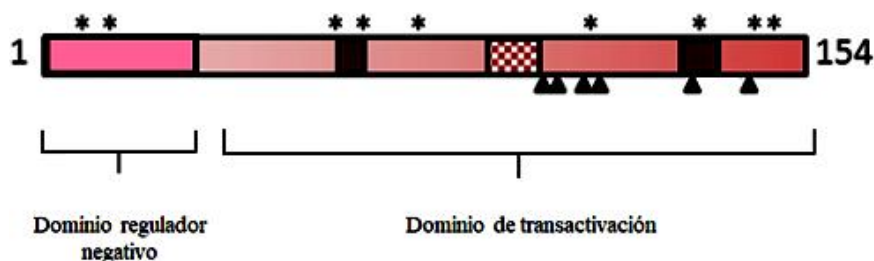


Figura 13 Representación esquemática de la HBx. Los dominios funcionales que se muestran incluyen el dominio regulador negativo (aa 1-50) y el dominio de transactivación (aa 50-154). Dentro del dominio de transactivación, se señala el de unión a DDB1 (aa 88-100; área marcada con pequeños cuadros), que se encuentra a su vez dentro del dominio de unión del proteasoma (aa 90-154), los dos dominios tipo Kunitz representados en negro, los residuos de cisteína marcados con asteriscos (aa 7, 17, 61, 69, 78, 115, 137, 143 y 148) y con triángulos se indican los seis residuos de lisina (aa 91, 95, 113, 118, 130, 140)⁹⁴.

1.4.1.4.2 Localización

La detección y localización de la HBx es todavía controvertida; algunos autores describen una localización mayoritariamente citoplasmática, aunque con una ligera detección nuclear⁹⁵, mientras que otros refieren una expresión principalmente nuclear⁹⁶. Finalmente, se ha propuesto un modelo donde la HBx se localiza mayoritariamente en el núcleo cuando sus niveles de expresión son bajos, pero se acumula en el citoplasma al incrementar su síntesis^{97,98}. Además, una sobreexpresión de la HBx conduce a una distribución mitocondrial anormal⁸⁷. Esta dinámica de localización parece que puede ser clave en las distintas funciones que la HBx es capaz de desarrollar a lo largo del ciclo vital del VHB.

1.4.1.4.3 Funciones

La HBx se caracteriza por ser capaz de modular diversas funciones celulares y víricas, ejerciendo una amplia variedad de funciones, que hacen que tenga un papel esencial en la biología y patología del VHB (figura 14)⁸⁷:

1. **Función transactivadora de genes celulares:** La HBx desempeña un papel decisivo en la respuesta inmunitaria del huésped y en el desarrollo del HCC mediante una de sus funciones mejor caracterizadas, su capacidad transactivadora. Así, la HBx es capaz de modificar la expresión de genes celulares implicados en carcinogénesis e inflamación, como c-fos, c-jun, c-myc, EGF-R, TGF- β , IL-6 o HLA-DR, entre otros, puede modular la actividad de la ARN polimerasa y asociarse con elementos de la maquinaria de transcripción como TFIIB, TFIIF y RBP^{94,99}, e interacciona con otros factores de transcripción como CREB, ATF-2, ATF-3, NF-AT, C/EBP β , SMAD4, SREBP1 y AP-2,

aumentando su capacidad de unión al ADN, o incrementando su actividad transactivadora^{94,99}.

2. Función en la replicación del VHB: La HBx tiene un papel fundamental en mantener la replicación del virus a través de varios mecanismos⁸⁶. Existen experimentos que demuestran que el aumento de la transcripción y replicación llevada a cabo por la HBx es por una vía dependiente del proteosoma y que DDB1 está implicado, de manera dependiente de la HBx, en la transcripción y replicación del VHB. DDB1 tiene un papel importante en diversos procesos celulares como la síntesis de ADN, la expresión génica, la división celular y la apoptosis^{106,112}. La HBx (aa 88-100) interacciona con DDB1 en la misma región a la que se unen los DCAFs (*DDB1 Cullin Accesory Factor*), y a este complejo se une a diferentes factores de transcripción (FT) necesarios para la replicación viral. Recientemente se ha descrito la interacción de la HBx con el complejo DDB1-CUL4-ROC1 (CRL4) que contiene la ubiquitina ligasa E3, para que marque para su degradación el complejo Smc5 / 6 (*Structural maintenance of chromosomes 5 and 6*). Dicho complejo inhibe la transcripción de elementos extracromosómicos como el ADNccc, impidiendo su expresión génica, por lo tanto su degradación es una de las funciones primarias de la HBx^{101,102}. Sin embargo, aún no se conoce el mecanismo por el cual Smc5 / 6 suprime la transcripción del VHB ni cómo se expresa inicialmente el HBx¹⁰³. La interacción de la HBx-DDB1 está conservada a lo largo de todos los hepadnavirus de mamíferos lo que refuerza el importante papel de esta interacción en la replicación viral^{93,100}.

Además se sabe que durante la replicación viral la HBx es reclutada al minicromosoma de ADNccc con una cinética paralela a la de la propia replicación¹⁰⁴. Además de la degradación del complejo Smc 5 / 6, la HBx también regula la replicación del genoma viral a través del control epigenético de la transcripción del ADNccc mediante el estado de acetilación de las histonas H3 y H4 unidas al minicromosoma de ADNccc¹⁰⁵. En ausencia de la HBx estas histonas son desacetiladas rápidamente por desacetilasas de histonas y disminuye significativamente la transcripción del ARNpg¹⁰⁴.

3. Función de transcripción de genes virales: También tienen un papel destacado en la replicación viral de los factores de transcripción la familia *Cyclic-AMP-response element binding protein* (CREB), que actúan sobre el metabolismo y la proliferación celular¹⁰⁶. La HBx estabiliza la unión de CREB con el minicromosoma de ADNccc al EN1, que regula la propia expresión del ORF X, y también a preS2¹⁰⁶, por lo tanto, esto puede estar relacionado con la activación de la transcripción del ORF X. Además, la HBx

también interacciona con el complejo *CREB-binding protein* (CBP) / p300, para potenciar la acción de CREB¹⁰⁷, y con *Protein Phosphatase 1* (PP1), para inhibir la desfosforilación de CREB y evitar la supresión de su actividad transcripcional. También actúa sobre otros factores de transcripción celular que estabiliza la unión al minicromosoma de ADNccc como el *CCAAT / enhancer-binding protein- α* (C / EBP α), que potencia la transcripción del ARNpg. Por otra parte, la HBx bloquea la *protein arginin N-methyltransferase 1* (PRMT1) que también interacciona con el genoma del VHB y tiene una actividad inhibidora de su replicación.

4. Función de apoptosis: La HBx tiene funciones anti-apoptóticas a través de la inhibición de la actividad transcripcional de proteínas supresoras de tumores como p53 y vías apoptóticas dependientes e independientes de caspasas. Paradójicamente la HBx también tiene funciones pro-apoptóticas como, por ejemplo, activando la muerte celular a través del aumento de los niveles citoplasmáticos de calcio, alterando las membranas mitocondriales a través de la unión con la chaperona Hsp60 y *human voltage-dependent ion channel 3* o aumentando la susceptibilidad de las células a la apoptosis mediada por *tumor necrosis factor α* (TNF α)^{108,109}. Por otra parte, la HBx sería capaz de modular las vías PKB/Akt, ERK, SAP/JNK y p38 en el citoplasma, según se ha descrito en diversas líneas celulares. Asimismo, a través de la familia de receptores no tirosina-cinasa-Src, la HBx modula la señalización de las vías Ras-Raf-MAPK, JNK, JAK/STAT, la acción de factores de transcripción AP-1 y NF- κ B, y la apoptosis mediada por *Transforming growth factor β* (TGF- β)^{110,111}.

5. Función en el ciclo celular: El mecanismo por el cual la HBx regula la activación de la quinasa- Src, la cual activa la polimerasa del VHB estimulando la replicación viral a nivel de la transcripción reversa, consiste en la despolarización de la membrana mitocondrial y la consiguiente movilización de calcio intracelular. La HBx desencadena la liberación de Ca²⁺ desde la membrana mitocondrial al citoplasma al interactuar con un componente del poro mitocondrial. Los niveles elevados de Ca²⁺ citoplasmático activan Pyk2, que une y activa las quinasas Src^{87,112}. Además, durante la mitosis, la HBx puede causar anomalías cromosómicas, a través de la unión con *Hepatitis B X-interacting protein*, o de *mitotic checkpoint serine / threonine-protein kinase BUB1 beta*, un componente del complejo del punto de control mitótico. Estudios reciente han demostrado que durante la división celular, el minicromosoma podría estar distribuido asimétricamente o incluso perderse, lo que conduce a una dilución de ADNccc¹³.

6. Modulación de vías celulares de señalización: La HBx tiene la capacidad de modular vías celulares de señalización para producir cambios en el metabolismo de la célula huésped, como por ejemplo en el metabolismo de lípidos y la gluconeogénesis. La HBx altera el metabolismo lipídico interactuando con PPAR- γ (*peroxisome proliferator-activated receptor- γ*) y SREBP1 (*sterol regulatory element binding protein 1*), que puede llevar a una acumulación lipídica en los hepatocitos, y aumenta la β -oxidación bajo condiciones de glucosa limitantes. También se ha relacionado con un metabolismo alterado de la glucosa, aminoácidos y ácidos nucleicos^{113,114}. La HBx también puede activar la hepatocarcinogénesis a través de la activación de los factores de transcripción NF- κ B y AP-1, que activan la proliferación celular, al igual que ocurría con proteínas de superficie con truncamientos a preS2. Por otra parte, la HBx también induce de la angiogénesis. Finalmente, la HBx puede actuar como una proteína inmunomoduladora actuando sobre la vía del (TGF- β), implicada en procesos inflamatorios. La sobreexpresión de TGF- β puede dar lugar a metástasis. Esta proteína también puede regular los procesos inflamatorios a través de la interacción con la fosfatasa de treonina / serina PP2C α y la respuesta inmunitaria uniéndose directamente al *interferon β promoter Stimulator-1 factor* (IPS-1) e inhibiendo la expresión del interferón- β (INF- β) inducida por ADN de doble cadena (dsADN) como el del VHB¹¹⁵.

7. Inhibición de la reparación de lesiones en el ADN y vías de degradación proteica: El marcaje de proteínas celulares para su degradación por el proteosoma es crítico para el funcionamiento celular normal. La HBx interacciona por su parte C-terminal con las subunidades del proteosoma PSMA7, una subunidad α del proteosoma 20S y PSMC1, una subunidad ATPasa-like en el componente regulador 19S del proteosoma 26S, inhibiendo así la actividad proteolítica del proteosoma, lo que sugiere que la función de la HBx como co-activador de muchos factores transcripcionales pueden ser el resultado indirecto de las alteraciones en la actividad proteolítica¹¹⁶. La HBx también puede incidir en la ubiquitinización a través de su interacción directa con DDB1, uno de sus principales ligandos, que reprograma la actividad ubiquitin-ligasa del complejo *Cullin4A-DDB1 ubiquitin E3 ligasa*. Inicialmente se describió DDB1 como una subunidad del complejo UVDDDB, que tiene la función de unirse al ADN dañado e iniciar el mecanismo de reparación por escisión, que es inhibido la unirse la HBx a DDB1¹¹⁷. HBx también puede regular la respuesta a ADN dañado y la reparación del ADN porque se une directamente a la helicasa XBP (Xeroderma pigmentosa B) o XPD (Xeroderma pigmentosa D) en el FT general IIIH (TFHII)⁹².

8. Influencia de la HBx sobre la respuesta inmune: Aunque el VHB es no citopático, la cronicidad de la infección provoca un daño hepático derivado de la respuesta inmunitaria del paciente. Esto sucede cuando el sistema inmunitario del individuo infectado no es capaz de neutralizar la infección y acaba desarrollando hepatitis crónica. El VHB ha desarrollado diferentes estrategias para evadir el sistema inmunitario. Sin embargo, es paradójico que la HBx induzca factores proinflamatorios como el factor de necrosis TNF- α , la sintetasa inducible del óxido nítrico (iNOS), las interleucinas (IL)-6 o IL-8, en lugar de citocinas antiinflamatorias como la IL-10^{118,119}. El TNF- α es una citocina clave en muchos procesos inflamatorios que desempeña una amplia variedad de funciones tanto en la inmunidad innata como en la adaptativa. En los hepatocitos, el TNF- α inicia diferentes cascadas de señalización capaces de inducir tanto la proliferación como la muerte celular por apoptosis. De esta manera, la HBx altera el equilibrio entre las señales de proliferación y apoptosis, impidiendo la muerte de la célula y favoreciendo su división al inducir la expresión de iNOS en los hepatocitos. Al igual que TNF- α , iNOS podría inhibir la replicación vírica, pero favorecería, junto a la IL-6, la proliferación de los hepatocitos inducida por TNF- α ^{86,120,121}. Estudios recientes han descrito que la HBx disminuye la respuesta innata al favorecer la degradación de la proteína de señalización antivírica mitocondrial (MAVS). Esta proteína es crítica en la respuesta inmunitaria frente a infecciones víricas ya que es capaz de activar factores de transcripción como NF- κ B y el factor regulador del INF tipo III, estimulando la producción celular del INF tipo I. Consecuentemente, la degradación de MAVS provocaría un déficit de la síntesis de INF por parte de las células de la respuesta innata, evitando la muerte de los hepatocitos infectados, permitiendo la supervivencia del virus y su escape del sistema inmunitario¹²².

9. Efectos profibrogénicos: Aún no se conoce mucho la implicación de la HBx en la fibrosis hepática. La activación de las células estrelladas hepáticas, (HSC) es esencial para el desarrollo de fibrosis y cirrosis. Dado que no existen evidencias de la infección de las HSC por el VHB¹²³, se ha sugerido que el efecto profibrogénico de la HBx reside en su influencia sobre la síntesis y secreción de factores profibrogénicos por los hepatocitos infectados. Así, la HBx induce la activación y proliferación de las HSC por la vía del factor de crecimiento transformante β (TGF- β)¹²⁴. Hay estudios *in vitro* que demuestran que la HBx estimula la secreción de colágeno I y la metaloproteinasa de matriz extracelular-2 (MMP-2) en las HSC¹²⁴ a través de la síntesis de factores profibrogénicos, comprometiendo el equilibrio entre los fenómenos de síntesis y degradación de la matriz extracelular. De este modo, la HBx potenciaría el avance de la lesión hepática al favorecer la activación de las HSC y, por tanto, la fibrosis y cirrosis.

10. Angiogénesis y HCC: La angiogénesis patológica característica de numerosas enfermedades hepáticas crónicas provoca una profunda alteración de la arquitectura y funcionalidad del hígado al producirse la «capilarización sinusoidal», que impide el correcto intercambio de oxígeno y factores entre el hígado y la sangre. La hipoxia tisular que se origina estimula la formación desorganizada de nuevos vasos, agravando el proceso. En el caso de la CHB, la HBx ejerce un papel proangiogénico ya que regula la transcripción del factor inducible por hipoxia-1 (HIF-1). De este modo, la HBx promueve el incremento de la transcripción del factor endotelio del crecimiento vascular (VEGF), y de la anhidrasa carbonica-9 (CA9), uno de los genes más sensibles a hipoxia^{112,125,126}. Además, la HBx favorece la formación del complejo MTA1/HDAC (proteína 1 asociada a metástasis/histona deacetilasa), incrementando la estabilidad de HIF-1 y sus acciones proangiogénicas¹²⁷.

Aunque se ha descrito que la HBx es capaz de transformar líneas celulares y células primarias¹²⁸, hay otros estudios que nos indican que la HBx provoca la apoptosis de células transformadas por distintos oncogenes¹²⁹. Se ha observado que la HBx favorece la activación del prooncogén PTTG1, sobreexpresado en HCC y directamente implicado en metástasis¹³⁰. Se ha demostrado que la HBx coopera con c-myc y diversos agentes químicos en la inducción de proliferación y apoptosis¹³¹. La HBx favorece in vivo la hiperplasia celular al activar la vía Wnt-β-catenina y coopera en la inducción tumoral, generada por NRAS¹³². Además de su capacidad transactivadora, que permite la inducción de distintos prooncogenes, la HBx interviene en otros procesos celulares que le confieren su potencial procarcinogénico¹¹². Así, la HBx interacciona con la maquinaria de reparación de ADN a través de DDB-1, que, junto a la inducción de agentes nocivos para el genoma como el óxido nítrico, promueven mutaciones en el ADN del huésped¹³³. Además, la HBx induce la *up*-regulación de la actividad de las MTA1/HDAC y ADN (citosina-5)-metiltransferasa 3A (DNMT3A) e induce la hipermetilación del promotor de algunos genes supresores de tumores como p16^{INK4A}, CDH1, IGFBP-3 (*insulin-like growth factor binding protein 3*) y GSTP1 (*glutathione S-transferase P1*). Por otro lado, la HBx aumenta la expresión de oncogenes a través de la hipometilación de sus promotores como: ASPP2 (*ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region-containing protein 2*), MT1F (*metallothionein 1F*), IL-4R, ALDH1 (*retinal dehydrogenase 1*), RBP (*plasma retinol-binding protein precursor*), CRBP1 (*cellular retinol-binding protein 1*) y CDH6 (*cadherin 6*)¹³⁴. La HBx puede controlar multitud de ARNm de diferentes maneras para promover la progresión del HCC¹³⁴. La HBx causa una formación excesiva de centrosomas y la aparición de husos mitóticos multipolares llevando a defectos en la segregación cromosómica y a la inestabilidad

genética. Esto se ha asociado con la unión de la HBx a HBXIP (*HBx-interacting protein*), proteína que es requerida para el crecimiento y supervivencia de los hepatocitos. La HBx también contribuye a la inestabilidad genética alterando varias vías de reparación de ADN, la HBx se une y secuestra p53 en el citoplasma, inhibiendo su función como supresor de tumores, bloqueando su acción en NER y evitando su acción apoptótica^{83,112}. Provoca inestabilidad genética alterando varias vías de reparación de ADN, la HBx se une y secuestra p53 en el citoplasma, inhibiendo su función como supresor de tumores, bloqueando su acción en NER y evitando su acción apoptótica. La HBx contribuye a los estadios tardíos de la tumorigénesis induciendo MMP, promoviendo la transición epitelio-mesénquima (EMT) y facilitando la metástasis^{83,112}. Por otra parte, también participa en los mecanismos de control del ciclo celular, acelerando el tránsito por los puntos G0/G1 y G2/M50, e interfiere con los mecanismos de apoptosis, favoreciendo la supervivencia del virus⁸³. Todas estas actividades demuestran el papel de la HBx como cofactor tumoral en la transformación de los hepatocitos humanos y en el desarrollo y establecimiento del HCC, independientemente de su discutido potencial oncogénico intrínseco.

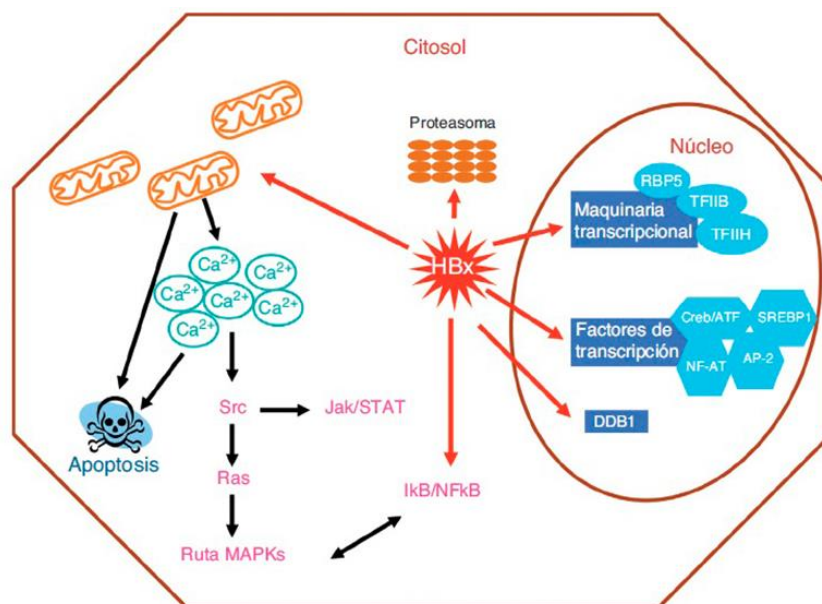


Figura 14 Efectos de la HBx en diversos factores celulares para realizar sus funciones⁸⁷.

1.4.1.4.4 HBwx (*whole HBx*)

En 2016 se identificó un ORF adicional que empieza 168 nt *upstream* del codón de inicio normal del gen X, con actividad transcripcional y que puede ser traducido en pauta con el gen X para formar un gen X completo (wX: *whole-X gene*) de 630 pb que codifica para una proteína X completa de 210 aa (HBwx). Se ha encontrado principalmente en pacientes asiáticos con HCC y genotipo C, con lo que podría estar asociado con la hepatocarcinogénesis. De hecho, HBwx favorece la tumorigénesis y crecimiento *in vivo* e *in vitro* así como migración celular e invasión¹³².

1.5 Regulación de la transcripción.

Aunque el VHB tiene un genoma pequeño, tiene un gran número de secuencias reguladoras transcripcionales que tienen varios roles. La transcripción del genoma y la replicación puede ser estimulada o suprimida por numerosos factores del huésped. En particular, los factores hepáticos nucleares (HNF) que están presentes en hepatocitos en elevadas concentraciones, interactúan con componentes virales y juegan un papel importante en la replicación viral¹³⁵.

1.5.1 Transcritos, promotores y enhancers.

El VHB contiene varios promotores, dos potenciadores y una secuencia de poliadenilación. Los cuatro promotores son responsables de la transcripción de los cuatro ORFs, que dan lugar a 5 ARNm diferentes: ARN_{pc} y ARN_{pg} se transcriben desde el promotor Precore/ pregeómico y son más largos que el propio genoma del VHB (Tabla 1). Los tres ARNm subgenómicos son preS1 y preS2 / S, que se transcriben a partir de los promotores preS1_p y Sp respectivamente y X que se transcribe desde el promotor X (Tabla 1).

En el VHB, además, encontramos dos regiones que actúan como potenciadores y que son llamados s ENH I y ENH II, los cuales tienen actividad potenciadora de la transcripción de los ORFs. El ENH I es una región de aproximadamente 300 nucleótidos localizados entre los ORF S y X¹³⁵ (figura 4). ENH II, que se encuentra antes del promotor basal de core (BCP) y se superpone con el ORF X, *upstream* de la secuencia reguladora (figura 4). La actividad de estos promotores y ENs está regulada¹³³ por Factores de Transcripción (FT) generales y específicos de los hepatocitos que se unen a ellos (tabla 3)⁹². Los factores de transcripción generales que se unen a los promotores o potenciadores incluyen

CAPÍTULO 1. INTRODUCCIÓN

el NF1, la proteína específica1 (SP1), la proteína activadora 1 (AP1), la proteína de unión a TATA (TBP) que está relacionada con la proteína homeobox 1 (PROX1), la proteína de unión a elementos de respuesta a c-AMP (CREB), el NF- κ B, el factor de transcripción de octámero 1 (OCT1), y el factor respiratorio nuclear 1 (NRF1)¹³⁵. Además, en el hígado también está enriquecido con los factores de control de la transcripción específicos, los cuales desempeñan papeles esenciales en el ciclo de vida del VHB. Los ejemplos más representativos incluyen HNF1 α , HNF3 β , HNF4 α , HNF6, y proteína de unión a CCAAT / enhancer α (C / EBP α).

Como los ARNm eucariotas, los ARNm del VHB tienen una capucha o casquete ("cabeza") en su extremo 5' que consiste en la adición de 7-metilguanosa trifosfato, y una cola de poliadeninas en su extremo 3'¹³⁶. La transcripción de estos ARNm es unidireccional y termina en la señal de poliadenilación común (secuencia TATA). Esta señal se encuentra entre los nt 1916 y 1921 (hace referencia a la secuencia AJ344117.1), por tanto, solapada con el extremo 5' de la región que codifica el extremo C-terminal de HBcAg de la ORF PC / C (figura 4). Así pues, todos los ARNm anteriores tienen la misma secuencia en su extremo 3' mientras que el extremo 5' es variable porque depende de la localización de su promotor.

Elemento regulador	Factores de transcripción ubicuos	Factores de transcripción específicos
preS1p	OCT1, Sp1, TBP (<i>TATA-binding protein</i>), NF1	HNF1 y 3 (hepatocyte nuclear factor)
Sp	Sp1, NF1, NF-Y, CREB	HNF1, CBF, COUP-TF, HNF3
ENHII, Cp	Sp1	C/EBP (CCAAT-enhancer-binding protein), HNF3 y 4, PPAR-RXR α , COUP-TF
ENHI, Xp	OCT2, NF1, CREB/ATF2 (<i>activating transcription factor</i>), AP1	C/EBP, C/ENB, HNF1, 3 y 4, PBF, RFX1, RXR α , PPAR (<i>peroxisome proliferator-activated receptor</i>), COUP-TF

Tabla 3 FT que se unen a promotores y enhancers. ⁹².

1.5.1.1 Regulación del transcrito preCore/pregenómico:

El promotor preCore/pregenómico (nt 1613-1849) (hace referencia a la secuencia AJ344117.1) dirige la transcripción de dos ARNs: el ARN_{pc}, que codifica la proteína preCore que es la precursora del HBcAg y el ARN_{pg} (Tabla 1), que es bicistrónico y codifica para la proteína core y la polimerasa y, además, actúa como molde para la

CAPÍTULO 1. INTRODUCCIÓN

transcripción reversa que da lugar al ADNrc¹³⁷.

El Cp está formado por el BCP, una pequeña región de unos 108 pb que es suficiente para dirigir la iniciación correcta del ARNpg y ARNpc y una secuencia *upstream* del BCP que regula su actividad y que se conoce como URR (*upper regulatory región*)¹³⁸. El Cp es extremadamente débil por sí mismo y necesita la activación por parte del ENHI¹³⁹. Además el ENNII tiene la secuencia solapada con este promotor y también puede estimular su expresión¹⁴⁰.

El BCP se encuentra entre el extremo 3' del ORF X y el extremo 5' del ORF PC / Core y se solapa con el DR1. Aunque el ARNpg y el ARNpc se producen a partir de esta región, su síntesis comienza a partir de posiciones diferentes y se regulan independientemente¹⁴¹. La URR contiene elementos reguladores positivos CURS (*core upstream regulatory sequence*) y negativos NRE (*negative regulatory element*) de la actividad del BCP¹³⁸. A su vez el elemento CURS está formado por CURS A (nt 1636-1702), que contiene tres dominios reguladores positivos: *box* α , γ y δ , y CURS B (nt 1702-1742), que contiene un dominio regulador negativo: *box* β ¹³⁸.

El BCP contiene cuatro regiones ricas en AT llamadas *TATA-box like* (TA1 nt 1750-1755, TA2 nt 1758-1762, TA3 nt 1771-1775 y TA4 nt 1788-1795) (hace referencia a la secuencia AJ344117.1). Las tres primeras controlan la síntesis del ARNpc y la cuarta *TATA-box like* controla la síntesis del ARNpg. El ARNpc es ligeramente más largo y la secuencia de su extremo 5' es más variable que la del ARNpg. De manera homóloga a lo que pasa en los promotores celulares de la ARNpol II, el ARNpc y ARNpg tienen sus respectivos elementos iniciadores de la transcripción, que son elementos promotores mínimos y que potencian el efecto de las *TATA boxes-like*^{18,142}.

Toda la secuencia del promotor preCore/pregenómico (BCP y URR) contiene sitios de unión para factores de transcripción celular ubicuos como Sp1 y específicos del hígado como C/EBP, HNF 3 y 4, PPAR, RXR, COUP-TF¹⁰⁶. Estos son responsables de la especificidad de este promotor para expresarse en el hígado y permiten que el ARNpc y el ARNpg se regulen de manera coordinada o diferenciada. Así, por ejemplo, mientras que Sp1 estimula selectivamente la síntesis de ARNpg, HNF4 inhibe selectivamente la síntesis de ARNpc¹⁰⁶. Además, el segundo lugar de unión de HNF4, que se extiende por la secuencia *TATA-Box like* que regula la expresión del ARNpc, también es la diana de otros factores de transcripción, como el heterodímero formado por PPAR-RXR α que activa la síntesis de ARNpg o COUP-TF1, que reprime coordinadamente la síntesis tan de ARNpg como de ARNpc¹⁰⁶ (Tabla 3).

1.5.1.2 Regulación del transcrito preS1 y preS2/S:

La transcripción de las diferentes proteínas de superficie es regulada por dos promotores diferentes (preS1p y Sp). El ENHI tiene poco efecto sobre su actividad, en cambio el ENHII sí puede estimular ambos promotores de manera independiente de su orientación¹⁴³.

El ARN preS1 de 2,4 Kb está regulado por el promotor preSp (Tabla 1), que está solapado con la secuencia que codifica el dominio TP de la polimerasa en la ORF P (Tabla 1). Este promotor es el único del VHB que contiene una secuencia *TATA-box* clásica¹⁴³, donde se une el factor de transcripción ubicuo TBP, que junto con un sitio de unión para HNF -1 situado *upstream*, son los dos únicos motivos que necesita este promotor para su actividad transcripcional basal¹⁴⁴. Aparte de estos dos motivos esenciales, la actividad de preSp también está regulada por factores de transcripción ubicuos como Sp1 y NF1 (que también se une a la *TATA-box*) y específicos del hígado como HNF-1 y 3¹⁴⁵ (Tabla 3).

El ARN preS2 / S de 2,1 Kb se transcribe desde el promotor Sp (tabla 1), que está situado dentro del mismo ORF S, entre las regiones preS1 y preS2. Este no contiene ningún motivo *TATA-box*¹⁴⁶, simplemente contiene un motivo CCAAT y dos elementos de inicio de la transcripción que pueden actuar independientemente¹⁴⁶. El motivo CCAAT, al que se unen los factores de transcripción NF-Y y CBF¹⁴⁷, regula las concentraciones relativas de las diferentes proteínas de superficie a nivel transcripcional¹⁴⁸. En la secuencia del promotor Sp se han identificado siete elementos transcripcionales que regulan el promotor (A-G) cada uno de los cuales contiene dianas para factores de transcripción ubicuos (Sp1, NF1 y NF-Y) o específicos del hígado (CBF, COUP-TF, HNF -3, CREB y HNF-1) (Tabla 3)¹⁴⁹.

1.5.1.3 Regulación del transcrito X:

El promotor X está solapado al extremo 3' del ENHI que estimula su actividad¹⁵⁰⁻¹⁵². De hecho, estos dos elementos forman la unidad que regula la transcripción de la ORF X y sus límites no están bien definidos, aunque se asume que la secuencia de esta unidad reguladora se encuentra totalmente solapada al ORF P, en la región entre los ORFs S y X. La actividad del promotor X también es estimulada por la ENHII¹⁴³. Los transcritos a los que da lugar este promotor tienen diversos orígenes de transcripción y por lo tanto tienen extremos 5' heterogéneos y diferentes tamaños^{151,153}.

CAPÍTULO 1. INTRODUCCIÓN

Este promotor carece de *TATA-box*, pero contiene diferentes secuencias para la unión tanto de factores de transcripción ubicuos (NF-1, AP1, CREB/ATF2, OCT2) como específicos del hígado (C/EBP, C/ENB, PBF, RFX1, RXR α , PPAR) (Tabla 3). Los dos sitios de unión más alejados del inicio de transcripción son 2c y EP, los dos esenciales para la actividad ENHII y constituyen su módulo basal¹⁵⁴. Los otros dos sitios más cercanos al inicio de transcripción y ya solapados con el promotor X, forman un módulo accesorio que aumenta la actividad de la ENHI unas 10 veces pero no son esenciales, y sirven como punto de unión para los factores de transcripción C/EBP, AP-1, CREB y NF1¹⁵⁴. Además otros factores implicados en el control del ciclo celular y la apoptosis, incluyendo p53, c-Abl y RFX-1 se unen específicamente al ENHI y regulan su actividad¹⁵². Recientemente se ha propuesto la existencia de un nuevo promotor situado en medio de la región codificante del gen X¹⁵⁵. Dicho promotor produciría un transcrito del ORF X que empezaría por la posición 1524 del genoma viral, que daría lugar a una HBx más corta, que se traduciría a partir de un segundo codón ATG (*in frame* con el primer ATG e igual de conservado) situado cerca de la posición 1600 del genoma viral.

1.5.2 Otros elementos reguladores.

1.5.2.1 Elemento de respuesta a glucocorticoides (GRE):

Corresponden a una secuencia corta entre 10-20 nt, 735 *upstream* de la diana EcoRI, que sirve de diana para la unión de glucocorticoides. Este activa la transcripción y la replicación viral, y explica la reactivación del virus en tratamientos basados en esteroides¹⁵⁶.

1.5.2.2 Elemento regulador post-transcripcional (PRE)

En el genoma del VHB se ha descrito una secuencia que actúa a nivel post-transcripcional inhibiendo el proceso de *splicing* de los ARN virales y facilitando el transporte y utilización de los transcritos del VHB, lo que es necesario para la expresión de productos génicos del VHB a niveles elevados. Sin embargo, su efecto varía según el tipo de ARN viral¹⁵⁷. Por ejemplo, PRE tiene poca influencia en la exportación al citoplasma del pgARN, por lo tanto, solo regularía su *splicing* mientras que en los ARN subgenómicos (preS1, preS2/S i X) PRE podría también regular su exportación al citoplasma. PRE tiene una longitud aproximadamente de entre 365 y 533 nt y está en la región común en el extremo 3' de todos los ARNm virales¹³⁶.

1.5.2.3 Elementos de repuestos a hormonas (ERH)

Se han identificado varios ERH en el genoma del VHB y funcionan a través de la unión a miembros de la superfamilia de receptores nucleares (RN). HFN4 , RXR α (*Retinoid X Receptor α*) y PPAR α (*Peroxisome proliferator activated Receptor α*) se unen al mismo ERH localizado en el ENHI y activan la transcripción^{158,159}.

1.6 Ciclo replicativo.

El ciclo replicativo del VHB se puede dividir en dos partes. La primera parte incluye desde la unión de las partículas Dane con el receptor hasta la formación de ADNccc. La segunda parte involucra desde la transcripción y traducción de genes específicos del virus hasta la secreción de los viriones y partículas subvirales⁷⁸ (figura 15).

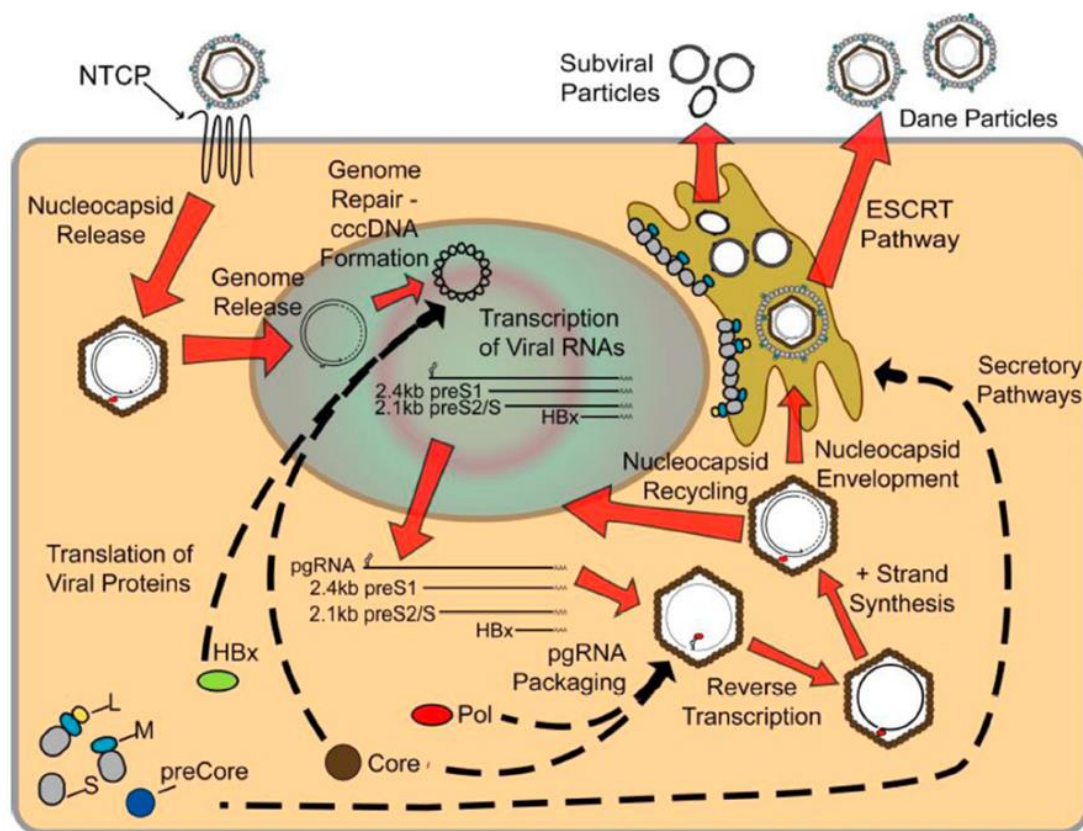


Figura 15 Ciclo replicativo del VHB. Los viriones maduros del VHB entran en los hepatocitos a través del polipéptido cotransportador de sodio-taurocolato (NTCP situado en la membrana celular de los hepatocitos. Después de la liberación de la envuelta viral, la nucleocápside se transporta al núcleo donde se repara el genoma para formar el ADNccc. Utilizando este ADNccc como molde, los ARN virales se transcriben y exportan al citoplasma donde se traducen para formar las proteínas virales. Además, el ARNpg se empaqueta con las proteínas core, junto con la polimerasa, y el genoma viral se replica a través de la transcripción inversa del ARNpg para formar la cadena (-), seguido de la síntesis parcial de la cadena (+). Las nucleocápsides maduras se pueden reciclar de nuevo al núcleo para mantener ahí la población de moléculas de ADNccc, o se pueden envolver y secretar a través de la vía ESCRT. Ver texto para una descripción más detallada de este ciclo¹⁶⁰.

1.6.1 Entrada del virus

El VHB inicia el proceso de infección con una interacción de baja afinidad a los heparán sulfato proteoglicano (HSPG), los cuales son muy abundantes en la membrana plasmática de los hepatocitos, este proceso es esencial para la su infectividad¹⁶¹⁻¹⁶³. Esta interacción depende sobre todo de los residuos del determinante “a” (presente en todas las proteínas de superficie) R/K122 y K141¹⁶⁴. Después de interactuar con HSPG, las partículas virales interactúan con elevada especificidad con su receptor específico, el NTCP, que interactúa específicamente con dominio de unión al receptor de preS1 (figura 16). El gen que codifica el NTCP es miembro de la familia de genes de proteínas transportadoras SLC10 y también se denomina *solute carrier family 10A1* (SLC10A1). Este receptor se expresa principalmente en el hígado y se localiza en la membrana plasmática basolateral de los hepatocitos¹⁶⁵, que está enriquecida además en HSPG.

Después de la interacción de la región preS1 de las partículas virales con el NTCP, las partículas virales son absorbidas en el interior de la célula y forman un endosoma en el interior de la célula hepática¹⁶⁶. Aún no se ha caracterizado el subcompartimento endosómico desde el que las nucleocápsidas son liberadas en el citoplasma, pero se cree que el dominio TM1 de la SHBs (común a la secuencia de las tres proteínas de superficie) tiene un papel esencial en este paso, que podría tener lugar por hemifusión de membranas¹⁶⁵. La composición lipídica de la envuelta viral, y en particular de colesterol, tiene un papel importante en la entrada de las partículas virales⁷⁸. Así pues, al final de este proceso la nucleocápside es liberada y permanece en el citoplasma con la molécula de ADNrc en el interior unida a la proteína P covalentemente¹⁶⁶.

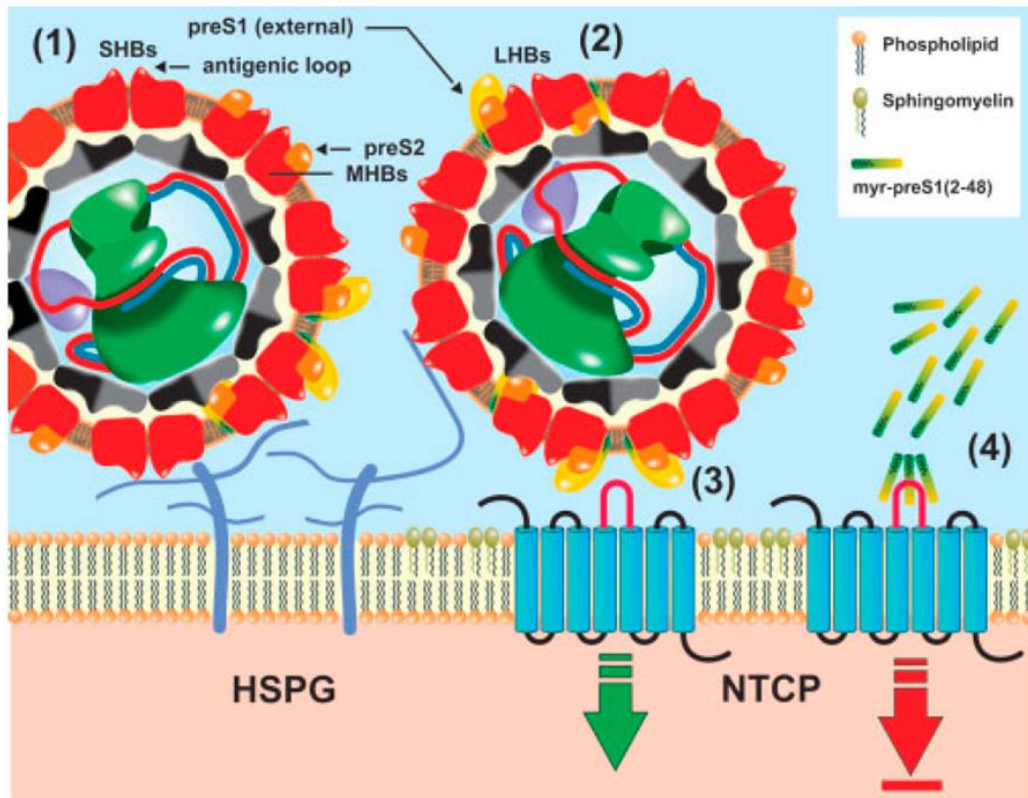


Figura16 Modelo actual para la unión del receptor del VHB a hepatocitos susceptibles, dando como resultado la infección ⁷⁸.

1.6.2 Transporte de las cápsidas a través del citoplasma hacia el núcleo y liberación del ADNrc en el núcleo

Una vez liberadas en el citoplasma, las cápsidas se dirigen hacia los complejos del poro nuclear (NPC) a través de microtúbulos (MT) por un mecanismo de transporte activo (con gasto de energía por parte de la célula)¹⁶⁶⁻¹⁶⁸. Los dímeros de HBcAg tienen una señal de localización nuclear (NLS) que exponen al citoplasma todas las nucleocápsidas que han madurado (es decir, las que la síntesis del ADNrc haya finalizado en su interior) gracias a la cual puede interactuar con proteínas de la superfamilia de la importina β , unos receptores de transporte nuclear que facilitan su interacción con la membrana nuclear a través del complejo del poro nuclear⁴⁸. En la cara nuclear de este complejo se encuentra la canasta nuclear, una estructura en forma de jaula donde las nucleocápsidas quedan retenidas⁴⁸. Esta retención tiene lugar mediante la interacción con la nucleoporina Nup 153⁷⁸, un mecanismo único entre los virus que atraviesan el complejo del poro nuclear. En la canasta nuclear, las nucleocápsidas con un genoma viral en forma de ADNrc (maduro) se desacoplan y liberan el ADNrc y el HBcAg en el carioplasma¹⁶⁹

1.6.3 Formación del ADNccc

Una vez en el carioplasma, el ADNrc se convierte en ADNccc, una molécula de ADN circular covalentemente cerrada de doble cadena y estable¹⁶⁶. Para ello el ADNrc debe sufrir una serie de modificaciones: la liberación de la polimerasa viral unida covalentemente al extremo 5' de la cadena (-) y la eliminación de la redundancia terminal 5' o 5'r a la que esta estaba unida, la eliminación del oligonucleótido de ARN del extremo 5' de la cadena (+) y la finalización de la síntesis de esta cadena (incompleta en el ADNrc) y la ligación covalente de los extremos 5' y 3' de las dos cadenas de ADN viral, lo que da lugar a una estructura similar a un plásmido (figura 17). Finalmente, el ADNccc forma nucleosomas para dar lugar a un minicromosoma.

Estas modificaciones requieren varias actividades enzimáticas, pero los responsables y la forma en que se llevan a cabo aún se desconocen. Parece claro que el primer paso en la formación del ADNccc es la desproteíntización del ADNrc, es decir, la escisión de la polimerasa del extremo 5' de la cadena (-), esta escisión da lugar a la acumulación de una forma intermediaria libre de proteínas o *protein-free rcDNA* (P-Free RC)¹⁷⁰. La polimerasa viral establece un enlace covalente con el ADNrc estructuralmente muy similar a los aductos causados por las topoisomerasas, unas enzimas que regulan el superenrollamiento del ADN. Esto sugiere que el VHB podría utilizar las enzimas que reparan los aductos causados por las topoisomerasas para escindir su propia polimerasa. Uno de los mecanismos de reparación de estos aductos se basa en las enzimas llamadas tirosil-ADN fosfodiesterasas (TDP), que cortan específicamente el enlace tirosil-fosfodiéster, como el que se establece entre el resíduo Y63 del dominio TP de la polimerasa y el genoma viral. Así Königer y colaboradores demostraron *in vitro* que la TDP2 es capaz de cortar enlaces sintéticos que imitan al tirosilfosfodiéster y también es capaz de liberear la polimerasa del ADNrc extraído de cápsides de VHB humano y de pato (DVHB)¹⁷⁰. Pero estos resultados son controvertidos, ya que otros estudios han observado que la enzima TDP2 podría no ser imprescindible para la formación del ADNccc *in vivo*¹⁷¹.

CAPÍTULO 1. INTRODUCCIÓN

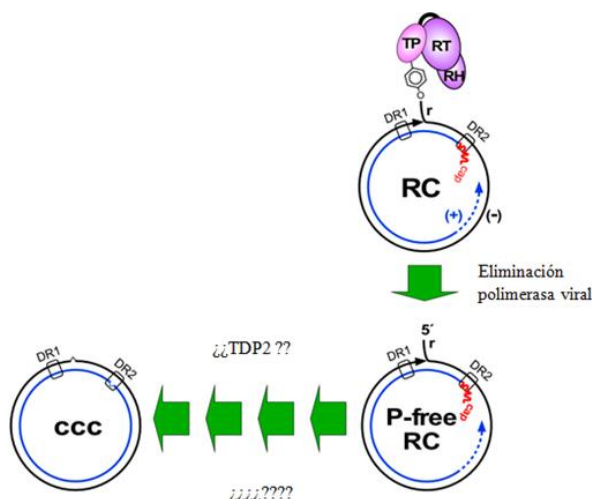


Figura17 Formación del ADNccc. Los responsables y la forma en que se llevan a cabo aún se desconocen. Imagen modificada¹⁷².

Una vez reparado el ADNrc, éste se asocia a histonas y a HBcAg formando nucleosomas y adopta la típica estructura en forma de "collar de perlas" de la cromatina, lo que le permite organizarse como un mini cromosoma de la célula infectada³³. Estas estructuras tienen una vida media de 10-100 días²¹, lo que explicaría la gran dificultad de eliminar la infección¹⁷³. Esta organización permite al ADNccc utilizar mecanismos de regulación epigenética (que permiten regular el acceso al ADN y por lo tanto la expresión génica sin alterar la secuencia). Así se ha demostrado que las histonas H3 y H4 del ADNccc se acetilan para estimular la expresión de los ORFs del VHB¹⁰⁵. La proteína reguladora HBx también tiene un importante papel en la regulación epigenética de la transcripción del ADNccc⁸²(ver en el Capítulo 1 el apartado *funciones de 1.4.1.4. ORF X y la proteína X*).

Una pequeña parte del cccADN no proviene del rcADN sino del ADNdsl. Esta forma del genoma se circulariza probablemente mediante vías alternativas de reparación del ADN nuclear a través de unión de extremos no homólogos (NHEJ). El mecanismo de NHEJ tiene tendencia a cometer errores¹⁷⁴ y la secuencia donde se unen los dos extremos puede contener en muchos casos pequeñas deleciones y / o inserciones, mayoritariamente dentro y alrededor de r, con lo cual impide que el ARNpg sintetizado a partir del ADNdsl circularizado se retrotranscriba correctamente¹⁷⁵.

1.6.4 Transcripción Viral, Formación de Partículas Core

La replicación se inicia en el citoplasma con la unión de la ADN polimerasa viral al motivo *stemloop* llamado señal Epsilon o señal de encapsidación ϵ . Esta señal de encapsidación es el elemento *cis* mejor conocido del ARNpg, corresponde a unas

secuencias de 120 pb redundantes en el extremo 5' y 3' de este ARN, que adoptan una estructura de tallo-bucle. Además contiene las DRs, implicadas en la replicación viral,²⁸ (figura 18). Esta unión ADN polimerasa-señal de encapsidación es esencial para la replicación ya que la unión de la polimerasa a 5'-ε desencadena la encapsidación del ARNpg, seguido de la síntesis del ADN dentro de la cápside (figuras 19 y 20)⁵².

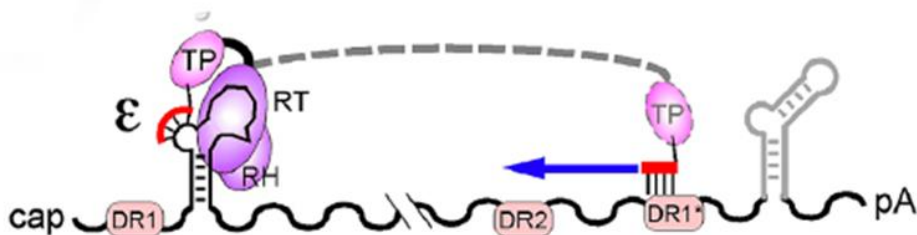


Figura18 Establecimiento del enlace 5'-fosfotirosilo. La proteína P usa un residuo de Tyr en su dominio TP para iniciar la síntesis de un oligonucleótido de ADN utilizando como plantilla la protuberancia ε; su posterior extensión después de la translocación a 3' DR1 * produce el ADN de cadena (-) que lleva una redundancia terminal³⁷.

La polimerasa se une con la ayuda de chaperonas y co-chaperonas como Hsp90, Hsp70, Hsp40, Hop/p60 y p23 a la señal εpsilon en 5' del ARNpg para establecer y mantener a la polimerasa en una conformación adecuada (Figura 19). Para la interacción con ε es necesario que se dé un proceso dinámico dirigido por la hidrólisis de ATP^{52,53}. La formación de este complejo ribonucleoproteico polimerasa-ARNpg-chaperonas desencadena la encapsidación del complejo induciendo la adición de 240 unidades de HBcAg al complejo²⁸. Este proceso de encapsidación depende de fuertes interacciones electrostáticas entre las cargas negativas del ARNpg y las cargas positivas de los residuos de Arg en el dominio C-terminal del HBcAg (también llamado protamina)¹⁷⁶.

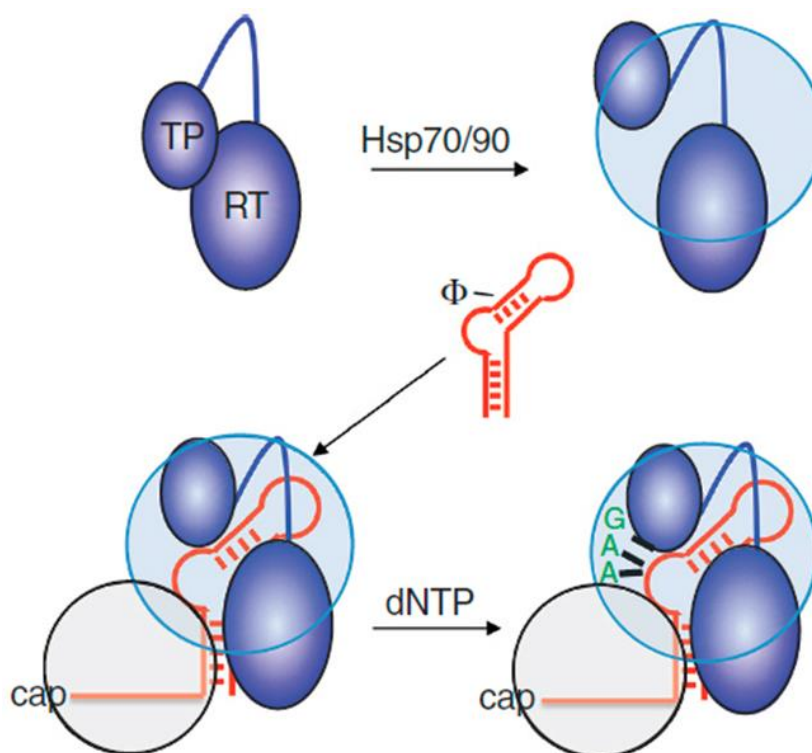


Figura19 Primeros pasos del empaquetamiento pregenómico de ARNpg y cebado de proteínas de la transcripción inversa. La unión del ARNpg a la proteína terminal y el dominio de RT de la polimerasa se ve facilitado por chaperones (círculo azul). Los factores proteicos adicionales que se cree que juegan un papel en el empaquetamiento de ARN al unirse a la estructura ribonucleoproteica formada por el ARNpg y diferentes proteínas (círculo gris)¹⁷⁷. * *dNTP*, desoxirribonucleótido trifosfato; *Hsp90*, proteína de choque térmico 90.

La transcripción inversa es iniciada una vez el ARNpg queda encerrado dentro de la capsida, mediante el cebado por el dominio PT de la polimerasa. Esta proteína cataliza la generación de un *primer* formado por cuatro nt, usando la protuberancia en 5'-ε como molde. La síntesis de este *primer* es posible gracias a la presencia de un grupo -OH en la tirosina 63 de la PT, que permite añadir un residuo dGMP usando el sitio activo YMDD del dominio RT polimerasa y finalmente lleva a la síntesis de un oligonucleótido de cuatro nucleótidos TGAA (4NT) unido covalentemente al dominio PT (figura 18)⁵². Este *primer* 4NT, queda unido covalentemente a la polimerasa y es translocado a 5'UUCA en la región DR1* en el extremo 3' del ARNpg, lo que convierte esta región en el sitio aceptor (AS). Para evitar hibridaciones inespecíficas del *primer* 4NT existen elementos que definen DR1* como el AS correcto: φ (secuencia a 30 nt del extremo 5' de DR1*)¹⁷⁶ y ω (secuencia que se solapa con el extremo 3' de DR1*)¹⁷⁸. Así el 4NT actúa como primer para la síntesis de la hebra(-) a partir del ARNpg por la actividad retrotranscriptasa de P³¹. Mientras que se sintetiza la hebra(-) (figura 20), la actividad ARNasa H de P degrada el ARNpg, excepto una pequeña secuencia de 11-16 nucleótidos de su extremo 5' (11-16NT)⁷⁸, que está unida a la “cabeza” de este ARN.

CAPÍTULO 1. INTRODUCCIÓN

Este pequeño fragmento resistente a la ARNasa H, contiene la secuencia DR1, cosa que permite su traslocación por hibridación a la DR2 localizada en el extremo 5' de la hebra(-) y actúa como primer para la síntesis de la hebra(+) mediante la actividad polimerasa de la polimerasa viral⁵³. Este último paso está limitado por la disponibilidad de los nt dentro de la cápside para la replicación, por eso durante la maduración de las cápsides la hebra(+) queda incompleta y el genoma del VHB adopta una conformación de doble hebra parcial (ADNrc) (figura 20)⁵³.

Es en este último punto donde se puede originar el ADN lineal de doble cadena (ADNdsl) por *priming in situ*, el primer de ARN en vez de translocar 11-16NT a DR2 y se extiende la hebra (+) desde su posición original. Por lo que esta forma minoritaria se puede encapsidar y dar lugar a ADNccc, que por retrotranscripción, puede dar lugar a nuevos ADNdsl¹⁷⁵, aunque de hecho esta forma del genoma del VHB se integra con más frecuencia en el genoma del huésped que la forma de ADNrc¹⁷⁵.

Para la síntesis de la hebra (+) del ADNrc, es imprescindible la intervención de las redundancias terminales 5' y 3'. Éstas se yuxtaponen debido a la conformación que adopta la hebra (-) después de su síntesis, ya que la señales DR1 y DR2 se acercan. De este modo, el extremo 3' de la nueva hebra (+) puede salvar la discontinuidad de la hebra (-) pasando de la r en 5' a r en 3' (un tercer cambio de molde (figura 20).

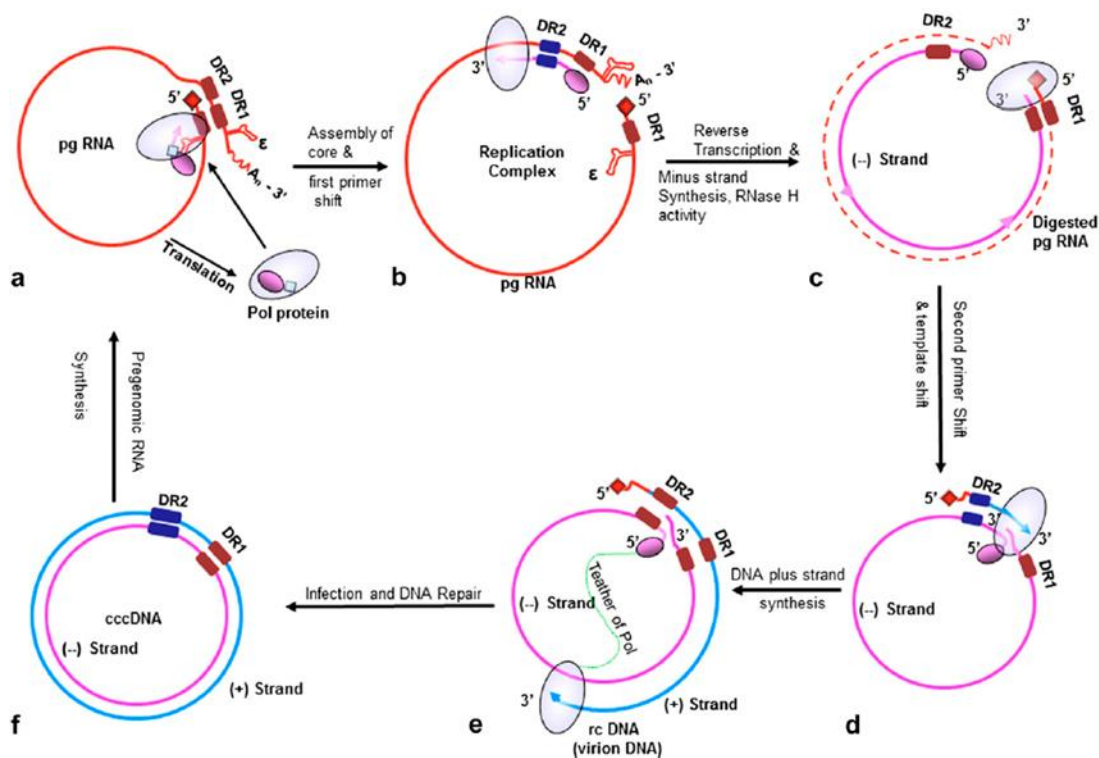


Figura 20 Pasos consecutivos de la replicación del genoma del VHB. (a) Transcripción inversa de ARNpg e iniciación de la síntesis de cadena (-) utilizando el dominio TP de Pol y la región de protuberancia de 'ε' como plantilla. (b) Primer desplazamiento del cebador: translocación de la hebra (-) naciente hasta el extremo opuesto del ARNpg por el apareamiento del primer 4NT con DR2. (c) Terminación de cadena (-) usando ARNpg como molde y degradación simultánea de ARNpg. (d) Segundo cambio de molde y desplazamiento-reasignación del iniciador del ARN (*primer* 11-16NT) al extremo de la cadena (-) (DR2), seguido de circularización del genoma debido al tercer cambio de molde gracias a las secuencias redundantes presentes en la hebra (-). (e) Formación de ADNrc, elongación incompleta del ADN de la hebra (+) usando la hebra (-) como modelo. (f) Reparación de ADNrc después de la infección y entrada en el núcleo, para crear ADNccc³⁷.

1.6.5 Maduración y secreción de viriones y partículas subvirales

Un aspecto intrincado del ciclo de vida del hepadnavirus es la formación de la envuelta alrededor de las cápsides. Éstas adquieren la envuelta lipoprotéica en MVB y la gemación se cree que depende de la maquinaria ESCRT (*endosomal sorting complexes required for transport*)¹⁶⁶, sin embargo, las señales responsables del ensamblaje con las proteínas de la envoltura, que residen en las membranas de ER, no se conocen. Las formas maduras de las nucleocápsides pueden envolverse ya que son las únicas capaces de interactuar con el dominio específico de las proteínas de superficie LHBs. Parece que el estado de fosforilación de las Arg de la región protamina de HBC forman parte de la señal que indica que una cápside es madura¹⁶⁶. Sin embargo, queda por esclarecer como se envuelven las partículas vacías y las de ARN. Por otro lado, también se pueden liberar cápsides “desnudas” (sin envuelta), tanto vacías como las que contienen el genoma viral,

CAPÍTULO 1. INTRODUCCIÓN

por un mecanismo que parece estar regulado por la *apoptosis linked gene II interacting protein X*¹⁷⁹. El papel de estas cápsides desnudas se desconoce.

El destino de las cápsides formadas *de novo* depende de la etapa de la infección del hepatocito. En estadíos iniciales de la infección, algunas de las cápsides *de novo* pueden reinfectar el núcleo, aumentando el pool de ADNccc nuclear y la variabilidad de este¹⁶⁶. Los hepatocitos infectados de manera natural por hepadnavirus contienen entre 1 y 50 copias de ADNccc, si a esto le sumamos la larga vida media de los hepatocitos se asegura que el ADNccc no se pierda durante la división celular a pesar de no repartirse por igual entre las células hijas¹⁷. En fases más tardías de la infección, las nucleocápsides se unen a las proteínas de la envuelta y son secretadas al torrente sanguíneo. En el modelo de DVHB, se ha observado que el destino de las nucleocápsidas está regulado por LHBs¹⁸⁰: Si LHBs está ausente, las nucleocápsidas liberan el genoma viral en el núcleo, por eso el reciclaje nuclear tiene lugar sobre todo en la infección inicial. Por contra, si la LHBs está presente, las cápsides se envuelven formando partículas virales y se segregan.

2. Variabilidad genética.

2.1 Complejidad de quasiespecies

2.1.1 Definición y significado clínico.

El VHB se replica con tasas de error en torno entre 10^{-4} a 10^{-5} nucleótidos (substituciones/tamaño/año)¹⁸¹ Como consecuencia de la limitada fidelidad de copia de su material genético y de su elevada tasa de replicación es poco probable que se pueda copiar el genoma viral completo sin introducir mutaciones, además el genoma del VHB adquiere un alto nivel de mutaciones espontáneas que se incrementan aún más con los medicamentos antivirales o el estado de activación inmunológica. Como resultado de estas tasas de error tan elevadas en las poblaciones virales, aparecen una multitud de genomas mutantes cuya frecuencia relativa varía dependiendo del ambiente y de las interacciones que se establecen entre los genomas (y sus productos de expresión) de la misma población. A las distribuciones de genomas estrechamente relacionados entre sí, aunque no idénticos, y con un origen evolutivo común (un solo genoma ancestral o unos pocos que establecieron una infección y de estos irradiaron todos los demás) se les denomina quasiespecies virales (QS)¹⁸². Las Nubes o espectros de mutantes son el conjunto de mutantes que en un momento dado forman una QS. En general, las nubes de

CAPÍTULO 1. INTRODUCCIÓN

mutantes no son estáticas, sino cambiantes en el tiempo. Ello se debe a que el ambiente en el que se multiplican los virus en un organismo infectado suele variar tanto por alteraciones internas del propio organismo (respuesta inmune, alteraciones metabólicas, fiebre, etc.) como por intervenciones externas (tratamientos antivirales, inmunoterapia), con lo que varía la capacidad replicativa de los mutantes que forman la nube (a veces denominada de modo más general aptitud biológica) o *fitness*. Este parámetro es muy importante para entender la evolución de los virus. A la variación en el tiempo de la composición de los espectros de mutantes se denomina dinámica de QS¹⁸². Por un lado, esta dinámica es una consecuencia de los principios darwinianos de la generación de variantes genéticas de un virus, es decir la competición y el dominio de las más adaptadas a cada ambiente particular. Pero además de esta selección positiva, existe, la deriva genética, la cual juega un importante papel en el proceso de diversificación del VHB. La deriva genética se acentúa durante la transmisión de virus, proceso que a menudo constituye un cuello de botella poblacional, cuya intensidad depende del modo de transmisión. La dinámica de QS y el efecto de cuellos de botella poblacionales se representan de manera esquemática en la figura 21

Es decir, tanto por la alta mutabilidad asociada a la baja fidelidad de su maquinaria replicativa como por la dinámica de transmisión con cuellos de botella poblacionales que actúan sobre poblaciones altamente heterogéneas, el VHB sufre un continuo proceso de variación cuya fase inicial es la diversificación de espectros de mutantes en todos y cada uno de los individuos infectados. A lo largo del tiempo, el proceso de los sucesivos cuellos de botella poblacionales y los episodios tanto de selección positiva como de selección negativa (eliminación de las variantes con menor *fitness*) conduce a las formas dominantes que se han observado con mayor frecuencia en la práctica clínica y que han servido como base para la clasificación de los virus. La propiedad que cuantifica la diversidad y la frecuencia de haplotipos (conjunto de genomas con la misma secuencia nucleotídica), independientemente del tamaño de la población viral se define como complejidad de una QS¹⁸³. La complejidad indican el potencial patogénico, la evolución clínica, la seroconversión a anti-HBe o la respuesta al tratamiento antiviral¹⁸⁴⁻¹⁸⁸, con lo cual es importante poder cuantificarlos, esto se logra mediante los índices de diversidad. De manera ideal, la complejidad de la QS debería basarse en todo el repertorio de mutantes de la población pero esto no es posible con la metodología actual y se extrapolan los resultados del análisis de una muestra limitada¹⁸⁹.

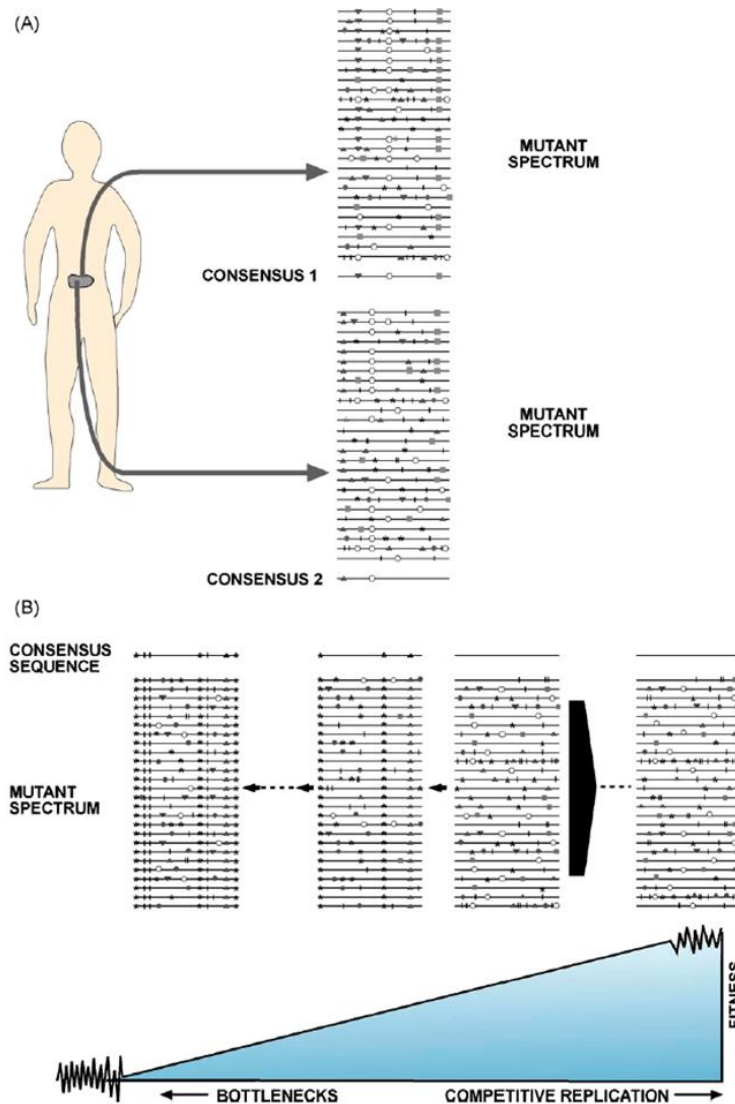


Figura 21. Representación esquemática de la quasispecies y variaciones en el *fitness*. La evolución de la quasispecies está guiada por el gradiente de *fitness* del ambiente, es decir, los mutantes con mayor *fitness* pueden dominar por replicación competitiva^{183,188}. A: De manera simplificada se describe una quasispecies con dos espectros de mutantes con su secuencia consenso en un individuo infectado. Cada línea horizontal representa un genoma y cada símbolo un tipo de mutación. Una quasispecies real puede contener miles de genomas continuamente mutando para formar un conjunto enorme. B: La quasispecies actúa como una unidad de selección, las flechas pequeñas indican pasos repetidos de cuello de botella que resultan en un descenso del *fitness*. La flecha grande representa pasos a poblaciones más grandes que generalmente resultan en una ganancia de *fitness*.

La naturaleza de la QS del VHB le permite adaptarse rápidamente a los cambios del ambiente, la QS sería una unidad de selección, un espectro de mutantes que poseen diferentes niveles de *fitness* en determinados ambientes¹⁸². Por eso, muchas de las propiedades importantes del virus no se pueden explicar por una secuencia consenso y requieren el conocimiento de las variantes presentes en la población viral¹⁹⁰.

2.1.2 Índices para el estudio de la complejidad de la quasiespecie.

El paralelismo entre las descripciones de la diversidad ecológica y de la QS de la población viral han conducido a la introducción de la sistemática utilizada en ecología para medir la complejidad de la QS en una infección viral. En un alineamiento múltiple de haplotipos se muestran las entidades que están presentes en la QS viral. Acompañando a los haplotipos suelen ir indicadas las frecuencias relativas que nos informan de la abundancia de esas entidades, las cuales están estrechamente relacionadas con el *fitness* de cada haplotipo en el momento en el que se analiza la QS. Cualquier índice de diversidad individual proporciona una visión parcial de la información completa sobre la complejidad de la QS, por lo que ningún índice individual puede capturar la complejidad de la población viral en su totalidad. Así pues, se ha sugerido la utilización de un análisis multivariado, esto quiere decir que se recomienda examinar el mismo conjunto de datos usando diferentes índices¹⁸⁹

Los índices más comúnmente utilizados para comparar la complejidad de las quasiespecies virales son la entropía de Shannon (H_{SN}), el índice de Gini-Simpson (H_{GS}), la frecuencia de mutación (M_f) y el índice de diversidad nucleotídica (π)¹⁸⁹. La H_{SN} y el H_{GS} miden la complejidad teniendo en cuenta el número de haplotipos y sus frecuencias relativas en la QS, por lo que se denominan índices de abundancia. En cambio el M_f y π miden la complejidad de la QS en función de las diferencias entre los haplotipos que la forman, por lo que los consideremos índices funcionales¹⁸⁹.

1. **Índices de abundancia:**

Entropía de Shannon normalizada H_{SN} : mide la diversidad o uniformidad de una QS cuando es normalizada a la diversidad máxima. La H_{SH} permite medir la diversidad de la QS teniendo en cuenta la cantidad de haplotipos o variantes que se han encontrado y la frecuencia relativa entre ellos. Varía entre 0 y 1, valor mínimo cuando todas las variantes son idénticas y máximo cuando todas son diferentes, pero no tiene en cuenta la heterogeneidad de las diferentes variantes de la QS porque es insensible al número de mutaciones¹⁸⁹.

$$H_{SN} = - \sum_{i=1}^H p_i \log p_i / \log(H) *$$

* p_i : frecuencia del haplotipo i en la población de QS. H : número de Haplotipos en la población de QS

Gini- Simpson H_{GS} : mide la probabilidad de que dos individuos, dentro de un mismo genoma seleccionados al azar pertenezcan a un haplotipo diferente. Puede variar entre 0 y 1, el valor mínimo se da cuando todas las variantes son idénticas y el máximo cuando todas son diferentes¹⁸⁹.

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

* p_i : frecuencia del haplotipo i en la población de QS

2. Índices Funcionales:

Mutation Frequency (Mf): Proporción de sitios mutados en una población de genomas virales con respecto al haplotipo dominante (máster) o la secuencia consenso, el valor es normalizado por el total de nt de la secuencia. Así pues, mide la heterogeneidad genética respecto a al haplotipo máster o una secuencia consenso de la QS, Cuanto mayor es el valor más diferentes son los individuos de la población con respecto a ésta¹⁸⁹.

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

* d : Proporción de mutaciones relacionadas con la secuencia master

Nucleotide Diversity (π): mide la heterogeneidad genética global de la población teniendo en cuenta la media del número de mutaciones entre cada pareja de haplotipos de la QS. Es decir mide el promedio de nucleótidos diferentes entre dos genomas cualesquiera de la QS¹⁸⁹.

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

* p_i : frecuencia del haplotipo i en la quasispecies viral, p_j : frecuencia del haplotipo j en la quasispecies viral, d_{ij} : distancia de Hamming (número de mutaciones que diferencia el haplotipo i y j).

2.1.3 Estudio de la conservación de la quasiespecie.

El estudio de la conservación de la QS se basa en el contenido de información (CI). El CI se define como el número de decisiones binarias (número de preguntas cuya respuesta es sí / no) requerido para encontrar el elemento correcto en un conjunto de N elementos. Como la probabilidad de encontrar un nt/aa en una posición determinada de un alineamiento múltiple de nt o aa es incierta, Shannon decidió ampliar el CI introduciendo esta incertidumbre por lo que podemos definir el CI basado en la entropía de Shannon para un alineamiento múltiple de secuencias de nt/ aa como una medida clásica de conservación, y se define como:

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

**j: posición j en la alineación y, p_j: frecuencia del haplotipo j en la quasiespecie viral, en la posición j. N el número de posibilidades (4 nt y 22 aa). IC varía desde 0, lo que significa incertidumbre máxima o variabilidad hasta 2 bits en nt y 4.322 bits en aa, lo que significa máxima información o conservación.*

La conservación de la QS se verá influenciada por la forma en que consideramos una mutación en ella. En genética humana, una mutación se considera fijada en una población si se encuentra al menos en el 1% de la población. Esto no es equiparable si hablamos de entidades como las QS donde las mutaciones pueden aparecer en cualquier abundancia en un momento dado. Debido a esta dificultad, en los estudios de esta tesis se consideraron dos escenarios que proporcionan valores límite:

- Primer escenario: consideremos fijadas solo las mutaciones que aparezcan en la secuencia de consenso para cada paciente. Aquí solo consideramos el nucleótido más abundante en cada posición dentro de cada paciente. Los valores de CI serán un límite superior de conservación.
- Segundo escenario: consideremos todas las mutaciones, siempre que aparezcan en haplotipos que no sean inferiores al 0,25% en el alineamiento múltiple por paciente. Los valores de CI serán un límite inferior de conservación.

Para el estudio de la conservación y basándonos en el CI en esta tesis hemos utilizado lo que se conoce como *Sliding window*, una herramienta que utilizamos para analizar fragmentos de la secuencia de los haplotipos (ventanas) de N nt o aa, comenzando desde la primera posición en el alineamiento múltiple, y avanzando en pasos de 1 nt o aa. Es

decir, cada punto del *sliding window* nos indica la media del IC de N nt o aa, el siguiente punto representaría otra ventana de N nt o aa en los que se incluirían todos los nt o aa de la ventana anterior excluyendo el primero y añadiendo el siguiente nt o aa del alineamiento y así sucesivamente (Figura 22A). Para cada ventana se calcula la media del IC de cada posición dentro de la ventana. El objetivo es representar gráficamente las medias obtenidas cada ventana (Figura 22B), lo cual puede ser útil para ubicar el fragmento de N nt o aa con el CI más alto (más conservado) dentro del alineamiento múltiple. En el caso de nt hablaremos de ventanas de 25 pb y en los aa en ventanas de 10.

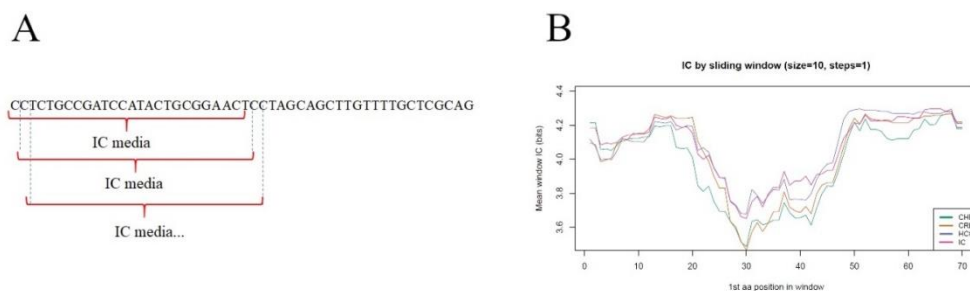


Figura 22. A Representación gráfica de las ventanas donde se calcula la media del IC. Representación de cómo se va aumentando un nucleótido en cada ventana, para el cálculo medio de la IC. B Representación gráfica de las medias calculadas en cada ventana mediante un *sliding window*.

Para obtener una representación gráfica del CI, usamos los logos de secuencias que nos ayudan a representar gráficamente la conservación posición a posición de un fragmento de interés dentro de un haplotipo, ya sea representando nucleótidos o aminoácidos, esto lo conseguimos gracias a la fórmula que nos proporciona Rseq (figura 23).

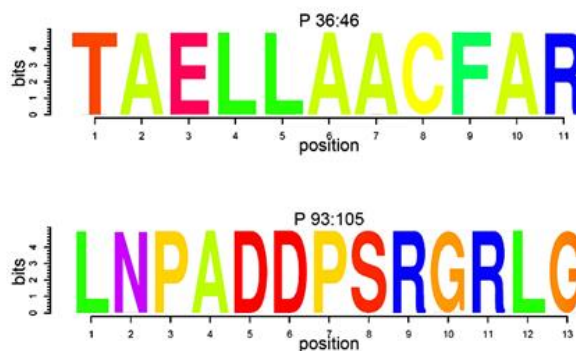


Figura 23 Representación de la conservación de un fragmento de la secuencia de un haplotipo (de aminoácidos), mediante logos de secuencia.

2.2 Herramientas para el estudio de la quasispecies

El estudio de la dinámica de la QS del VHB junto con la progresión la enfermedad es determinante para predecir la patogenia y la selección del mejor tratamiento disponible¹⁸⁴. Este tipo de estudio puede ser realizados por métodos clásicos (métodos indirectos y técnicas de secuenciación de primera generación) o por enfoques más potentes (secuenciación masiva o *next generation sequencing* (NGS)).

2.2.1 Métodos indirectos:

Existen algunos métodos moleculares indirectos que se han utilizado ampliamente, los cuales están principalmente enfocados a la detección de mutaciones específicas. El más común y el más usado dentro de los métodos indirectos es el LiPA, INNO-LiPA DR (LiPA Innogenetics, Ghent, Belgium) el cual puede detectar las variantes específicas aunque se encuentren en pequeñas proporciones (5-10%)^{185,191,192}. Otros, métodos menos utilizados para detectar estas variantes son el análisis de fragmentos de restricción de longitud polimórfica (RFLP)¹⁹³⁻¹⁹⁶, los ensayos de 5'-nucleasa¹⁹⁵, el análisis de curvas de fusión (*melting curves*)¹⁹⁷, tecnología de chip de ADN¹⁹⁸ y la PCR en tiempo real usando cebadores específicos de mutaciones¹⁹⁹. Aunque estos métodos pueden diferenciar entre mezclas de poblaciones²⁰⁰, no permiten la detección simultánea de diferentes sustituciones en el mismo genoma viral. Además de esto, el hecho de que solo permitan detectar mutaciones específicas no permite estudiar adecuadamente las variantes complejas, por lo que podemos decir que no son adecuados para describir la composición de las quasispecies.

2.2.2 Técnicas de secuenciación de primera generación o secuenciación directa:

Durante los últimos 50 años, el método más utilizado para la secuenciación del ADN ha sido el de Sanger²⁰¹⁻²⁰⁴, que permite determinar la secuencia de fragmentos de ADN de entre 300 y 900 nt. La limitación principal de esta técnica es que no permite determinar secuencias de nt de haplotipos individuales, si no que la secuencia obtenida, es la media de los haplotipos presentes en la QS en una proporción ≥ 20 % de éstas²⁰⁵. Por lo tanto, la secuenciación por Sanger presenta dos principales limitaciones: baja sensibilidad para la detección de haplotipos minoritarios en la QS e incapacidad para realizar análisis de haplotipos individuales²⁰⁶. El método tradicional utilizado para analizar secuencias individuales en un entorno de heterogeneidad poblacional genómica, como es la QS viral, es la clonación en un plásmido de fragmentos del genoma viral obtenidos por PCR y su

posterior secuenciación¹⁸⁴. Esto supone un proceso costoso y laborioso ya que requiere de múltiples experimentos y complejos pasos, que acaban proporcionando una resolución limitada en lo que a espectro de mutaciones y frecuencias se refiere. Para obtener una representación significativa de una QS viral, el análisis de una QS por clonación requiere una gran cantidad de clones (> 100) lo que significativamente dificulta su uso²⁰⁶. A pesar de su destacado éxito, estas técnicas que se consideran de primera generación (técnicas de secuenciación directa) carecen de un alto rendimiento, y se caracterizan por su limitada adaptabilidad al laboratorio clínico ya que no son coste-efectivas para la secuenciación de múltiples muestras y/o genomas largos²⁰⁷.

2.2.3 Secuenciación masiva o next generation sequencing (NGS)

El escenario de la secuenciación del genoma ha cambiado mucho, principalmente debido al desarrollo de técnicas NGS²⁰⁸. La NGS permite el análisis de alto rendimiento de fragmentos de ADN, secuenciando simultáneamente miles de ellos que han sido clonalmente amplificados. Por lo tanto, son herramientas poderosas para caracterizar la diversidad genética en poblaciones virales²⁰⁷. Las plataformas más extendidas y actualmente disponibles en el mercado son el analizador de genoma Solexa (Illumina)²⁰⁸, y el sistema Ion Torrent (Life Technologies)²⁰⁹. Es necesario mencionar en este apartado a la plataforma 454 FLX (Roche) (dada de baja desde diciembre 2017)²¹⁰, que fue una de las más utilizadas hasta su desaparición. Estas plataformas aplican diferentes estrategias experimentales de secuenciación (enzimológico, químico), poseen una ingeniería compleja (óptica de alta resolución, hardware y software), y proporcionan un ahorro de tiempo significativo y mínimos equipos asociados en comparación con la clonación.

Las tecnologías NGS buscan la amplificación de cadenas sencillas de ADN agregando cebadores (*primers*) por unión química o por PCR específica y realizan reacciones de secuenciación en los filamentos amplificados. Esto es lo que permite la secuenciación de genomas virales individuales sin requerir un paso previo de clonación en un plásmido. Lo que permite obtener un rendimiento mucho mayor que el método Sanger, permitiendo secuenciar al mismo tiempo miles de moléculas a en un menor que el método de Sanger. Sin embargo, a diferencia de éste, la tasa de error es más alta, lo que obliga a un filtrado bioinformático de los datos obtenidos, previo a su interpretación.

2.2.3.1 NGS de fragmentos cortos:

Secuenciación por síntesis (SBS): se usa una polimerasa y una señal, como un fotón de

luz o un cambio en la concentración iónica, identifica la incorporación de un nucleótido en la secuencia que se amplifica clonalmente en una superficie sólida. Dentro de SBS podemos encontrar dos enfoques: los terminadores cíclicos reversibles (CTR) y los que adicionan nucleótidos individuales (SNA)²⁰⁸.

Pirosecuenciación 454 o UDPS (Ultra-deep pyrosequencing): El primer estudio que forma parte de esta tesis se ha basado en la tecnología de UDPS, la cual utilizan las plataformas GS-FLX + y GS-Junior distribuidas por Roche (ahora discontinuadas). Por esta razón haremos especial énfasis en su descripción. Estas plataformas están basadas en SBS por SNA y una de sus principales ventajas es el hecho que secuencia amplicones relativamente largos en comparación con las otras plataformas. Esto permite caracterizar de manera más completa el perfil de cambios de nt y aa de las diferentes variantes de la QS.

La obtención de las lecturas de las secuencias de las moléculas de ADN consiste en:

1. **Creación de una biblioteca de amplicones de la región de interés**: Al igual que la secuenciación por método de Sanger, el primer paso del procedimiento de secuenciación es delimitar la región de la que interesa determinar la secuencia, teniendo en cuenta la limitación de la longitud máxima del fragmento que permite secuenciar la plataforma: Esta región se amplificará mediante PCR, en el último paso de amplificación se deben utilizar unos *primers* denominados por el fabricante *fusión primers* los cuales llevan unos adaptadores en sus extremos 5' y 3' que incorporaran a la secuencia de los amplicones obtenidos y les permitirán interaccionar con los diferentes elementos del sistema de UDPS (figura 24). El conjunto de amplicones obtenido es lo que se denomina biblioteca de amplicones.

A:5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-GTTGTA AACGACGGCCAGT-secuencia específica template-3'
 B:5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-CACAGGAAACAGCTATGACC-secuencia específica template-3'

Figura 24 Los adaptadores usados en la tecnología 454 Roche son primers de 55 nt añadidos a cada extremo de los template (secuencias molde, de las que nos interesa determinar la secuencia), los primeros 25 pb en el extremo 5' son los adaptadores que permiten la unión de los amplicones a los elementos del sistema de secuenciación (amarillo), y las 4 pb en el extremo 3' TCAG, son una secuencia no palindrómica "llave". Hay dos clases de adaptadores A y B, que se diferencian en la secuencia de nt²¹⁰. Después está la secuencia MID (*Multiplex Identifier*), un identificador único de 10 nt establecidos por el fabricante y que se utiliza durante el análisis bioinformático para identificar la muestra a la que pertenece cada lectura obtenida, el cual va unido al M13 (secuencia de 20 nt que permite unir esta secuencia con el template (azul)). Opcionalmente, permite mezclar en el mismo experimento los amplicones de diferentes muestras. En el extremo 3' está la secuencia específica del template

1. Amplificación clonal en la PCR en emulsión (emPCR). Mediante esta técnica las moléculas de ADN de la librería se desnaturalizan para obtener el template monocatenario (ADNmc), esta molécula se hibrida con esferas (*beads*) cubiertas con la secuencia complementaria al adaptador (A o B) de manera que una sola molécula de ADN se une a cada *bead*²¹⁰. Se diluye para garantizar la ratio: una molécula de ADN por cada *bead*²¹⁰, para ello previamente la librería de ADN es purificada, cuantificada y comprobada su calidad. Una vez que la librería está inmovilizada en las *beads*, es emulsionada con los reactivos necesarios para la amplificación (dNTPs, primers y la ADNpol) en gotas acuosas embebidas en aceite (micelas), de manera que cada gota contiene una única *bead* con una única molécula de ADN y actúa como un microreactor para la PCR²¹⁰. Con la emulsificación se consiguen aproximadamente 1000 microrreactores de PCR por microlitro y gotas de diámetro 100-150 μm ²¹⁰, dentro de las cuales se produce la amplificación por PCR para crear *beads* recubiertas con miles de copias de la misma secuencia (figura 25)²¹¹.

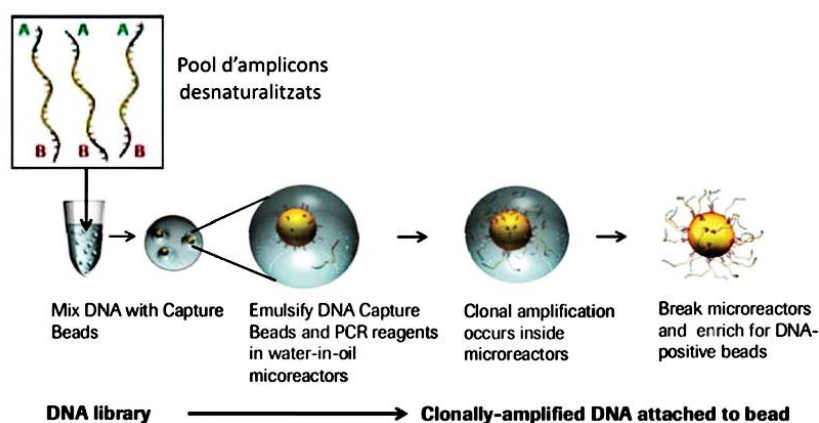


Figura 25 . Amplificación clonal en la PCR en emulsión. Figuras proporcionadas por el fabricante y modificadas (Roche, 454 Life Sciences, Branford, USA).

Después de la emPCR se rompe la emulsión y se cuentan las *beads* recuperadas, la recuperación óptima está entre el 65 y 85%. Se obtienen dos tipos de *beads*, *beads* con ADN amplificado y *beads* nulas sin producto amplificado. A continuación, se lleva a cabo el proceso de enriquecimiento, que consiste precisamente en capturar las *beads* con ADN (que podemos secuenciar) y eliminar las nulas, para ello se añade un buffer de hibridación que contiene un primer de enriquecimiento: 5' Biotin-18-atom 72mul-etilenglicol espaciador 5'-CGTTTCCCCTGTGTGCCTTGCCATCTGTCCCTCCCTG-

TC-3' complementario al adaptador en el extremo 3' del template inmovilizado en las *beads*. Esta mezcla de *beads* con ADN y nulas se incuba con *beads* paramagnéticas recubiertas de estreptavidina que se unirán al *primer* de enriquecimiento biotinilado. El tubo se introduce en un imán para que se unan las *beads* paramagnéticas y el sobrenadante con *beads* nulas se elimina. A continuación, se libera la hebra no biotinilada y las *beads* que permanecen en el imán son usadas como librería de *templates* de cadena única. A las *beads* recuperadas se les añade el buffer de hibridación y el primer de secuenciación (5'-CCATCTGTTCCCTCCCTGTC-3'), se cuantifican y pueden almacenarse a 4°C. El enriquecimiento óptimo está entre el 5-20% de los 10 millones de *beads* originales añadidas a la emulsión (500.000-2.000.000 *beads*). Las *beads* se depositan en una placa de fibra óptica PTP (*PicoTiterPlate*), de manera que solo entre una *bead* por pocillo (figura 26)²¹⁰.

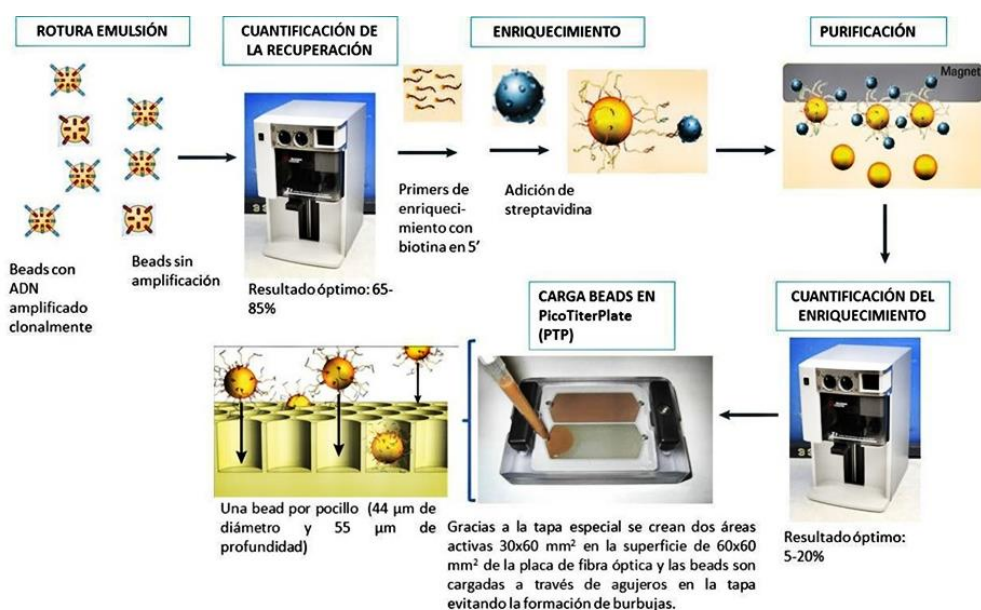


Figura 26 Pasos previos a la reacción de secuenciación en el instrumento 454™. Figuras proporcionadas por el fabricante (Roche, 454 Life Sciences, Branford, USA).

2. Pirosecuenciación: Seguido a la carga de las *beads* con ADN amplificado en los pocillos de la PTP, se añaden *beads* recubiertas de sulfúrida y luciferasa y dNTPs que fluyen a través de los pocillos y se depositan en ellos, en las concentraciones idóneas para que pueda ocurrir la reacción de pirosecuenciación. Esta reacción se realiza en cinco pasos (figura 27)²¹⁰:

3.

CAPÍTULO 1. INTRODUCCIÓN

a) El ADN de cadena simple, unido a las *beads* con DNA recuperadas después del enriquecimiento, hibrida con el *primer* de secuenciación y se incuba con las enzimas ADN polimerasa, ATP sulfurilasa, luciferasa y apirasa, más los sustratos adenosina-5'-fosfosulfato (APS) y luciferina, en los pocillos de la placa PTP.

b) La adición en ciclos de uno en uno de los cuatro dNTPs inicia el segundo paso, en el que la ADN polimerasa cataliza la incorporación del dNTP al molde si es complementario. Si hay incorporación se libera una cantidad de pirofosfato inorgánico (Ppi) equivalente a la cantidad de dNTPs incorporados.

c) La ATP-sulfurilasa convierte cuantitativamente el Ppi en ATP en presencia de APS. El ATP generado permite la conversión de la luciferina en oxiluciferina por acción de la luciferasa, generando luz visible en cantidades proporcionales a la cantidad de ATP presente. La luz emitida es detectada por una cámara de alta resolución CCD (*charge-coupled device*). Cada señal luminosa es proporcional a la cantidad de nucleótidos incorporados.

d) Para continuar con la secuenciación, es esencial la degradación de aquellos dNTPs que no han sido incorporados. La apirasa es la enzima encargada de ello.

e) Finalmente se libera un nuevo dNTPs que puede ser añadidos para iniciar un nuevo ciclo.

3. **Captura de imagen.** El instrumento 454TM consiste en tres sistemas principales: el sistema de fluido, el sistema óptico y la cámara de flujo²¹⁰. La placa se monta en la cámara de flujo diseñada para crear un canal 300 μm encima de la apertura de los pocillos a través de la cual fluyen los reactivos de secuenciación. La parte de debajo de la placa de fibra óptica está directamente en contacto con la cámara CCD que permite la detección de la luz generada cuando tiene lugar la reacción de pirosecuenciación. El orden e intensidad de los picos de luz se graban en un *flowgrama* (figura 27) que revela la secuencia de ADN²¹¹.

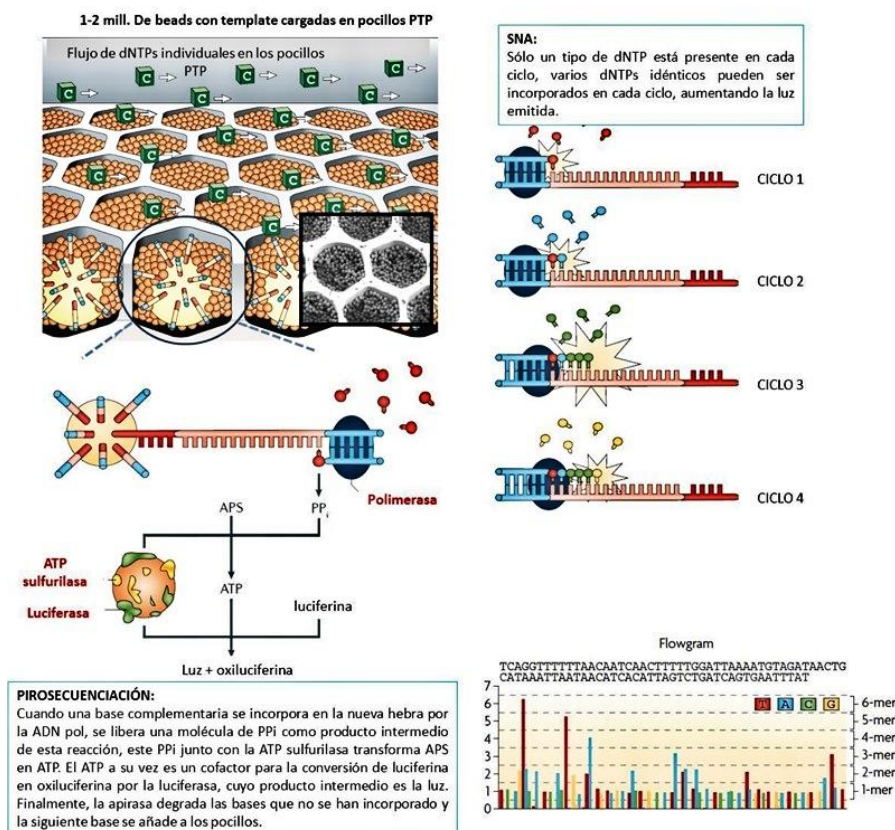


Figura 27 Pirosecuenciación ²⁰⁸

4. Análisis de la imagen. Cada flujo en cada pocillo resulta de la incorporación de uno, dos, tres o más nt, la intensidad de señal, calibrada a partir de la secuencia “llave” (TCAG) (figura 27), se usa para calibrar la señal que indica directamente el número de nt incorporados, esta linealidad en la señal existe hasta al menos homopolímeros de ocho nt. La plataforma 454TM dispone de un software (Amplicon Variant Analyzer, AVA, Roche) que le permite convertir estos datos en lecturas en un proceso conocido como asignación de base (*base calling*), los parámetros de cada imagen (intensidad de señal, ruido de fondo), se analizan por el software que aplica una serie de pasos de normalización, corrección y filtrado de la señal obtenida y la transforma en una gráfica de flujos (*flowgrama*) (figura 27), el cual da lugar a una lectura de secuencia. Pero a veces se requiere de softwares más avanzados para la eliminación de bases de baja calidad y filtrado de los errores. Un proceso muy efectivo en reducir la tasa de error es aumentar la cobertura, es decir, cada lectura de secuencia se alinea a una secuencia de referencia con un bajo nivel de exigencia para detectar mutaciones y variaciones genómicas, una vez que este alineamiento se ha llevado a cabo, se hace una media de todas las señales de varias lecturas que corresponden a la misma localización, y se identifica cada base individual²¹⁰.

Secuenciación por síntesis de Illumina: Esta tecnología de secuenciación masiva que utiliza la plataforma MiSeq, es la que hemos utilizado para la realización del segundo estudio incluido en esta tesis. Por lo tanto, igual que en el apartado anterior nos detendremos más exhaustivamente en su descripción. La tecnología de Illumina se basa en el enfoque de SBS por CTR, por lo tanto, consiste en la incorporación de nucleótidos marcados con terminadores de cadena reversibles de manera que, en cada ciclo de ligación, solamente uno de los cuatro nucleótidos se une de forma complementaria al ADN molde, emitiendo una señal luminosa que es captada por un sistema óptico altamente sensible. Posteriormente, el terminador es eliminado para permitir la incorporación del siguiente nucleótido y así ir elongando la cadena complementaria a la secuenciada en ciclos sucesivos de secuenciación.

La obtención de las lecturas de las secuencias de las moléculas de ADN con la tecnología Illumina es un proceso dividido en diferentes pasos que se resumen a continuación:

1. Creación de una biblioteca de amplicones de la región de interés: Al igual que la secuenciación por UDPS, el primer paso del procedimiento de secuenciación es delimitar la región de la que interesa determinar la secuencia, teniendo en cuenta la limitación de la longitud máxima de la plataforma. Para ello se preparan las bibliotecas de amplicones con los *primers* que se muestra en la Figura 28. Las bibliotecas obtenidas para cada muestra individual se agrupan por *pools*, en un número variable, en función del número de amplicones y de su longitud. Dentro de cada *pool* de las bibliotecas, se encuentran los amplicones de cada muestra individual que previamente fueron identificados con una secuencia MID. La optimización del protocolo de secuenciación de estos amplicones nos ha llevado a normalizar la concentración de los amplicones de estos *pools* a 8nM. Para comprobar la verdadera concentración de nuestro *pool* de las bibliotecas de amplicones, se cuantifica mediante real-time qPCR con el kit Kapa library quantificacion Kit, esta determinación se realizará en el LightCycle (Roche).

En concreto en nuestro estudio se han utilizado cebadores personalizados por lo tanto las bibliotecas obtenidas se han preparado para el sistema MiSeq mediante el *custom TruSeq-style*, por lo que es necesario ligar en los extremos de los amplicones de las bibliotecas oligonucleótidos-indexados que identificaran cada *pool* de bibliotecas de amplicones. Para ello se deben transformar los extremos cohesivos de los amplicones, que contienen la secuencia del oligonucleótido-universal de la figura XX, en extremos romos, para que así puedan ligarse a los oligonucleótidos-indexados. Esto se ha realizado con la ayuda del Kapa Hyper Prep Kit (Roche) y acto seguido se procede a la ligación de los

oligonucleótidos-indexados mediante PCR que se ha realizado con Nimblegene SeqCap Adapter Kit (Roche).

Forward: 5'-	TruSeq Adapter, Index 1-	MIDCCCATCTGTTCCCTCCCTGTC	- Secuencia específica template-3'
Reverse: 5'-	TruSeq Adapter, Index 1-	MIDCCTATCCCCTGTTGCGTGTC	- Secuencia específica template-3'

Figura 28: Primers utilizados para crear las bibliotecas de amplicones. Los adaptadores usados en la tecnología SBS de Illumina son primers de 90 nt añadidos a cada extremo de los templates (secuencias molde, de las que nos interesa determinar la secuencia), los primeros 60 pb en el extremo 5' son los oligonucleótidos-indexados que permiten la identificación de cada pool de bibliotecas (amarillo). Estos oligonucleótidos-indexados están compuestos por secuencias de 60 nt, y están marcadas con *index*, que están formados por 6 nt. Actualmente hay dos versiones del kit, cada una con 12 adaptadores diferentes que incorporan etiquetas de *index* únicas, lo que permite multiplexar diferentes muestras en la misma secuencia de ejecución. Después está la secuencia MID (*Multiplex Identifier*). Un identificador único de 10 nt establecidos por el fabricante y que se utiliza durante el análisis bioinformático para identificar la muestra a la que pertenece cada lectura obtenida, el cual va unido al M13 (secuencia de 20 nt que permite unir esta secuencia con el template (azul)). En el extremo 3' está la secuencia específica del template.

2. Formación y cuantificación del pool de pools. A continuación, todos los *pools* de bibliotecas amplificadas se unen en un solo pool y se cuantifica de nuevo la concentración de este nuevo *pool de pools* mediante real-time qPCR con el kit Kapa library quantification Kit, esta determinación se realizará en el LightCycle (Roche) y se comprobará dicha cuantificación mediante la realización de diferentes diluciones del pool inicial con una segunda real-time qPCR. Una vez cuantificado se llevará el pool a una concentración de 4nM y lo desnaturizamos mediante NaOH 0.2 N y mediante calor. Para aumentar artificialmente la diversidad de las secuencias determinadas en cada experimento de secuenciación, se mezclará el pool con una biblioteca de control de ADN genómico del fago phi (PhiX). Esto es necesario para no saturar la señal obtenida en los diferentes ciclos de secuenciación debido a la secuenciación de bibliotecas de “baja diversidad”, con lo que se evitan problemas de enfoque y eliminación de fases²¹²²⁰⁸.

El siguiente paso es cargar este *pool de pools* final con nuestros amplicones en la plataforma MiSeq, para ello seguiremos la siguiente figura (figura 29):

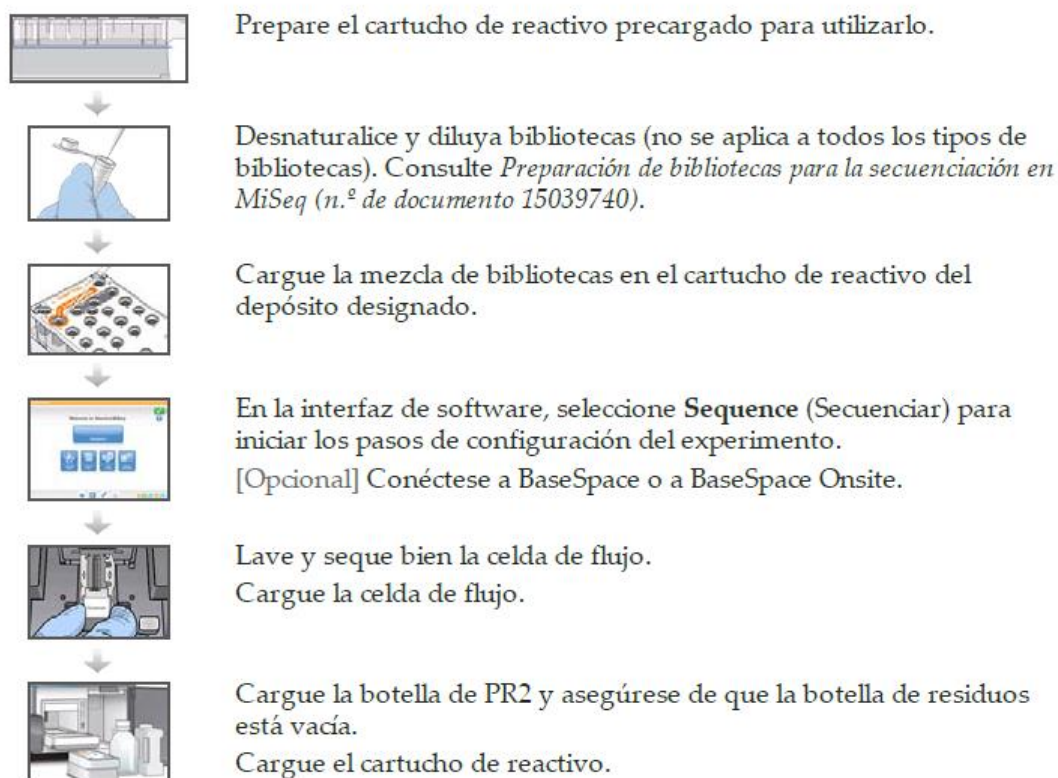


Figura 29 Flujo de trabajo en la plataforma Illumina MiSeq²¹³.

3. Generación de clústers. Ya dentro de la plataforma MiSeq con la muestra, los reactivos y la celda de flujo (*78mul cell*) cargados, el proceso de secuenciación comienza con el paso de amplificación de ADN mediante la generación de puentes a partir de los amplicones de las bibliotecas. Estos amplicones se unirán a sondas fijadas en la superficie de la celda de flujo por complementariedad de los oligonucleótidos índice de sus extremos con estas sondas tal como se muestra en la figura 30. La adhesión de estos amplicones es necesaria para llevar a cabo el proceso de amplificación mediante el empleo del ADN polimerasa, encargada de producir múltiples copias de los fragmentos de ADN que forman los puentes (*bridge PCR*), dando lugar a clústers derivados de la molécula que inició la amplificación (figura 30).

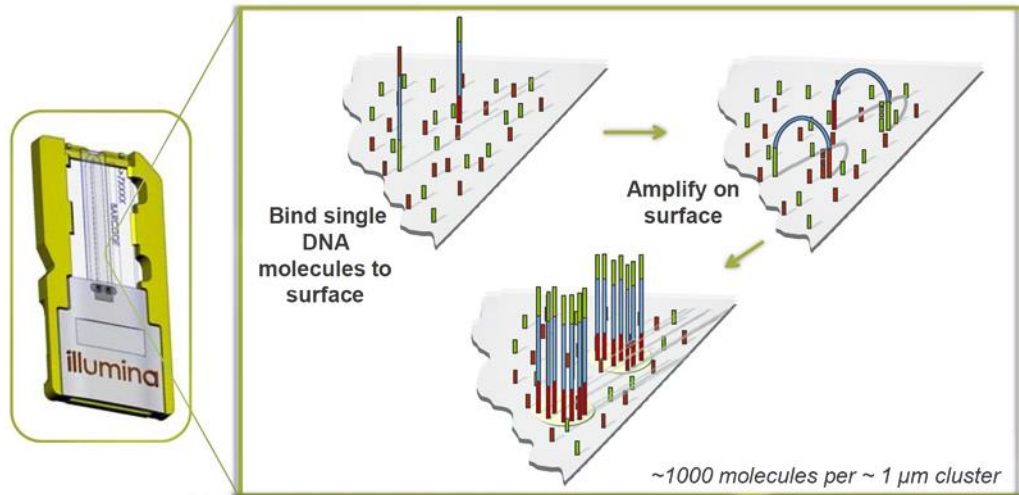


Figura 30 Secuenciación por síntesis de illumina. Los fragmentos monocatenarios de ADN se unen al interior de la superficie de los canales que forman la célula de flujo. Se añaden nucleótidos no marcados y enzima para iniciar la amplificación en puente (*bridge PCR*) y la desnaturalización de las cadenas dobles²¹¹.

4. Secuenciación por síntesis: Cada clúster contiene aproximadamente un millón de copias del fragmento original, suficiente para emitir una señal con la intensidad óptima para que sea detectada cada vez que se incorpora un nucleótido durante el proceso de secuenciación. La secuenciación por síntesis requiere la incorporación de una mezcla de los cuatro nucleótidos en cada canal de la célula de flujo junto con la ADN polimerasa y los *primers*.

La clave de este sistema de secuenciación, basado en CTR, es que los nucleótidos se encuentran modificados con un fluoróforo reversible en el extremo 3'-OH. Esto hace que estén químicamente bloqueados, impidiendo que otro nucleótido se añada a la nueva cadena hasta que este no sea eliminado (figura 31). Al final de cada ciclo se realiza un proceso de captación de imagen, en el que se identifica el nucleótido que se ha agregado en el ciclo actual en cada una de las celdas de la placa. A continuación, se elimina el fluoróforo de los nucleótidos y se comienza el ciclo desde el principio, continuando la síntesis de la hebra de ADN y la secuenciación.

El número de ciclos dependerá de la longitud del fragmento que el usuario necesita e influirá en el tiempo de secuenciación y en la cobertura media final.

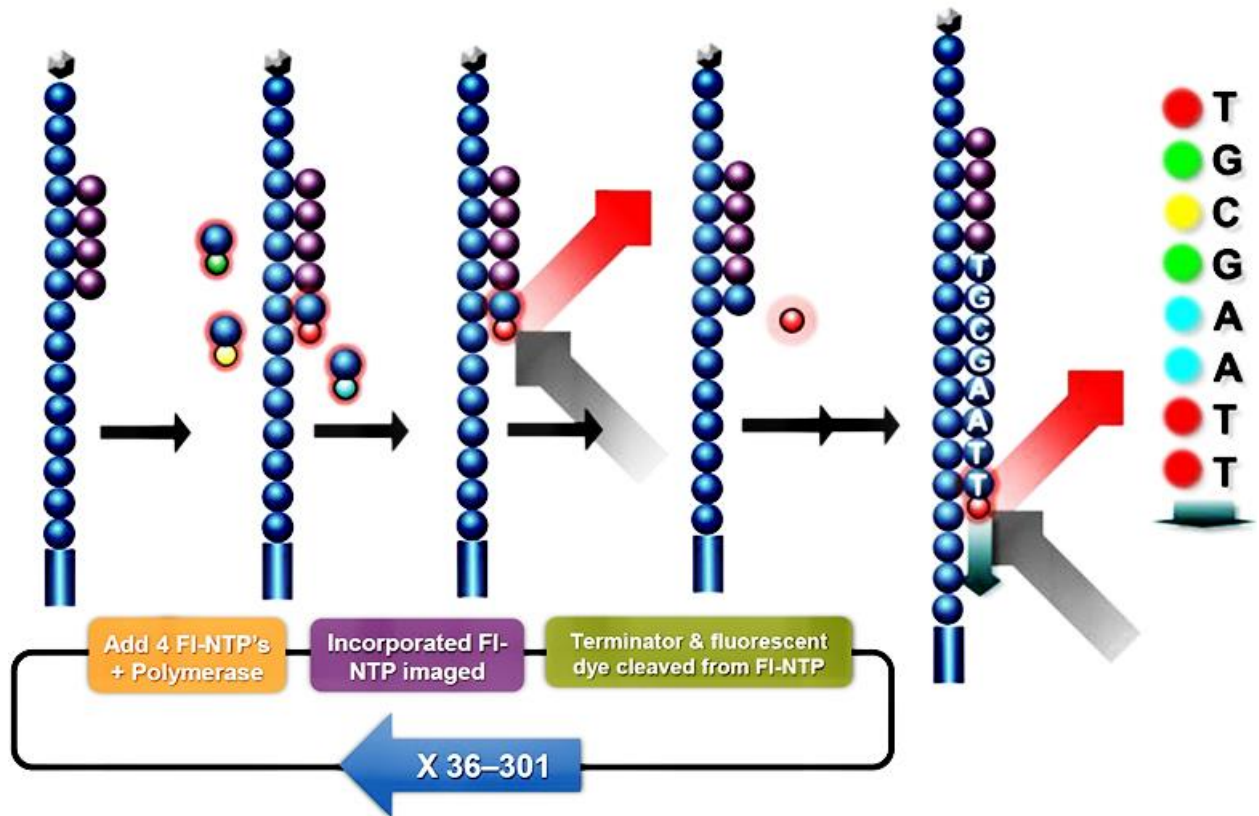


Figura 31 Para iniciar el primer ciclo de secuenciación, se añade a la célula de flujo: los cuatro terminadores reversibles marcados, los primers y la ADN polimerasa. Tras la excitación con láser, se captura la imagen de fluorescencia emitida por cada uno de los clústers de la célula de flujo. Se identifica y se graba la primera base de cada clúster. Antes de comenzar el siguiente ciclo químico, se eliminan los terminadores 3' bloqueados y el fluoróforo de cada base incorporada. La lectura de la secuencia se realiza durante múltiples ciclos químicos. Se repiten ciclos de secuenciación para determinar la secuencia de bases de un fragmento dado²¹¹.

5. . **Análisis de los datos:** El enfoque de la secuenciación Illumina se basa en la lectura simultánea de grandes cantidades de secuencias. El muestreo en profundidad y una cobertura uniforme entre las diferentes bibliotecas de amplicones asegura una alta precisión en la determinación de diferencias genéticas. Cada lectura de base tiene una puntuación asignada según la escala de calidad *phred score* por lo que el software de Illumina puede hacer diferencias y generar puntuaciones seguras²⁰⁸.

Un vídeo explicativo de cómo funciona esta plataforma puede observarse en <https://www.youtube.com/watch?v=womKfikWlxM>

Ion Torrent, (SNA): La secuenciación de Ion 80mulsió utiliza un enfoque de secuenciación por síntesis basada en SNA, parecido al de la UDPS. Al igual que en ésta, la amplificación clonal tiene lugar por emPCR, sin embargo, esta plataforma utiliza un

chip semiconductor que detecta los iones de hidrógeno producidos durante la polimerización del ADN. Puede secuenciar fragmentos de hasta 400 pares de bases.

El Ion Torrent es la primera técnica comercial para no usar fluorescencia y escaneo de cámara; por lo tanto, es más rápido y barato que muchos de los otros métodos. El gran problema de esta técnica es que el cambio de pH detectado por el sensor es poco sensible al número de nucleótidos detectado, lo que conduce a una precisión limitada en la medición longitudes de homopolímero²¹⁴.

2.2.3.2 NGS de fragmentos largos (secuenciación de tercera generación):

Existen dos tipos de tecnologías para secuenciar fragmentos largos, por un lado, las basadas en secuenciación a tiempo real de moléculas única (SMRT) y, por otro lado, las tecnologías basadas en lecturas largas formados de forma sintética (SRL).

SMRT, single-MolecularReal-Time:

Secuenciación de Pacific Biosciences (PacBio): La secuenciación se realiza sobre un chip que contiene muchas guía de onda de “modo cero” (ZMW). Dentro de cada una de ellas se encuentran fijadas una ADN polimerasa activa y una molécula de ADN molde de cadena sencilla. La ZMW permite que la luz penetre y se cree un espacio de visualización que permite monitorizar la actividad de la ADN polimerasa. La señal procedente de la incorporación de un nucleótido por parte de la ADN polimerasa se detecta durante la síntesis de ADN, de ahí que se trate de una secuenciación a tiempo real²¹⁵.

MinION from Oxford Nanopore Technologies (ONT): a diferencia de otras plataformas, los secuenciadores nanopore no monitorean incorporaciones o hibridaciones de nucleótidos guiados por una plantilla de cadena de ADN. Si no que detecta directamente la composición del ADN, para ello el ADN es forzado a pasar a través de un poro de proteína lo que provoca cambios en el voltaje, los cuales son característicos de cada nt²¹⁶.

SRL, Synthetic Long-Reads:

Illumina. Genomic DNA templates are fragmented: es una tecnología innovadora para generar lecturas largas que combina un nuevo método de preparación de bibliotecas y herramientas de análisis genómico. La tecnología divide el ADN en fragmentos más pequeños que se secuencian en las plataformas de secuenciación estándar de Illumina

para su posterior ensamblaje en lecturas de fragmentos largo utilizando un sistema informático patentado²⁰⁸.

10X Genomics' 82mulsión-based sequencing: esta innovadora tecnología 10X Genomics es una plataforma que se basa en la creación de bibliotecas de fragmentos cortos etiquetados con más de 10^6 códigos de barra para su posterior unión, todo esto es llevado a cabo en el interior de una emulsión. Estas lecturas vinculadas se pueden utilizar para el estudio del genoma completo, la detección de variantes estructurales y el ensamblaje del genoma de novo²¹⁷

2.2.4 Secuenciación de cuarta generación.

En esta década estamos entrando en el comienzo de una nueva fase de análisis genómico en la que la secuenciación masiva se realiza directamente sobre la célula. Recientemente se han descrito dos métodos para la secuenciación de ARN in situ ISS (In situ sequencing), uno dirigido y otro no dirigido, que se basan en la química de la ligadura. Esta cuarta generación de tecnología de secuenciación abre perspectivas para el análisis transcriptómico, la validación de biomarcadores, el diagnóstico y la estratificación del paciente²¹⁸.

2.2.5 Aplicaciones y limitaciones de NGS

La capacidad de secuenciar clonalmente y en paralelo la secuencia de miles o incluso millones de moléculas de ADN de las técnicas de NGS le da una gran ventaja a la hora de analizar mezclas complejas de moléculas de ADN en un solo experimento de secuenciación. Por este motivo, las plataformas de NGS han sido utilizadas con éxito en el análisis de quasispecies del VHB en varios estudios, sobre todo para cuantificar las variantes minoritarias del VHB portadoras de mutaciones de resistencia al tratamiento antiviral, de escape al tratamiento antiviral, o para estudiar su complejidad²¹⁹⁻²²⁶. Además, el enfoque de la NGS ha permitido el estudio dinámico de la evolución clonal en células cancerosas, detectando mutaciones somáticas en subclones raros a una tasa de 1 en $5 \cdot 10^3$ copias²²⁷. Las tecnologías de NGS son de alto rendimiento y permiten generar de 10^3 a 10^4 veces más información que la secuenciación de primera generación (Sanger), a un coste bajo para megabase secuenciada²²⁸ y con un ahorro considerable de tiempo y trabajo. Sin embargo, varios factores deben ser recordados en relación con estas tecnologías. En primer lugar, la longitud de una lectura de secuencia con las plataformas de NGS más comunes hoy en día es más corto que el de la secuencia de Sanger (aunque

es mucho más largo en el caso de las plataformas de tercera generación) y, en segundo lugar, la plataforma NGS tiene un modelo de error único que difiere del modelo de error establecido para lecturas de secuencia capilar. En consecuencia, algoritmos complejos y precisos son necesarios para procesar los datos brutos de NGS, y deberían ser específicos para cada plataforma. El filtrado bioinformático ha sido el principal método desarrollado para validar variantes que llevan sustituciones (es decir, distinguir entre variantes artefactuales del sistema de secuenciación y variantes reales de la quasispecies analizada), lo cual es especialmente importante en regiones homopoliméricas, donde la tasa de error de la mayoría de las plataformas de NGS es mayor^{207,229}. Otro factor que debe tenerse en cuenta en el análisis de datos de NGS es la posibilidad de recombinación que ocurren durante la amplificación con PCR²³⁰ por diferentes mecanismos; por ejemplo, amplicones cortos incompletos que pueden actuar como cebadores para diferentes secuencias presentes en el quasispecies, o los cruce de secuencias durante su amplificación, que pueden ocurrir de manera similar a los eventos de recombinación cromosómica. Estos fenómenos pueden ser una limitación potencial de las NGS, pero también pueden ocurrir en clonación clásica y cualquier otra técnica que incluye un paso de amplificación con PCR. Por último, el límite de detección de un haplotipo determinado por NGS no se ha establecido, y las diferencias entre los estudios de diferentes grupos deben tenerse en cuenta a la hora de analizar los resultados.

2.3 Genotipos y subgenotipos: impacto en la biología y la patogénesis.

Actualmente se estima que el genoma del VHB evoluciona con una tasa de error de entre 10^{-4} a 10^{-5} nucleótidos (substituciones/tamaño/año)¹⁸¹. Esta elevada tasa de evolución hace posible la aparición y fijación de mutaciones en diferentes poblaciones humanas repartidas a lo largo del mundo. Por consiguiente, los genomas del VHB se pueden clasificar filogenéticamente según su divergencia genética en grupos llamados genotipos. Actualmente, se considera que los diferentes genotipos del VHB poseen una divergencia intergrupala $> 7,5\%$ del genoma completo y existen 9 genotipos, A-I²³¹, además de un supuesto décimo genotipo, 'J', que tan solo ha sido aislado en un individuo²³¹. Dentro del mismo genotipo, cuando esta diferencia intergrupala se encuentra entre el 4 y 8% del genoma completo nos conduce a la diferenciación en subgenotipos de los genotipos A-D, F, H e I (no se han encontrado subgenotipos en los genotipos G y H). Se conocen al menos 35 subgenotipos¹⁸² (figura 32).

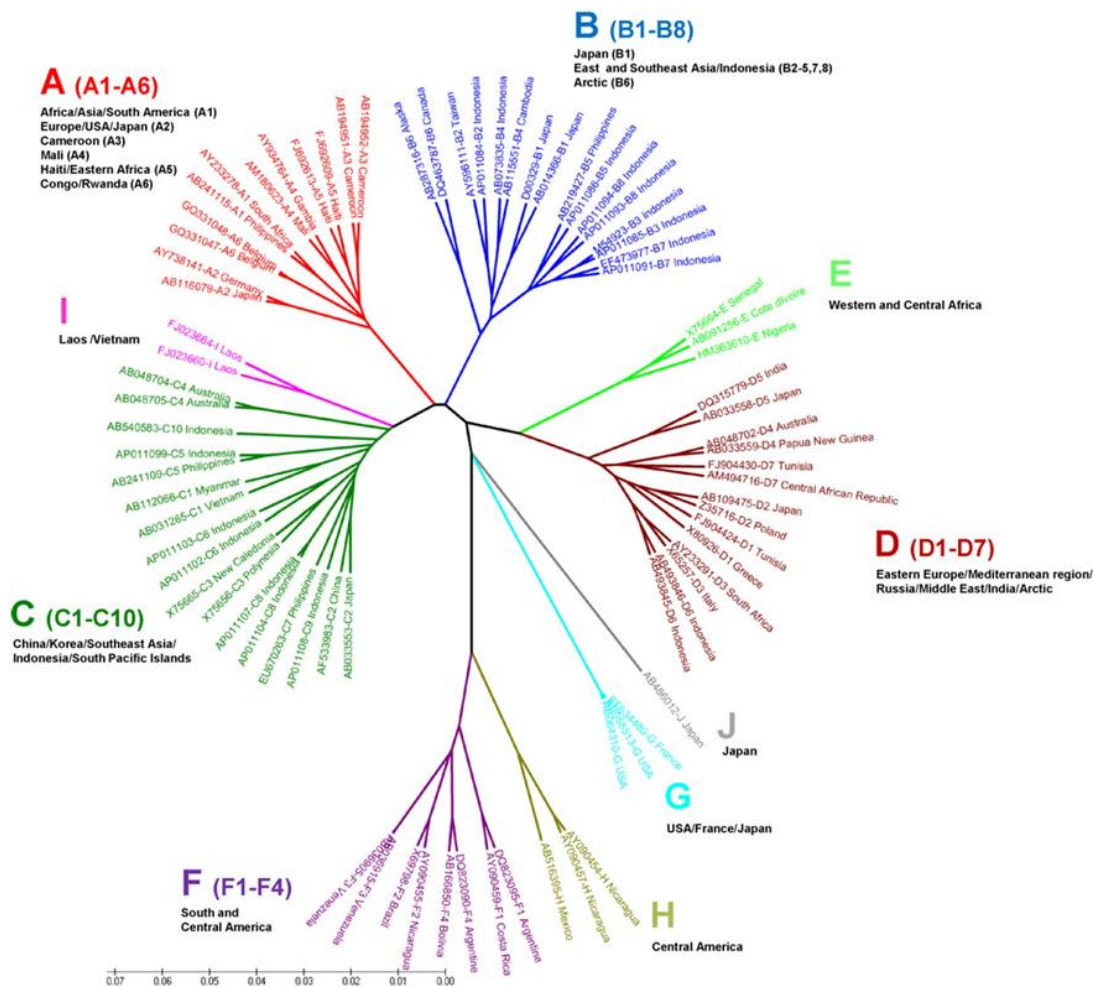


Figura 32. Árbol filogenético, en el que se muestran las relaciones filogenéticas de los diferentes genotipos y subgenotipos²³²

Los genotipos difieren en la longitud del genoma, en el tamaño de los marcos de lectura abiertos, y en algunos polimorfismos de las proteínas traducidas²³³, así como en el desarrollo de diversas mutaciones¹⁸². Con anterioridad a los genotipos, las poblaciones de VHB de diferentes orígenes se clasificaban basándose en la heterogeneidad inmunitaria de su HBsAg²³³, en función de sus epítomos “a”, “d/y” y “w/r”, que posteriormente se subclasificaron en “w1-4” y “q” han sido identificados 10 subtipos serológicos: ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw3, adw4, adr y adrq. Actualmente se conoce que existe una correlación estadísticamente significativa entre subtipos serológicos y genotipos: adw está asociado con genotipos A, B, F, G y H, adr con C, y ayw con D y E, pero existen muchas excepciones (tabla 4)²³³.

Genotipo	Subtipo serológico
A	ayw1, adw2, ayw2, adw4*
B	ayw1, adw2, adr*
C	adr, adrq, ayw2, ayw3, adw2, ayr*
D	ayw2, ayw3, ayw4*, adw3*
E	ayw4
F	adw4, adw2*, ayw4*
G	adw2
H	adw4
I	adw2, ayw2
"J"	ayw3

Tabla 4 Correlación entre subtipos serológicos y genotipos. *subtipo serológico raro para el genotipo

El método *gold 85standard* para el genotipado del VHB es el análisis filogenético de las secuencias de nucleótidos de todo el genoma del VHB²³⁴. Este tiene como inconveniente que es relativamente caro y consiste en un proceso lento y laborioso, por lo que el análisis filogenético se suele realizar en genes individuales o fragmentos del genoma viral^{235,236}. Además, también se pueden utilizar métodos indirectos como los explicados en el Capítulo 1 apartado 2.2.1 *Métodos indirectos*, que, aunque son menos fiables que la secuenciación también es más beneficiosos en cuanto a coste y tiempo de trabajo. Algunos de los métodos indirectos más usados son la hibridación reversa por INNO-LiPA, uno de los métodos más ventajosos en lo que a coste-beneficio se refiere²³⁷. También se ha usado el análisis de fragmentos de restricción de longitud polimórfica que depende de la amplificación por PCR (PCR-RFLP)²³⁸, entre otros.

2.3.1 Distribución geográfica

En la figura 33 y en la tabla 5, se muestra la distribución de los genotipos en todo el mundo. La identificación del genotipo del VHB es importante por muchas razones ya que existen claras asociaciones entre el genotipo y la progresión de la enfermedad o la eficacia del tratamiento con interferon- α (IFN- α) en pacientes con CHB²³⁹. En algunos casos la distribución de genotipos/subgenotipos está relacionada con el origen étnico, por ejemplo B6 está confinado a poblaciones indígenas del Ártico canadiense, C4 a los aborígenes australianos y C3 y C5-C10 a población indígena de Indonesia, por eso se ha propuesto que los (sub)genotipos han evolucionado en diferentes corrientes de migraciones humanas²³⁵. No obstante, la distribución geográfica de los genotipos se está

alterando debido a la movilidad de las poblaciones humanas por viajes o migraciones, como se demuestra por la aparición de genotipos europeos, asiáticos y africanos en países de baja endemicidad como Australia²⁴⁰

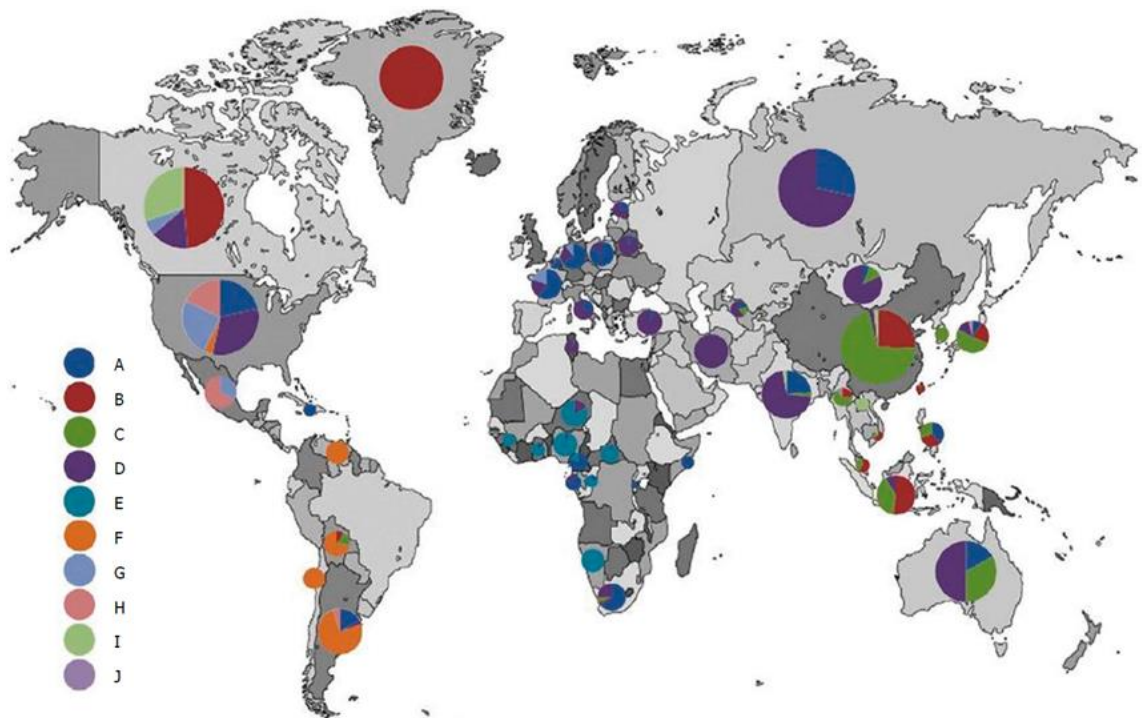


Figura 33 Distribución geográfica de los genotipos del VHB²⁴¹

CAPÍTULO 1. INTRODUCCIÓN

País	Genotipo	Subgenotipo
China	B, C	B2, C1, C2
Indonesia	C, B	C1, B3, B7, C10, B9 Y C8
Túnez	D, F	-
Turquía	D,	D2, D1, D3
Brasil	A, F	A1, F2a, A2, F4
Vietnam	B, C, I	B2-5, C5-16
Taiwán	B	B2, B5
Sur Corea	C	-
Hong Kong	C, B	-
Gambia, Nigeria, Haití, Congo, Ruanda y Camerún	A	A4, A5, A6, A7
Japón	A, C	C1, C2, C3
Filipinas	A, B, C	A1, B5, C5
India	A, C, D	-
África central	A, D, E	A1, D4
Arabia Saudí	D, E	D1
Irán	D	D1
Mongolia	D	-
Sudáfrica	D	D3
Tailandia	C, B	C1-5
Italia	D	-
Marruecos	D, A	D1, D7, A2
Argentina	F	F1, F2, F4
Egipto	D	D1
Paquistán	D	-
Australia	C, D	C4, D4
España	A, D, F	-

Tabla 5 Distribución de genotipos y subgenotipos en los países²⁴¹ .

2.3.2 Recombinación intergenotípica.

La recombinación en el VHB es el resultado de la coinfección de un huésped con más de una población viral de diferentes genotipos o subgenotipos, que pueden intercambiar su material genético dentro de las células hospedadoras, al igual que pasa entre cromosomas homólogos. Este fenómeno se ve favorecido en determinadas regiones geográficas sobre todo por 3 condiciones: que en una misma región geográfica se encuentren dos o más genotipos diferentes; que en esa región haya una elevada tasa de cronicidad de la infección y que el nivel de salud pública sea bajo. Así, la

CAPÍTULO 1. INTRODUCCIÓN

mayoría de las poblaciones recombinantes han sido reportadas en el este de Asia o África, donde la prevalencia de infección por el VHB es alta y la profilaxis y el control de la infección es bajo. Debido a la alta densidad de población, la posibilidad de co-infección o sobre-infección con distintos (sub)genotipos se potencia, por lo que el riesgo de recombinación es elevado. Por el contrario, los aislados recombinantes del VHB raramente han sido reportados dentro de las típicas cepas europeas (como los subgenotipos A2, D2 y D3). Esto corrobora la necesidad de un control de salud eficiente y medidas profilácticas para disminuir el riesgo de infección, coinfección y finalmente recombinación.

En todo el mundo se han documentado nuevas variantes generadas por recombinación entre diferentes genotipos, por ejemplo, el gen preCore/Core de B2-B5 tiene su origen en el genotipo C²⁴², el VHB dominante en el Tíbet es un recombinante C/D²⁴³, el genotipo I es un recombinante A/C/G y se han identificado recombinantes A/D, G/C y D/E en África²⁴⁴. También hay evidencias de recombinación entre genotipos de VHB humanos y VHB de simios, como en el genotipo J. Además el análisis a gran escala de secuencias completas en bases de datos ha evidenciado incongruencias en árboles filogenéticos construidos desde diferentes regiones del genoma del VHB, que han llevado a la detección de nuevos genomas recombinantes²³².

2.3.3 Diferencias genéticas, clínicas y evolutivas entre genotipos del VHB.

2.3.3.1 *Tamaño del genotipo y heterogeneidad genética:*

El elevado nivel de divergencia en las secuencias de nt en las variantes de los diferentes genotipos, es producto de un largo proceso evolutivo y adaptativo. Durante este proceso se han acumulado inserciones y deleciones polimórficas en el genoma de los distintos genotipos, lo que provoca que tengan diferentes longitudes: al genotipo del VHB prototípico podríamos asignarle una longitud de 3215 nt, pero el genotipo A tiene 3221 nt, el del D 3182 nt (este es el más pequeño de todos los genotipos del VHB descritos) i el del G 3248 (genotipo más grande)^{240,245}. Estas diferencias de tamaño polimórficas entre los diferentes genotipos del VHB se resumen en la tabla 6.

CAPÍTULO 1. INTRODUCCIÓN

Genotipo	Genoma (pb)
A	3221 (ins de 6 nt cerca del extremo 3' gen core)
B	3215
C	3215
D	3182 (del de 33 nt en extremo 5' preS1)
E	3212 (del de 3 nt en extremo 5' preS1)
F	3215
G	3248 (ins de 36 nt en extremo 5' gen core y del de 3 nt en 5' preS1)
H	3215
I	3215
J	3182 (del de 33 nt en 5' preS1)

Tabla 6: Longitud del genoma de los genotipos del VHB.

Además de estas diferencias de longitud, los diferentes genotipos también presentan diferencias en su heterogeneidad genética lo que hace que los genotipos A, B, C, D, F y I se hayan subclasificado en diferentes subgenotipos y que esta subclasificación no se haya producido en los genotipos E, G, y J²⁴⁶. En un análisis evolutivo para caracterizar el origen del VHB en humanos, se ha estimado que el tiempo desde el primer ancestro común entre los genotipos A, B, C, D y F se sitúa entre 8.9-12.7 10^3 de años atrás, en cambio el ancestro común de los genotipos E, G y H es relativamente reciente ya que tan solo tiene entre 0.7-6 10^3 de años²⁴⁷. Esto sugiere que los genotipos más recientes y menos heterogéneos podrían haber derivado a partir de los genotipos más antiguos y heterogéneos, una prueba de esto es que en un análisis filogenético podemos observar que los genotipos F y H están estrechamente relacionados, al igual que pasa entre los genotipos D y E (figura 32). Esto nos puede hacer pensar que los genotipos que aún no se han dividido en subgenotipos no han tenido tiempo suficiente para poder evolucionar. Sin embargo existen diferentes teorías acerca del origen del VHB en humanos, que apuntan a que la diversificación de éste en distintos genotipos podría situarse mucho antes en el tiempo, en paralelo a la diversificación de los primeros homínidos, alrededor de 3 10^7 de años atrás²⁴⁸.

2.3.3.2 Ruta de transmisión:

Los genotipos B y C se transmiten principalmente por la ruta perinatal, mientras que los genotipos A y D generalmente se adquieren por transmisión horizontal en la edad adulta (subgenotipo A2) o en la adolescencia (genotipo D). En el caso del genotipo G suele ocurrir por contacto sexual en hombres que tienen sexo con hombres²⁴⁰. No obstante, este aspecto no está del todo claro ya que estas son las vías de transmisión habituales en las regiones donde dichos genotipos son

CAPÍTULO 1. INTRODUCCIÓN

mayoritarios, por lo que no se puede confirmar que el genotipo sea el responsable de la vía de transmisión *per se* o simplemente que esté utilizando la vía más común en el área correspondiente.

2.3.3.3 Implicación del genotipo en la patogénesis.

Los genotipos e incluso los subgenotipos del VHB pueden ser responsables de diferencias en la historia natural de la infección crónica por VHB¹⁸² y, por lo tanto, pueden desempeñar un papel en la manifestación clínica de la infección¹⁸². Ejemplos de esto los vemos a nivel de la seroconversión del HBeAg; los pacientes infectados con los genotipos A, B, D y F muestran una seroconversión espontánea más temprana y más frecuente en comparación con aquellos infectados con el genotipo C, independientemente de la etnia¹⁸². Los pacientes infectados con genotipo E tienen una mayor frecuencia de positividad de HBeAg y cargas virales más altas en comparación con los pacientes infectados con genotipo D²⁴⁹. Los individuos infectados con el subgenotipo A1 pierden HBeAg mucho antes que aquellos infectados con el subgenotipo A2²³⁹. Aunque no se requiere HBeAg para la replicación viral, se ha demostrado que juega un papel en la transmisión del VHB, la persistencia viral y es esencial en el establecimiento de la hepatitis B crónica¹⁸². Los genotipos y subgenotipos también muestran diferencias en cuanto a la aparición de complicaciones graves de la infección; A1, C, B2-B4, F1 muestran un mayor riesgo de desarrollar cirrosis y HCC en comparación con A2, B1 y B5²⁵⁰. Las infecciones por genotipo C conllevan un riesgo más temprano de inflamación del hígado, fibrosis hepática, cirrosis y HCC que aquellas infectadas con otros genotipos²³⁵. En la India, el subgenotipo D1 está significativamente asociado con la enfermedad hepática crónica y D3 con infección oculta por el VHB²³⁹. Este tipo de infección se caracteriza por bajas cargas virales y también es común en pacientes infectados con genotipo H, en los que el desarrollo de HCC es poco frecuente²³⁹. Los habitantes de Alaska infectados con genotipo C o F son más propensos a volver a la positividad para HBeAg y, por lo tanto, tienen un mayor riesgo de desarrollar enfermedad hepática grave que aquellos infectados con genotipos A, B y D²³⁵. Finalmente, en Sudán, se encuentra que el genotipo E prevalece en donantes de sangre asintomáticos¹⁸², mientras que el genotipo D predomina en pacientes con enfermedad hepática¹⁸².

2.3.3.4 Implicación del genotipo en el tratamiento

Los genotipos del VHB también influyen en la respuesta a diferentes tratamientos antivirales. En pacientes con genotipo C tratados con IFN α estándar se define una respuesta sostenida a este tratamiento como la seroconversión del HBeAg en pacientes HBeAg positivos (pérdida de la detección del HBeAg y detección de sus anticuerpos específicos, anti-HBe), normalización de los niveles séricos de ALT (principal marcador bioquímico de lesión hepática) y niveles indetectables

CAPÍTULO 1. INTRODUCCIÓN

de ADN-VHB seis meses después de finalizar el tratamiento²⁵¹. Así, los niveles de respuesta sostenida a este tratamiento son significativamente mejores en pacientes infectados por genotipo A que en pacientes infectados por genotipo D²⁵². Además, tras el tratamiento con IFN- α también se ha observado una mayor pérdida sostenida del HBeAg y HBsAg en pacientes con genotipo A que en pacientes con otros genotipos²⁵³. En el caso del tratamiento con ANs, el subgenotipo B2 se asocia más frecuentemente a la selección de variantes con resistencias a lamivudina que el B1 y en algunos estudios se ha observado que el genotipo A selecciona variantes con resistencias a este tratamiento antiviral más rápido que el D²³⁵.

2.3.3.5 Impacto de los genotipos en las medidas profilácticas.

La vacuna actual es un péptido recombinante súper antigénico del HBsAg que cubre el determinante “a”. Se considera que un título de anti-HBs > 10 UI/l es protector y se desarrolla en el 95-99% de los niños sanos y adultos jóvenes²⁵⁴. A pesar del éxito en la mayoría de los casos, la vacunación a veces falla debido al genotipo. Las mutaciones de escape inmune en el gen S varían entre genotipos, de hecho, algunas de estas mutaciones de escape están en la secuencia original (*wild-type*) en algunos genotipos²³⁵. Por ejemplo, se ha descrito un caso en Europa en el que dos pacientes desarrollaron infección por el genotipo F a pesar de tener anti-HBs y no presentar mutaciones de escape inmune²³⁵.

2.4 Principales variantes genéticas de interés clínico

Además de las grandes diferencias entre poblaciones virales de diferentes orígenes (diferentes genotipos y subgenotipos) también juegan un papel crítico, en la modulación de la patogenia en la infección por el VHB las variantes genéticas (equivalentes a haplotipos) con determinadas sustituciones nucleotídicas y/o aminoacídicas o con inserciones / deleciones. Los cambios en la secuencia de ciertas regiones del genoma del VHB pueden influir en las tasas de seroconversión de HBeAg y HBcAg, en los niveles de viremia, en el escape inmune, en la aparición de mutantes, en la patogenia de la enfermedad hepática, y en la respuesta y resistencia a la terapia antiviral, y en la vacunación contra el virus (figura 34).

CAPÍTULO 1. INTRODUCCIÓN

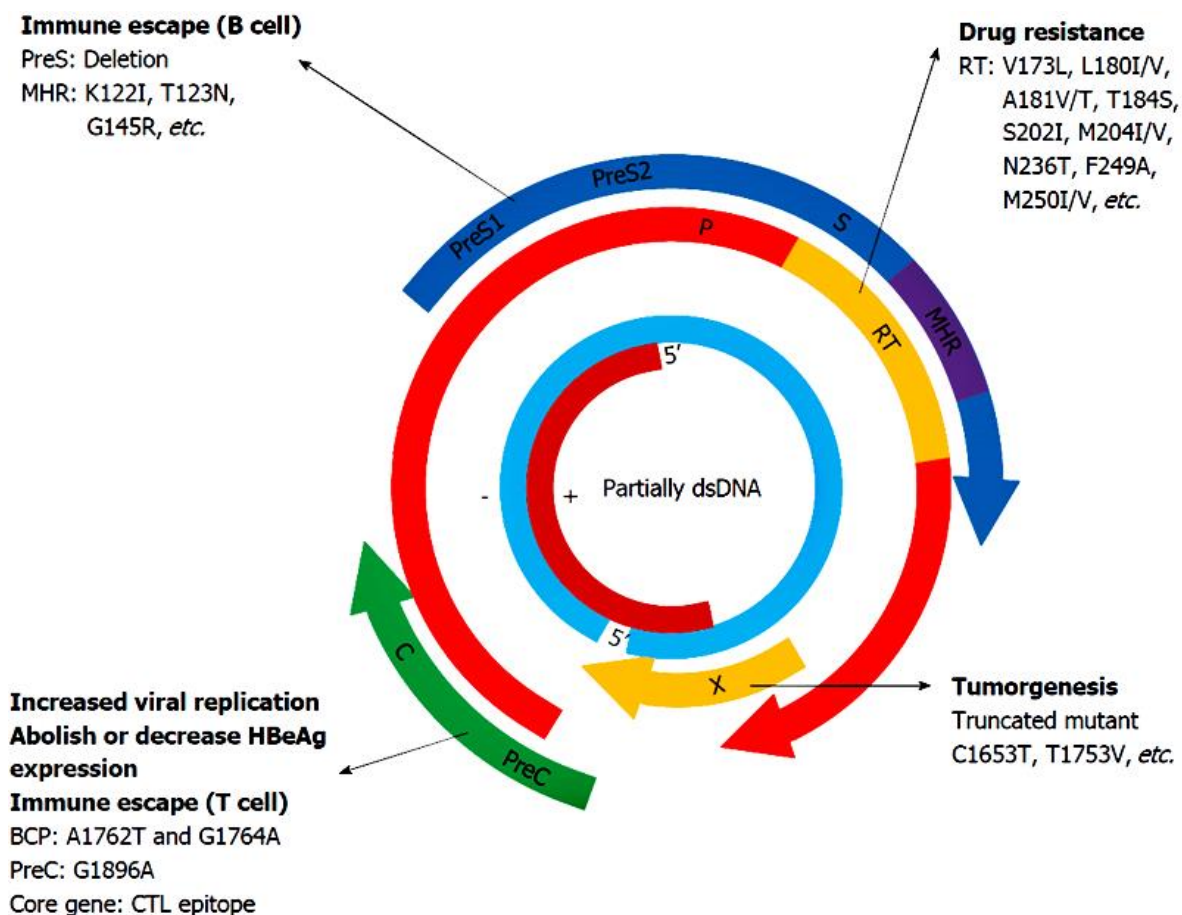


Figura 34 Principales efectos de los cambios en los diferentes ORFs del VHB y su significado clínico ²⁵⁵

2.4.1 Variabilidad en el ORF preS/S

Los ORF pre-S1 / S2 / S codifican tres proteínas de la envuelta (LHBs, MHBs y SHBs) que son determinantes para el ensamblaje del virus y la unión del virus a los hepatocitos (ver en el Capítulo 1 el apartado 1.4.1.3 ORF S y las proteínas de superficie). Estas 3 proteínas forman en conjunto el HBsAg, que contiene el principal epítipo de células B, el determinante "a" (aa 124-147) dentro de la MHR, que abarca desde los aa 99-169²⁵⁶. Mutaciones en la MHR, particularmente en el determinante "a" (contra el que se dirigen los anticuerpos contra el HBsAg o anti-HBs), se asocian con el escape inmune debido cambios conformacionales en epítipos. El cambio de aa G145R es la principal mutación de escape de respuesta inmune inducida por vacunación²⁵⁷ y se ha encontrado en niños que adquieren la infección por el VHB a pesar de la inmunoprofilaxis activa y pasiva²⁵⁵. Otras mutaciones de escape inmune son T116N, P120S/E, I/T126A/N/I/S, Q129H/R, M133L, K141E, P142S y D144A/E²⁵⁷. Hay estudios en los que se observa que, durante inmunosupresión,

CAPÍTULO 1. INTRODUCCIÓN

algunos pacientes, con resolución de la infección, muestran reactivación del VHB lo que en algunos casos pueden conducir a una hepatitis aguda grave, insuficiencia hepática fulminante y muerte. Hay estudios que han demostrado que un alto porcentaje de los pacientes reactivados con VHB fueron portadores de más de una mutación de HBsAg, la mayoría de las cuales fueron localizadas en el MHR M103I, L109I, T118K, P120A, Y134H, S143L, D144E y S171F. Otra pequeña cantidad también se detectó en epítomos de las células T pertenecientes a la clase C48G, V96A, L175S, G185E, V190A²⁵⁸.

En los últimos años, la infección oculta por el VHB (OBI) ha sido ampliamente investigada y se han detectado algunas mutaciones del HBsAg que se correlacionan con OBI²⁵⁷, como por ejemplo el cambio nucleotídico C695T conduce a un codón de parada en el aminoácido 181 esto podría ser responsable de la fuerte reducción de producción de HBsAg. También los codones stop prematuros en la región S parecen participar en el desarrollo de OBI, ya que pueden influir en la expresión, síntesis y secreción de HBsAg²⁵⁷. La selección de variantes debido al tratamiento antiviral con AN también puede dar lugar a la selección de variantes con codones stop prematuros o afinidad disminuida para anti-HBs debido al solapamiento entre los ORF S y P, como se explicará en el apartado siguiente (ver en Capítulo 1 el apartado 2.4.2 *Variabilidad en el ORF P*).

Se ha postulado que las mutaciones en las regiones preS1 y preS2 podrían contribuir a la hepatocarcinogénesis, una de ellas es la mutación W4P / R en preS1, que puede estar asociada con la gravedad de la enfermedad en pacientes masculinos crónicamente infectados con el genotipo C de VHB. Estos mutantes W4P / R se relacionaron significativamente con HCC y cirrosis hepática²⁵⁹. Sin embargo, las principales mutaciones de la región preS asociadas al desarrollo de HCC son las deleciones. Al menos tres mecanismos podrían estar implicados en la patogénesis del HCC asociado a estas deleciones en preS: una respuesta inmune alterada, ya que las regiones preS1 y preS2 contienen varios epítomos para las células B y células T²⁵⁷, la acumulación de LHBs en el RE, que desencadena estrés del RE, y los mutantes preS2 los cuales pueden activar directamente las vías promotoras de tumores como las vías VEGF/AKT/mTOR, p27/retinoblastoma/Cdk2/cyclin A, D, entre otras²⁵⁷.

2.4.2 *Variabilidad en el ORF P*

El ORF P codifica la polimerasa viral, que es una poliproteína multifuncional, implicada en la encapsidación, retrotranscripción y posterior degradación del ARNpg para formar la hebra (-) del genoma viral y finalmente en la síntesis de la hebra (+) del genoma viral a partir de la (-). Los medicamentos actualmente disponibles, aprobados para tratamiento de CHB en muchas partes del mundo, son un inmunomodulador (INF- α o peginterferón (Peg-INF) y 6 agentes antivirales que

CAPÍTULO 1. INTRODUCCIÓN

pertenecen a los ANs: lamivudina (LAM), telbivudina (LdT), entecavir (ETV); adefovir (ADV) y tenofovir disoproxil fumarato (TDF)²⁶⁰. El sexto agente antiviral es una nueva formulación del tenofovir (*Tenofovir Alafenamide* o TAF)²⁶¹. Los AN son inhibidores del dominio RT de la polimerasa y el tratamiento con estos fármacos es generalmente eficiente y bien tolerado. Sin embargo, la resistencia es un problema importante en algunos de estos agentes que afecta a la terapia a largo plazo. La resistencia está asociada a la selección de mutaciones adaptativas que reducen la sensibilidad de los mutantes a los efectos inhibidores del tratamiento²⁶⁰ las principales mutaciones se resumen en la tabla 7²⁶². Como se ha explicado en el apartado anterior, algunas mutaciones seleccionadas en el ORF P afectan al ORF S solapante. Por ejemplo, el mutante rtA181T seleccionado por ADV, LAM o LdT resulta en un codón stop en el gen S (sW172stop) causando un defecto en la secreción de HBsAg y esta mutación se encuentra con frecuencia en HCC²⁵⁷, las mutaciones de resistencia rtV173L + rtL180M + rtM204V resultan en las mutaciones sE164D + sI195M en HBsAg que reducen la unión a anticuerpos²⁵⁵.

Patrón	Sustituciones de aminoácidos en el dominio rt	LMV	LdT	ETV	ADV	TFV
	Wild-type	S	S	S	S	S
L-Nucleosido (LMV/LdT)	M204I/V	R	R	I	S	S
Facido fosfónico (ADV)	N236T	S	S	S	R	I
Compartido (LMV, LdT, ADV)	A181T/V	R	R	S	R	I
Doble (ADV, TFV)	A181T/V + N236T	R	R	S	R	R
D-Ciclopentano (ETV)	L180M+M204V/I ± I169 ± T184 ± S202 ± M250	R	R	R	S	S
Resistencia multiple	A181T+N236T+ M250V	R	R	R	R	R

Tabla 7. Patrones y vías de resistencia a los medicamentos antivirales en CHB en el contexto de la resistencia cruzada²⁶². *I, sensibilidad intermedia; R, resistente; S, sensible basado en cultivo celular y datos clínicos.*

2.4.3 Variabilidad en el ORF PC/Core

El gen del core del VHB se divide en dos regiones, la región PC y la región Core. Este ORF codifica para dos proteínas; el HBcAg, cuya principal función es formar la nucleocápside y el HBeAg que es una proteína de secreción cuya principal función es inmunomoduladora²⁵⁵.

El HBcAg del VHB es un potente estimulador del sistema inmune²⁶³, en el cual se han descrito múltiples regiones epitópicas para linfocitos B y T. Una alta conservación en la secuencia del HBcAg provoca una respuesta de las células T citotóxicas (CTL) más fuerte en comparación con otras proteínas virales. Estas células median una respuesta inmune multispecífica, policlonal y

CAPÍTULO 1. INTRODUCCIÓN

vigorosa durante la hepatitis B aguda y mínima o indetectable en CHB con persistencia viral y tolerancia inmune²⁶⁴. Así pues, las mutaciones en HBcAg pueden conducir a la producción de variantes de escape inmunológico no reconocidas por los CTL, lo que resulta en la persistencia del VHB²⁶⁵.

Se sabe que en el curso natural de la infección crónica por VHB, la pérdida de la expresión de HBeAg y la aparición de anticuerpos dirigidos contra HBeAg (anti-HBe) suelen asociarse al control inmunológico de la infección. Sin embargo, ciertas mutaciones en las regiones BCP y PC pueden dar lugar a la aparición de hepatitis B crónica HBeAg-negativa con presencia de anti-HBe, en las que continúa la replicación viral y el ADN del VHB permanece detectable en suero ($> 210^3$ UI / ml)²⁵⁵. La mutación de PC más prevalente es el cambio nucleotídico G1896A, que crea un codón de terminación TAG en el codón 28 de la proteína de PC (codificada por las regiones PC y core a través del ARN_{pc}) y elimina la expresión de HBeAg a nivel traducional²⁶⁶. Esta mutación se ha detectado tanto en estadios severos como en benignos^{219,267,268} y tiene una elevada prevalencia en el genotipo D, aunque también puede estar presente en los genotipos B y E. En cambio es raro en encontrarla en los genotipos A y H²⁶⁹.

El BCP en realidad se encuentra entre el extremo 3' del ORF X y el 5' del PC/Core, por lo tanto las mutaciones de esta secuencia reguladora pueden afectar al ORF X y viceversa, como se explicará en el siguiente apartado (Capítulo 1; 2.4.4 *Variabilidad en el ORF X*). La mutación más común en el BCP es el doble cambio de nucleótido A1762T y G1764A, lo que resulta en una disminución de hasta el 70% en la expresión de HBeAg²⁶⁶, ya que el BCP controla la síntesis del ARN_{pc} y el ARN_{pg}. Además de las mutaciones A1762T y G1764A, se han identificado otras mutaciones de BCP: el T1753C, y el C1766T. Básicamente, estas mutaciones reducen la síntesis de HBeAg y mejoraran las replicaciones virales en las células hepáticas, a menudo en asociación con manifestaciones más severas de la enfermedad y HCC²⁷⁰. Algunas de estas mutaciones (T1753C, A1762T y 1764A), junto con A1752G, A1846T, G1896A y G1899A, han sido correlacionadas significativamente con seroconversión del HBeAg²⁷¹.

2.4.4 *Variabilidad en el ORF X*

EL ORF X se solapa con la parte del ORF P que codifica el dominio RH en su extremo 5' y con el extremo 5' del ORF PC/Core en su extremo 3'. Este ORF codifica la HBx la cual, juega un papel crítico en la patogénesis de HCC y la replicación viral⁸⁶. La HBx promueve la carcinogénesis al interactuar con proteínas produciendo la desregulación de múltiples vías de señalización celular implicadas en la progresión del ciclo y crecimiento celular y en la apoptosis²⁵⁵. En pacientes con

CAPÍTULO 1. INTRODUCCIÓN

HCC a menudo se presentan formas de la HBx truncadas en su extremo C-terminal, como se ha reportado en varios estudios^{49,86,272-277}.

A continuación, se van a describir las mutaciones más comunes:

Mutaciones en ENHII y (o) en BCP, T1753V y C1653T:

Hay 3 mutaciones en la región donde se encuentran el ENHII y el BCP [una mutación en ENHII (H94Y: C → T de nt 1653) y dos en BCP (I127L, T, N, S: T → C/A de nt 1753, K130M y V131I: A → T de nt 1762 y G → A de nt 1764)] estos son "puntos calientes" mutacionales, independientemente del genotipo o distribuciones geográficas⁸⁶.

Los cambios de nucleótido A1762T / G1764A provocan los cambios de aa en la HBx K130M y V131I, que es la mutación más frecuentemente de esta proteína, identificada en un amplio número de estudios de asociación con el riesgo de HCC²⁷⁸⁻²⁸¹. La mutación causa una disminución en la expresión de HBeAg y mejora de la replicación del genoma viral, lo que contribuye a la progresión de la enfermedad hepática a través de aumento de la inflamación y la invasión viral^{282,283}. Además, puede contribuir a la hepatocarcinogénesis a través de reducción expresión p53, lo que lleva a una rápida e incontrolada proliferación celular²⁸¹. Las mutación BCP puede aumentar el riesgo de desarrollar HCC, particularmente en la población infectada con el genotipo C⁸⁶. Las otras dos mutaciones T1753V → I127T y C1653T → H94Y también están asociadas al desarrollo de HCC: la primera puede cambiar la afinidad de unión la HBx a BCL2, por lo tanto afecta la apoptosis inducida por la HBx²⁸⁴ y la segunda es un fuerte elemento de activación de EnhII / BCP; además de al desarrollo de HCC está asociado con la hepatitis fulminante y la exacerbación aguda de HCC⁸⁶.

Mutaciones en el dominio de regulación negativa de la HBx (aa 1-50) (A1383C, G1386A / C-V5M / L, C1485T-P38S):

Se ha asociado a HCC el cambio nucleotídico A1383T, que no causa un cambio de aminoácido en la HBx. Esta asociación fue hecha por primera vez en una cohorte coreana²⁸⁵ y es una de las mutaciones de la HBx identificadas como factores de riesgo en el genotipo C para el desarrollo del HCC⁸⁶. Por otro lado, el cambio G1386A / C que conduce a un cambio de aminoácidos de V5M/L es una mutación específica de HCC y posiblemente juegue un papel fundamental en la progresión de la cirrosis hepática a HCC⁸⁶. La mutación C1485T también está significativamente relacionada con HCC⁸⁶.

Inserciones/deleciones (indel) en la región C-terminal de la HBx:

Dado su papel transactivador, la región C-terminal de la HBx juega un papel clave en controlar la proliferación celular, la viabilidad y la transformación⁸⁶. Por lo tanto, las deleciones o inserciones en el C-terminal reducen la actividad de transactivación y los efectos inhibidores sobre la proliferación celular, por lo tanto, pueden contribuir a la generación de HCC⁸⁶. Tanto las deleciones como las inserción en el extremo C-terminal son una de las mutaciones reportadas con mayor frecuencia de la HBx y ha sido frecuentemente detectado en tejidos y muestras de suero de pacientes con HCC, independientemente de su genotipo o ubicación geográfica²⁸⁶⁻²⁸⁹. Se han reportado múltiples variantes por indel, algunas de las más relevantes son: Delección en 1763-1770²⁷⁸, Inserción en 1824-1825 (2nt) que provoca un cambio de aa en la posición, 176 (HBx más largo)²⁷⁴, Inserción en 1825 TA²⁷⁵ y delección en 1825²⁷⁵.

3. Epidemiología

La infección por el VHB es un importante problema de salud mundial, se estima que unos 2 10⁹ de personas presentan evidencias de haber estado en contacto en el pasado o el presente con el VHB y 257 millones son portadores crónicos del HBsAg. En función de la prevalencia de la enfermedad, existen regiones geográficas con endemidad alta, donde la prevalencia de portadores de HBsAg es igual o mayor del 8 % y en las que la infección se adquiere con frecuencia a edades tempranas, por transmisión perinatal (de madre a hijo); regiones con endemidad intermedia alta e intermedia baja (prevalencia de portadores de HBsAg entre el 2 y el 7 %), donde la transmisión horizontal precoz (habitualmente entre miembros de una misma familia o personas que conviven) es más frecuente; y regiones con endemidad baja (menos del 2 % de HBsAg), donde la enfermedad se adquiere predominantemente en la edad adulta y la transmisión está relacionada con prácticas de riesgo y se produce principalmente en personas con múltiples parejas sexuales y usuarios de drogas por vía intravenosa (figura 35)²⁹⁰.

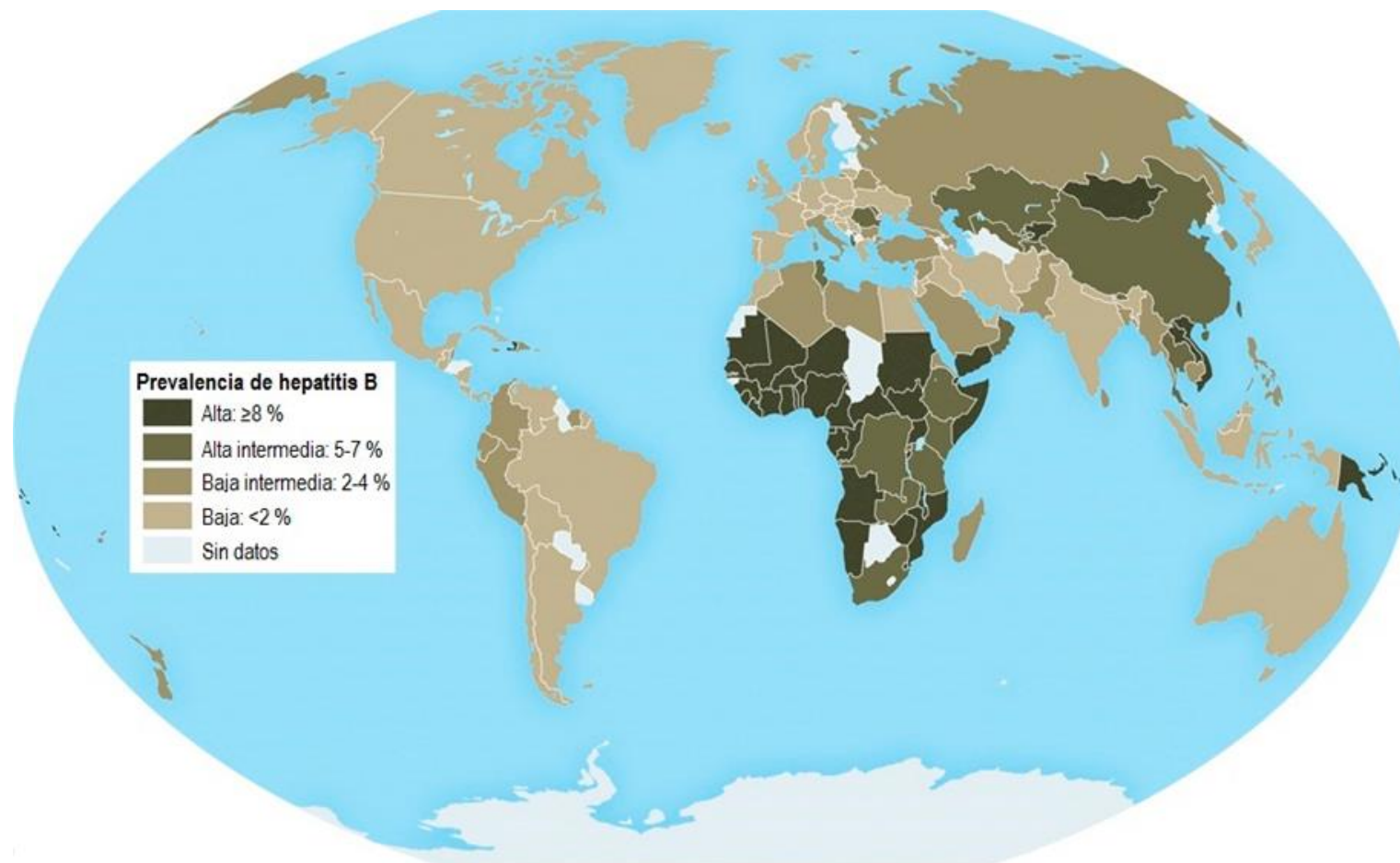


Figura 35 Prevalencia de la hepatitis B en el mundo²⁹¹

CAPÍTULO 1. INTRODUCCIÓN

En España, con una prevalencia de portadores < al 2 %, se considera un país de prevalencia baja²⁹². Anualmente se comunican alrededor de 700 nuevos casos de infecciones producidas por el VHB, en concreto en el 2017 se declararon a la Red Nacional de Vigilancia Epidemiológica (RENAVE) 645 casos²⁹² en la declaración agregada, lo que supone una tasa bruta de incidencia nacional de 1,48 casos/10⁵ habitantes. El periodo de mayor riesgo para la adquisición de la infección en nuestro país es al final de la adolescencia y los primeros años de la edad adulta. Gracias a la introducción de la vacuna contra el VHB en el sistema nacional de salud en la década de los 90, la incidencia de la infección por el VHB ha disminuido notablemente. La transmisión vertical es extremadamente rara debido a la baja incidencia y al cribado serológico sistemático de todas las gestantes para el VHB. En las mujeres embarazadas españolas se estima una prevalencia de HBsAg del 0,7-2 %, con algunas diferencias geográficas. En los últimos 10 años se han declarado tan solo 15 casos de hepatitis B en niños menores de 1 año, siendo 7 de ellos emigrantes²⁹¹.

En los últimos años la prevalencia de la infección por el VHB ha variado en algunas áreas geográficas, apreciando una disminución en los países desarrollados. Las causas principales son la vacunación universal y el establecimiento de medidas preventivas sanitarias. Otros cambios se producen por los movimientos migratorios provenientes de países en vías de desarrollo, confirmándose la mayor prevalencia del VHB entre la población inmigrante²⁹³. Estos cambios, producidos por los movimientos migratorios se han comprobado por el estudio de los distintos genotipos, modificándose las prevalencias de estos en áreas geográficas específicas.

3.1 Mecanismos de transmisión

El único reservorio del VHB humano es el hombre. La vía de transmisión es por contacto percutáneo o permucoso. La mayor concentración del virus en el organismo se encuentra en la sangre, aunque su presencia en otros líquidos biológicos como saliva, semen, secreciones vaginales y orina puede explicar ciertos contagios. El conocimiento de las principales formas de transmisión del VHB es crucial para desarrollar medidas de profilaxis primaria:

Vía parenteral: ocurre a través de agujas, productos sanguíneos contaminados, tatuajes o acupuntura. Actualmente esta vía está disminuyendo debido a una mejoría en los hábitos higiénicos, incluyendo el menor intercambio de jeringuillas entre adictos a

estupefacientes por vía parenteral y la generalización de técnicas de esterilización de material sanitario²⁹⁴.

Vía sexual: se trata de la vía de transmisión más frecuente en países desarrollados, siendo mayor el riesgo entre homosexuales y heterosexuales con conductas de riesgo. Las campañas de vacunación y medidas como el uso del preservativo disminuyen las posibilidades de contagio²⁹⁴.

Transmisión vertical o perinatal: este término se aplica a la transmisión de madre a hijo durante el parto y constituye el mecanismo principal en zonas de alta prevalencia. El riesgo aumenta en caso de que la madre sea positiva para HBsAg, HBeAg y ADN-VHB. En países en los que se realiza el cribado sistemático a la madre, la vacunación del recién nacido y el tratamiento con gammaglobulina anti-HBs, puede minimizar en gran medida esta transmisión²⁹⁴.

Transmisión horizontal por contacto no sexual: se produce a partir de objetos en los que el virus puede permanecer estable hasta 7 días, como por ejemplo el uso compartido de cepillos de dientes, material sanitario, etc.²⁹⁴.

3.2 Historia natural de la infección

El virus de la hepatitis B es capaz de producir enfermedad hepática aguda y crónica. El mecanismo fisiopatológico del daño hepático no es citopático directo, sino que es debido a la respuesta inmune celular contra los hepatocitos infectados y esa misma explicación se tiene para las manifestaciones extrahepáticas que en ocasiones acompañan el cuadro agudo^{295,296}. Esto ha sido confirmado por el seguimiento a pacientes con infección crónica HBeAg negativa o portadores asintomáticos, quienes, a pesar de mostrar niveles altos de replicación viral intrahepática, conservan las pruebas de función hepática normales, y por el hecho de que en los neonatos infectados, su sistema inmune inmaduro causa una mínima lesión hepática a pesar de la alta replicación del virus²⁹⁷. Sin embargo, hay que tener en cuenta que, a largo plazo, la infección crónica por el VHB contribuye al desarrollo del HCC a través de diferentes mecanismos, como la integración del ADN del VHB en el genoma del huésped y la expresión prolongada de la HBx y / o las versiones alteradas de las proteínas de la envuelta en la región preS / S. Esto hace que los HCC relacionados con el VHB puedan surgir incluso en hígados no cirróticos, lo que apoya aún más la idea de que el VHB desempeña un papel directo en la transformación hepática hacia HCC²⁹⁸.

3.2.1 Infección aguda

Las hepatitis B agudas normalmente son patologías auto limitadas en más del 95% de los adultos inmunocompetentes²⁹⁹, y se caracterizan por una inflamación hepática aguda y necrosis hepatocelular. Son fatales entre el 0,5 y el 1% de los casos²⁹³. Su periodo de incubación puede durar entre 30 y 180 días (figura 36). El primer marcador serológico en aparecer es el HBsAg, seguido poco después del anticuerpo contra el HBcAg, seguido por el anti-HBc, el cual es predominantemente del tipo IgM²⁹⁹. El HBeAg también puede ser detectado en la mayoría de los pacientes con infección aguda. Los niveles circulantes del ADN-VHB son altísimos, con valores que pueden llegar entre $2 \cdot 10^9$ - $1 \cdot 10^{15}$ UI/mL²⁹⁹, lo cual lo convierte en un virus con una capacidad de transmisión mucho mayor que la de otros virus como el virus de la hepatitis C y el VIH. Los niveles de ALT comienzan a aumentar cuando se establece la replicación viral, como resultado de la respuesta inmune contra los hepatocitos infectados. La forma aguda de la hepatitis por VHB se resuelve generalmente de forma espontánea entre 4 a 8 semanas sin necesidad de terapia. Inicialmente el cuadro suele ser atípico con síntomas inespecíficos como astenia, febrícula, anorexia, malestar general, dolor abdominal, náuseas, vómitos, cefalea, fotofobia, mialgias, artralgias, faringitis, tos y en ocasiones, coriza. Los pródromos pueden preceder en 1 ó 2 semanas a la fase ictericia de la hepatitis y persistir posteriormente hasta la resolución del cuadro agudo. De un 5% al 10% de los pacientes presentan manifestaciones extrahepáticas similares a la enfermedad del suero, con artralgias o artritis, erupción cutánea, y alteraciones renales con hematuria y proteinuria; en algunos casos puede aparecer una vasculitis tipo poliartritis nudosa. El 95% de los pacientes adultos se recupera de forma espontánea, resolviéndose el cuadro clínico agudo y la ictericia en un periodo de 1 a 4 meses. La forma que no presenta ictericia es la presentación clínica más frecuente de la hepatitis aguda B (70%), mientras que tan solo el 30% desarrollan una hepatitis ictericia³⁰⁰.

La hepatitis B aguda grave se caracteriza por la prolongación del tiempo de protrombina y una hiperbilirrubinemia notable, pero sin presentar encefalopatía hepática, por lo cual no cumple criterios de fallo hepático agudo. El fallo hepático agudo por hepatitis B es una entidad rara, ocurre en el 0,1% a 1% de los casos de hepatitis B agudas y lo favorecen la coinfección o sobreinfección por el virus de hepatitis Delta (VHD), la infección por los mutantes preCore o la presencia de daño hepático previo por cualquier causa³⁰¹.

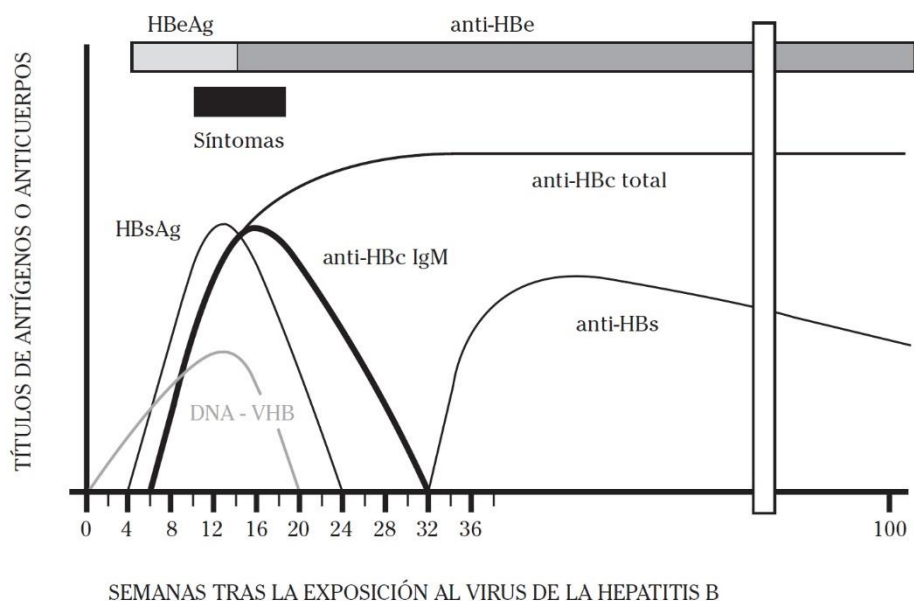


Figura 36 Marcadores serológicos de la infección aguda por VHB³⁰².

3.2.2 Infección crónica

La persistencia del HBsAg más de seis meses en suero define la progresión a hepatitis crónica²⁹⁹. Dicha progresión es inversamente proporcional a la edad de adquisición de la infección: un 90% si se trata de una infección perinatal, un 30% si se adquiere entre 1-5 años y hablamos de menos del 5% en el caso de los adultos inmunocompetentes³⁰³. Además de la edad de la infección, factores genéticos del hospedador como polimorfismos en el gen de la IL-18, el promotor del TNF- α y variantes en el HLA-DP se han asociado con la persistencia de la infección. Los síntomas en la infección crónica varían desde moderados y no específicos como fatiga y molestias en el cuadrante superior derecho, con daño hepático mínimo, hasta fibrosis progresiva que puede acabar dando lugar a descompensaciones hepáticas como ascitis, edema periférico y encefalopatía en pacientes con enfermedad hepática avanzada³⁰⁴. Además, como ya se ha comentado, la CHB aumenta en gran medida el riesgo de desarrollar HCC independientemente de la presencia de cirrosis²⁹⁸. La CHB es un proceso dinámico, que resulta de la interacción entre la replicación del virus y la respuesta inmunológica del hospedador, en la historia natural de la infección se pueden diferenciar cinco fases, no necesariamente secuenciales, y por las que no todos los pacientes atraviesan (figura 37)²⁹⁹. Las personas que desarrollan CHB presentan un HBeAg inicialmente positivo, acompañado de niveles altos de ADN VHB que pueden durar años o décadas. Eventualmente, el sistema inmune puede activarse y los pacientes pierden el HBeAg para desarrollar el anticuerpo anti-HBe²⁹⁶. En base a esto la nueva guía de tratamiento clínico *EASL 2017 Clinical Practice Guidelines*

*on the management of hepatitis B virus infection*³⁰⁵ se ha utilizado una nomenclatura nueva para referirse a las diferentes fases de la infección crónica por el VHB según si presenta HBeAg positivo o bien negativo (figura 38):

La Infección crónica por VHB con HBeAg positivo (anteriormente fase de inmunotolerancia) se caracteriza por tener el HBeAg en la sangre, una alta replicación viral (con niveles elevados de ADN-VHB), pero sin evidencia de necroinflamación hepática, es decir niveles normales de ALT (< 40 UI/mL) y fibrosis mínima o nula. El avance de la enfermedad es mínimo en las personas que se mantienen en esta fase, pero la posibilidad de infectar a otras personas es muy elevada. Esta fase es típica de la infección perinatal y estas personas suelen mantenerse estables hasta pasados 10-30 años. El genoma viral puede integrarse en el ADN del hepatocito, aumentando el riesgo en un futuro de desarrollar carcinoma hepatocelular³⁰³.

En la Hepatitis crónica por VHB con HBeAg positivo (anteriormente fase de aclaramiento viral) el HBeAg persiste, pero el ADN-VHB y la ALT son fluctuantes y se produce un aumento de la actividad histológica con progresión de la fibrosis. Esta fase aparece más rápidamente en personas infectadas en la edad adulta y su duración varía de semanas a años. En esta fase, también conocida como immunoactiva o de immunoeliminación, el sistema inmune reconoce el virus e intenta eliminarlo de las células hepáticas infectadas, y en muchos casos ocurre la conocida seroconversión del HBeAg a anti-HBe. Esto conlleva necroinflamación hepática de moderada a grave y fibrosis progresiva, por eso los pacientes que permanecen durante años en esta fase, sobre todo si presentan brotes repetidos de necrosis, tienen riesgo de desarrollar cirrosis y HCC³⁰³.

Después de la seroconversión a anti-HBe, la infección crónica por el VHB puede evolucionar hacia 2 fases o estados diferentes:

La Infección crónica por VHB con HBeAg negativo (anteriormente fase de portador inactivo) se caracteriza por una viremia muy baja (< 2 10³ UI /ml, aunque en algunos casos puede estar entre 2 10³-2 10⁴ UI/ml), ALT normal o con elevación leve. Estos niveles de ADN-VHB y ALT deben mantenerse así durante un año, determinándose cada 3 o 4 meses, para confirmar el paso a esta fase de la infección. El HBeAg es negativo y el anti-HBe es positivo. Esta fase de la infección por VHB tiene un buen pronóstico a largo plazo, con un riesgo bajo (aunque no nulo) de desarrollar cirrosis o HCC. En casos excepcionales (1-3%) se puede resolver la infección, desaparece el HBsAg y aparece el anticuerpo anti-HBs³⁰³.

La Hepatitis crónica por VHB con HBeAg negativo es la única fase de la historia natural de la hepatitis crónica por VHB cuyo nombre no ha cambiado en la nueva nomenclatura. Se caracteriza por una viremia $> 2 \cdot 10^3$ UI/ml y ALT elevada, aunque los niveles de estos dos marcadores suelen mostrar patrones fluctuantes. También se caracteriza por una hepatitis activa, lo que se evidencia por necroinflamación y fibrosis de moderada a grave a nivel histológico.³⁰³. Los pacientes en esta fase suelen tener variantes con mutaciones en la región PC del ORF C o en el BCP (ver en el Capítulo 1 en el apartado 3.2.2 *Variabilidad en el coreORF PC/Core*), lo cual hace que sus niveles de HBeAg sean bajos o inexistentes³⁰⁶. Así pues, los pacientes en esta fase de la historia natural de la infección crónica por VHB tienen una mayor probabilidad de progresar a cirrosis descompensada y HCC que los que presentan infección crónica.

La **fase de remisión** tras la pérdida de HBsAg se define por HBsAg negativo con o sin anti-HBs y se caracteriza por valores normales de ALT y valores indetectables de ADN-VHB en suero. Puede persistir baja replicación viral con ADN-VHB detectable en el hígado (OBI). Esta fase se asocia con mejoría del pronóstico si se llega a ella antes de desarrollar cirrosis. La relevancia clínica de la infección oculta no está bien definida³⁰³, sin embargo se ha reportado que en los pacientes que se encuentran en esta fase, la inmunosupresión puede dar lugar a la reactivación de la replicación viral³⁰⁷.

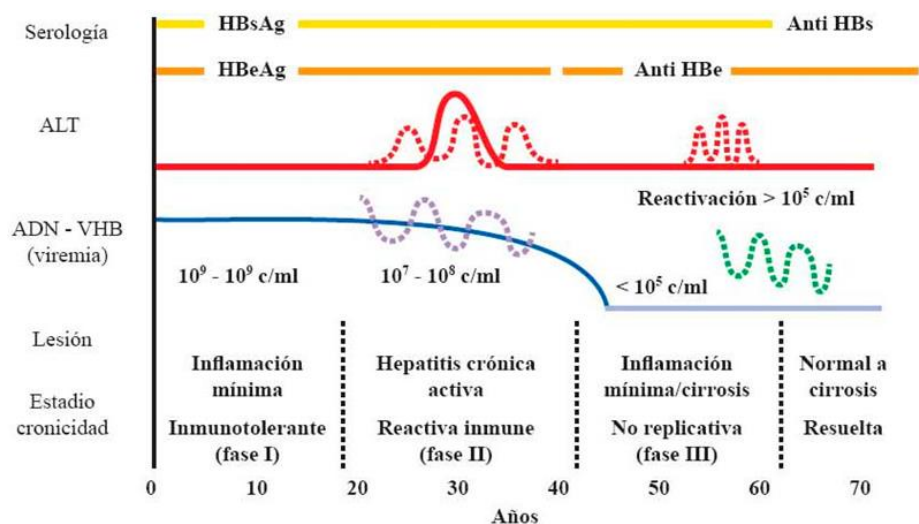


Figura 37 Fases de la infección crónica por VHB³⁰⁸. Fase I; Infección crónica por VHB con HBeAg positivo (anteriormente fase de inmunotolerancia), FaseII; Hepatitis crónica por VHB con HBeAg positivo (anteriormente fase de aclaramiento viral). Fase III La Infección crónica por VHB con HBeAg negativo (anteriormente fase de portador inactivo)

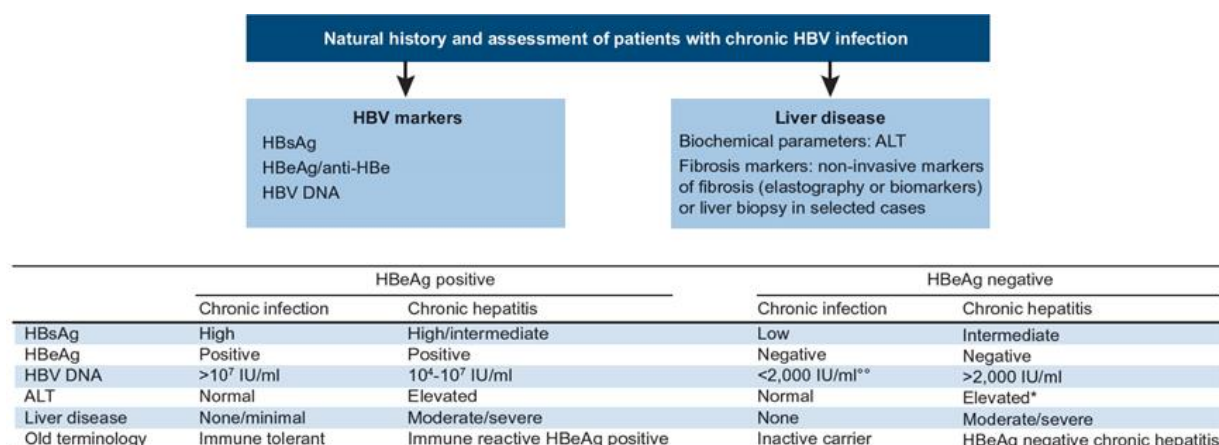


Figura 38 Fases de la historia natural y clasificación de pacientes con infección crónica por VHB basada en marcadores de enfermedad hepática y VHB según las nuevas guías de tratamiento clínico de la EASL. * Persistente o intermitentemente ADN del VHB los niveles pueden estar entre 2 10³y 2 10⁴ IU / ml³⁰⁵.

Esta clasificación de la infección crónica por el VHB en diferentes fases tiene importantes connotaciones desde el punto de vista pronóstico y terapéutico. En España predominan los pacientes con hepatitis crónica HBeAg negativo (un 75% de los casos), situación en la que la replicación viral es relativamente baja (niveles bajos y fluctuantes de viremia) y con un nivel poco elevado de las transaminasas. A veces, cuando estos pacientes HBeAg negativo pasan por una fase de baja replicación viral son casi indistinguibles de los portadores inactivos. Por el contrario, los pacientes con HBeAg positivo habitualmente suelen tener niveles circulantes de ADN-VHB relativamente elevados y más estables.

El riesgo de evolución hacia la cirrosis es del 2-5% personas/año en los pacientes HBeAg positivo y del 8-10% de las personas HBeAg negativo. La edad avanzada, los niveles elevados de viremia, la coinfección con otros virus como el virus de la hepatitis C (VHC), VHD o VIH, el abuso de alcohol, la obesidad, la presencia de fibrosis en el momento del diagnóstico y el grado de la actividad necro-inflamatoria (evaluadas en la biopsia hepática), son los principales factores que condicionan la evolución y progresión más o menos rápida de la enfermedad hepática hacia la cirrosis o HCC³⁰⁵.

3.2.3 Inmunopatogénesis del VHB

Como hemos señalado a lo largo de este trabajo el VHB no es directamente citopático a corto plazo: el daño hepático que se produce y el control viral están inmunomediados. El resultado clínico de la infección depende de la compleja interacción entre la replicación

del VHB y la respuesta inmune del huésped²⁹⁹. El VHB es un débil inductor de la respuesta inmune innata³⁰⁹. La resolución de la infección aguda se media principalmente a través de la respuesta inmune adaptativa; las personas con recuperación serológica de la infección aguda por el VHB tienen fuertes respuestas de linfocitos T a varios epítopos (policlonales) codificados en diferentes regiones del genoma del VHB. Por el contrario, los pacientes con CHB tienen respuestas débiles de los linfocitos T³⁰⁹. En particular, las células T CD8+ específicas del VHB a las que no se les ha presentado un antígeno viral con anterioridad, no logran diferenciarse en células de memoria efectoras citolíticas productoras de IFN- γ , es decir, muestran un fenotipo “agotado”³¹⁰.

La recuperación de la función inmune en los pacientes con CHB ha sido reportada en pacientes que presentaron una seroconversión de HBeAg espontánea y en aquellos que muestran respuestas virológicas a la terapia antiviral³¹¹, lo que sugiere que el tratamiento secuencial con la terapia antiviral seguida de la inmunoterapia podría aumentar la posibilidad de eliminación viral³¹¹. Las respuestas inmunes al VHB también pueden tener efectos negativos. Se piensa que la respuesta inmune demasiado agresiva es la causa de infección fulminante por VHB y de las exacerbaciones de la CHB, como son las concentraciones elevadas recurrentes de ALT, que aumentan el riesgo de desarrollar cirrosis y HCC. Además las células responsables de la respuesta inmunitaria son también responsables de las manifestaciones hepáticas y extrahepáticas, como la glomerulonefritis, la vasculitis y la poliartritis nudosa, provocadas por la acumulación de los inmunocomplejos circulantes³¹¹. Aunque algunos brotes se acompañan de caídas en las concentraciones del ADN-VHB y la posterior seroconversión de HBe (lo que sugiere un sistema inmunitario exitoso en la eliminación de hepatocitos infectados), otros brotes son acompañados de caídas transitorias en las concentraciones de ADN-VHB, pero sin seroconversión del HBe y por lo tanto representan aclaramiento inmune ineficaz³¹².

4. Tratamiento anti-VHB

4.1 ¿Qué se entiende por curación de la infección del VHB?

Los programas de prevención, la vacuna anti-VHB y los tratamientos disponibles (AN e IFN α), basados en la inhibición de la replicación viral, han conseguido disminuir el número de infectados y frenar la progresión de la enfermedad (fibrosis hepática y aparición de HCC). Sin embargo, aún no se ha encontrado una verdadera cura para erradicar la infección por el VHB. Esto se debe principalmente a que, una vez establecida

la infección, ningún tratamiento es capaz de acabar con la persistencia del ADNccc en los núcleos de hepatocitos infectados, lo que lo convierte en un reservorio natural del VHB.

Pero ¿Qué se entiende por curación de la infección del VHB? Pese a que se han organizado varios seminarios internacionales para intentar llegar a un acuerdo sobre la definición de lo que se debe considerar una verdadera cura de la infección^{313,314}, solo se ha podido acordar que una verdadera cura puede llegar a no ser factible debido a la persistencia del ADNccc en el hepatocito y a que el ADN del VHB se puede integrar en el genoma del huésped. Así pues, se han propuesto dos definiciones de cura de la infección por VHB para plantear objetivos más realistas para los nuevos tratamientos antivirales que se están desarrollando hoy en día.

1. **La cura total de la infección:** que viene definida por HBsAg indetectable en el suero y la erradicación del ADN del VHB, incluido el ADNccc intrahepático y el ADN integrado. Esta erradicación absoluta del VHB, en la actualidad se considera imposible³¹⁵.
2. **La curación funcional:** que viene definida por HBsAg y ADN indetectable en suero con o sin seroconversión a anticuerpos anti-HBs después de completar un ciclo finito de tratamiento. Esta definición no incluye la eliminación física del ADNccc o el integrado, pero sí su silenciamiento transcripcional, que aseguraría la respuesta sostenida al tratamiento aún después de interrumpirlo³¹⁵. Sería parecido a inducir la entrada de los pacientes en la fase de remisión. Este es el objetivo que se intenta conseguir con los nuevos tratamientos antivirales o con combinaciones de ellos.

4.2 Objetivo actual del tratamiento antiviral.

El objetivo principal de la terapia para pacientes con infección crónica por VHB es mejorar la supervivencia y la calidad de vida al prevenir la progresión de la enfermedad, y consecuentemente el desarrollo del HCC. Adicionalmente los objetivos de la terapia antiviral son prevenir la transmisión de madre a hijo, la reactivación de la hepatitis B y la prevención y el tratamiento de manifestaciones extrahepáticas asociadas al VHB³⁰⁵. Para alcanzar estos objetivos, es necesaria la supresión efectiva y sostenida de la replicación del VHB (conseguir una respuesta virológica sostenida), que reduce la actividad histológica de la hepatitis crónica y disminuye el riesgo de desarrollar HCC³¹⁶.

La probabilidad de lograr estos objetivos depende del momento en el que se administra la terapia antiviral durante el curso natural de la infección, pero también depende de la etapa de la enfermedad y de la edad de los pacientes en el comienzo del tratamiento. La regresión de la fibrosis y la cirrosis se puede considerar como un objetivo adicional de tratamiento en pacientes con avanzada fibrosis o cirrosis, aunque su impacto no ha sido completamente aclarado en resultados clínicos³⁰⁵.

4.3 Estrategias de tratamiento.

Actualmente, para el manejo de la CHB se cuenta con 2 tipos de terapias (una encargada de potenciar la respuesta inmune del huésped contra el virus y otra dirigida directamente contra el virus y su replicación):

Peg-IFN α -2a y α -2b: el interferón es una citoquina que tiene actividad antiviral directa e inmunomoduladora³¹⁷, su aplicación es subcutánea, una vía incómoda para los pacientes, y tiene muchos efectos adversos como por ejemplo síntomas generales, fatiga, neutropenia, trombocitopenia y depresión³¹⁸. Además, puede exacerbar enfermedades autoinmunes y está contraindicado en el fallo hepático, cirrosis avanzada o pacientes embarazadas. Por estos motivos, su uso está restringido a una población muy reducida³⁰⁵ y es una opción poco utilizada para tratar la infección crónica por VHB. Se considera como terapia la finita preferida en un grupo de pacientes seleccionados: estos deben de ser jóvenes, no cirróticos, con carga viral baja y marcada actividad inflamatoria. La tasa de seroconversión del HBeAg es del 29%-32% y la pérdida del HBsAg es de 3%-7%^{319,320}, unas tasas relativamente elevadas. Además, el Peg-IFN no presenta problemas de resistencia antiviral, ya que su actividad fundamental es inmunomoduladora³⁰⁵.

AN: inhiben la transcripción reversa del ARN pregenómico a ADNrc pero no tienen efecto directo sobre el ADNccc; es por esto que suprimen la replicación, pero no consiguen la curación total y rara vez la funcional de la infección. Por este motivo el tratamiento es a largo plazo y existe el riesgo de recaídas (reactivación de la replicación viral) después de su suspensión. Actualmente, hay 6 medicamentos aprobados en Europa: LAM, LdT, ETV, ADV, TDF y un nuevo tenofovir, *Tenofovir Alafenamide* (TAF)³⁰⁵. Todos ellos se administran por vía oral, tienen un efecto antiviral directo más o menos potente y en general son bien tolerados por los pacientes. Por estos motivos hoy en día los AN son los tratamientos contra la infección crónica por el VHB más utilizados. A pesar de esto, el tratamiento a largo plazo con AN también tiene desventajas, como por ejemplo que se desconocen sus efectos acumulativos sobre la salud de los pacientes³⁰⁵. Además al

actuar directamente sobre el VHB se pueden seleccionar variantes que contengan resistencias a su acción, es decir, mutaciones adaptativas que reducen la sensibilidad de estas variantes a los efectos inhibidores del tratamiento²⁶⁰ (ver en el Capítulo 1 el apartado 2.4.2 *Variabilidad en el ORF P*).

Los AN se caracterizan por diferentes patrones de potencia y barrera genética, es decir, por la probabilidad de seleccionar variantes resistentes a su acción. En general, actualmente no se usa la LAM como tratamiento de primera línea por su alta tasa de selección de variantes resistentes (barrera genética baja) y se emplea en algunos casos seleccionados donde se prevé que la terapia durará poco tiempo, como en la hepatitis aguda grave, fallo hepático agudo, profilaxis de transmisión vertical durante el embarazo o en los pacientes que serán inmunosuprimidos para evitar la reactivación³⁰⁵. La LdT y ADV tampoco se usan como tratamientos de primera línea porque tienen menor potencia y hay posibilidades de seleccionar variantes resistentes (barrera genética intermedia)³⁰⁵. Hoy en día se prefiere iniciar el tratamiento con los AN de alta potencia y baja o nula selección de variantes resistentes (barrera genética alta), como el ETV o TDF y en aquellos pacientes con riesgo de sufrir efectos adversos con el TDF el TAF³⁰⁵.

El tratamiento de un año de AN en pacientes HBeAg positivos conduce a la supresión del ADN viral hasta concentraciones indetectables entre el 21%-76% de los pacientes, con una normalización de las transaminasas hasta en un 77%, seroconversión del HBeAg entre el 12%-22% y pérdida del HBsAg del 0%-3%. La extensión del tratamiento hasta 5 años aumenta la seroconversión del HBeAg hasta el 48% y la tasa de pérdida del HBsAg hasta el 10%³²¹. En el caso de los pacientes HBeAg negativos los resultados son algo mejores la supresión del ADN viral hasta concentraciones indetectables puede llegar hasta el 47%-95% de los pacientes, la normalización de las transaminasas hasta en un 87% y la pérdida del HBsAg hasta el 5%^{322,323}.

4.4 Nuevas terapias antivirales

Recientes descubrimientos científicos que han permitido una mejor comprensión del ciclo de replicación del VHB y su interacción con el huésped, incluidos la identificación del receptor celular para la entrada del VHB, el descubrimiento de enzimas nucleares claves en formación ADNccc y en su control epigenético, la observación de degradación parcial de ADNccc inducida mediante IFN o las vías de señalización factor nuclear- $\kappa\beta$, y la identificación del papel de la HBx en la transcripción de VHB³²⁴. Estos descubrimientos junto con los principales avances en la investigación básica del VHB allanaron el camino

para la identificación de nuevos objetivos terapéuticos, que pueden conducir hacia una cura funcional, o incluso a una cura total en el futuro, de la infección por el VHB.

Para lograr el objetivo de una “cura” del VHB, las guías de tratamiento clínico de la EASL sugieren que probablemente serán necesarias combinaciones de terapia antiviral dirigidas a múltiples pasos en el ciclo de replicación del VHB (que supriman la replicación viral y la producción de antígenos virales) y terapia inmunomoduladora (para restaurar la respuesta inmune al VHB)³⁰⁵. A continuación, se explicarán los avances recientes en estas dos estrategias de tratamiento, en el momento en el que se escribe esta tesis.

4.4.1 Antivirales de acción directa (DAA) (Figura 39)

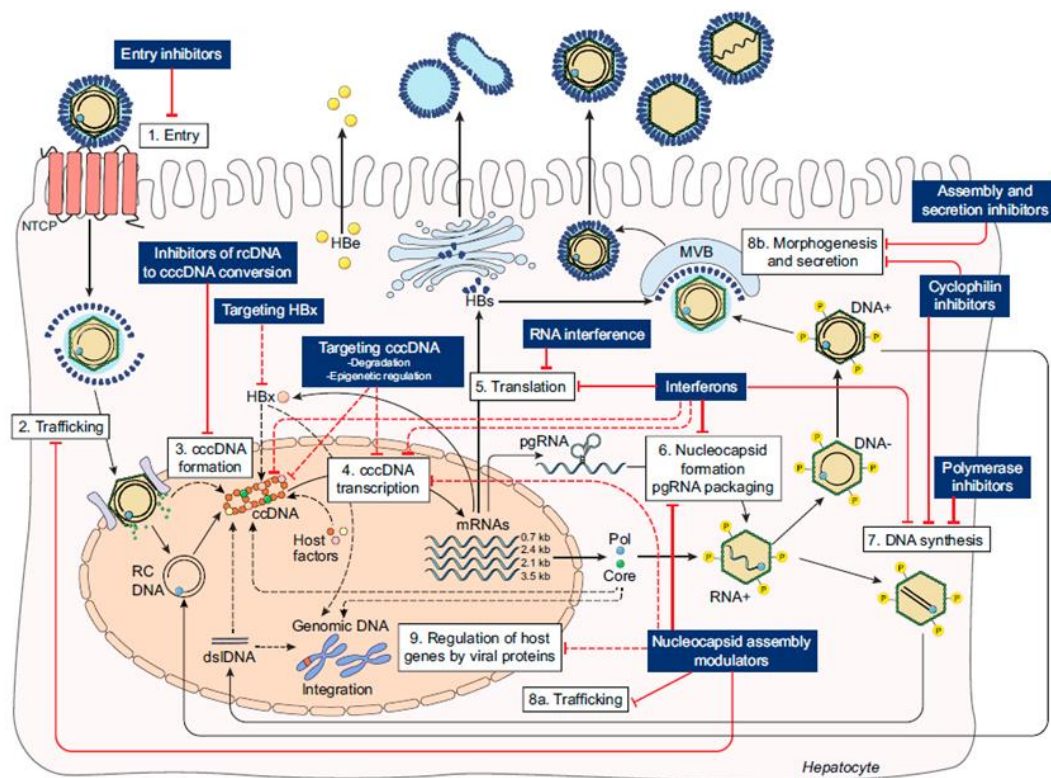


Figura 39 El ciclo replicativo del VHB y los puntos en los que podrían interferir los DAA. ³¹⁴.

4.4.1.1 Inhibidores de entrada al VHB

La vía por la cual el VHB entra en los hepatocitos implica primero la interacción del virus con los HSPG, permitiendo luego la unión del dominio preS de las LHBs a NTCP, seguido de fusión de membrana y la liberación de nucleocápside en el citoplasma de células infectadas (ver en el Capítulo 1 el apartado 1.6.1 *Entrada del virus*) (Figura 39). Los inhibidores de entrada se pueden clasificar según sus modos de acción: (1) anticuerpos neutralizantes del N-terminal de los epítomos en el dominio pre-S1. Estos anticuerpos son altamente específicos pero requieren de la administración parenteral, y son necesarias grandes cantidades de anticuerpos para neutralizar las partículas subvirales³¹⁴. (2) los inhibidores de la unión del virus son fármacos con carga negativa o positiva que se unen al virus (por ejemplo, la heparina) o a HSPG (por ejemplo, poli-lisina). Éstos son eficientes pero no específicos³¹⁴. (3) Los sustratos de NTCP que incluyen las sales biliares conjugadas (por ejemplo, taurocolato, ezetimiba) y otras moléculas pequeñas que son transportadas por NTCP; el requisito de altas concentraciones y el corto tiempo que permanecen en el receptor, limitan su aplicación clínica.³¹⁴ (4) Inhibidores de NTCP irreversibles: Myrcludex-B, la ciclosporina A y los derivados son inhibidores alostéricos de NTCP. Bloquean irreversiblemente la función del receptor.

Los inhibidores de entrada pueden prevenir la formación de novo del ADNccc en hepatocitos no infectados y pueden ser muy útiles en la prevención de la transmisión de madre al niño o en la reinfección después del trasplante de hígado³¹⁴. Entre estos tratamientos, la opción que se encuentra en una fase más avanzada de su evaluación en ensayos clínicos es Myrcludex-B, es un lipopéptido sintético (péptido miristoilado) derivado del dominio pre-S1 de la proteína grande de la envuelta del VHB dirigida al NTCP,³²⁵ el cual se ha demostrado que inhibe eficazmente la entrada del VHB³²⁵. En ratones inmunodeficientes humanizados infectados con VHB, el tratamiento con Myrcludex-B redujo la carga viral del suero y las concentraciones de HBsAg. Los resultados mostraron que Myrcludex-B no solo inhibe la amplificación de intrahepática del ADNccc, sino que también evita la propagación viral intrahepática³²⁶. Aunque los resultados del ensayo clínico de fase IIa están pendientes, el efecto combinado de Myrcludex-B y el interferon serían complementarios en el control Infección por VHB^{314,325}.

4.4.1.2 Degradación / inhibición de la síntesis de ADNccc

Varios estudios demuestran que podría ser posible dirigir el tratamiento antiviral específicamente al ADNccc a pesar de su ubicación en el núcleo de los hepatocitos. Estudios *in vitro* utilizando hepatocitos mostraron que varias citocinas (IFN- α , agonistas del receptor linfotoxina- β , El IFN- γ y el FNT- α) pueden modular las vías que conducen a la regulación positiva de las deaminasas APOBEC3A / B, que a su vez inducen degradación de ADNccc. Sin embargo, estos estudios *in vitro* solo mostraron una degradación parcial del ADNccc³²⁷.

También se podría eliminar el ADNccc mediante las enzimas de escisión de ADN, incluyendo endonucleasas o meganucleasas que son dirigidas a regiones específicas del genoma del VHB y que tiene como objetivo inhibir la replicación del VHB (Figura 39). Los tres diseños más comúnmente usados dirigidas a ADNccc son: nucleasas con dedos de zinc (ZFN), los *transcription activator-like effector nucleases* y sus derivados (TALENs) y el *clustered regularly interspaced short palindromic repeats* y sus proteínas asociadas (CRISPR/cas)³¹⁴:

1. ZFN: En el interior del hepatocito 2 ZFN reconocen partes opuestas de una secuencia del ADN viral y escinden la doble cadena. Es por esto por lo que teóricamente unos ZFN diseñados para atacar el ADN del VHB no tendrían efecto sobre el genoma del hospedador. Varios estudios *in vitro* han demostrado que mediante este mecanismo se disminuye la concentración de ARNpg y ADNccc significativamente^{328,329}. No obstante, el principal reto es lograr encontrar un mecanismo de entrega eficiente a los hepatocitos³³⁰.

2. TALEN: con una estructura similar a las ZFN, los TALEN cuentan con una secuencia de 33 a 35 aminoácidos que en la posición 12 y 13 reconocen un segmento específico del ADNccc. A diferencia de los ZFN, los TALEN escinden el ADNccc como monómeros³³⁰. Estudios *in vitro* y en ratones han demostrado su eficacia³³¹, por lo que si se lograra que ingresaran a los hepatocitos *in vivo*, los TALEN podrían ser una cura definitiva para la infección por VHB³³².

3. CRISPR/cas: es un sistema bacteriano con actividad nucleasa que está involucrado en la defensa contra bacteriófagos y plásmidos. Está compuesto por los genes cas y el ARN no codificante que determina la especificidad del sistema (ARN guía). Recientemente se ha demostrado su efectividad para disminuir la replicación del VHB *in vitro*^{333,334}, lo que abre las puertas a futuras investigaciones clínicas sobre este grupo de

medicamentos contra el virus. Algunos de estos estudios sugieren que la edición de ADN de VHB por escisión de nucleasa asociadas a CRISPR/Cas es más eficiente que la de las citocinas, mediada por la desaminasa APOBEC3/B después del tratamiento de células infectadas con IFN³³².

Adicionalmente también se podría disminuir el *pool* de ADNccc inhibiendo su síntesis desde ADNrc. Este proceso dentro del ciclo replicativo del VHB incluye diferentes pasos que se cree que se basan en varias enzimas nucleares del huésped que aún no se han identificado por completo (ver en el Capítulo 1 el apartado 1.6.3 *Formación del ADNccc*). Además, podría haber redundancia entre estos factores del huésped, como se ha observado en el caso de TDP2¹⁷¹. Esto podría limitar seriamente la posibilidad de desarrollar estrategias específicas contra una diana concreta. Sin embargo, se ha demostrado que dos compuestos de sulfonamida disustituidos estructuralmente relacionados bloquean la formación de ADNccc y podrían servir como prueba de concepto de fármacos candidatos para prevenir la formación inicial y / o el mantenimiento de ADNccc³³⁵.

4.4.1.3 Silenciamiento funcional

El ADNccc viral está organizado en una especie de estructura de cromatina y está sujeto a regulación epigenética. Los modificadores epigenómicos pueden bloquear el ADNccc a nivel de la actividad transcripcional (Figura 39). Sin embargo, si no son específicos, pueden inducir efectos nocivos en las células del huésped. Algunas citoquinas, incluyendo IFN- α , se ha demostrado que disminuyen la transcripción de ADNccc a través de modificaciones epigenéticas, en modelos preclínicos³³⁶. Se ha demostrado que la HBx es necesario para la transcripción de ADNccc y para la replicación viral; por lo tanto, esta proteína puede ser una atractiva diana viral para silenciar no solo de la transcripción ADNccc³³⁷ sino también de otros mecanismos dependientes de la HBx, relacionados con las interacciones celulares del virus (immunomodulación, hepatocarcinogénesis, etc) (Figura 39). Los ensayos actuales para detectar ADNccc carecen de especificidad y un método estandarizado para medir la transcripción de ADNccc es necesario.

4.4.1.4 ARN pequeños de interferencia (*small interference RNA* o *siRNA*)

Los siRNA son un tipo de ARN interferente con una longitud de 20 a 25 nucleótidos altamente específicos para la secuencia de nucleótidos de su ARNm diana. Los siRNAs promueven el silenciamiento de genes mediante su unión a una diana específica en el

ARNm, activando así el Complejo de Silenciamiento Inducido por el ARN (RISC) y promoviendo la degradación de la diana del ARN (Figura 40). Debido a que el genoma del VHB es compacto y los ORF se superponen en él, el uso de los siRNA es alentador ya que un solo ARN interferente puede unirse a los transcritos de más de un ORF. Así, algunos estudios han demostrado que siRNA dirigidos a los transcritos de *core* o del *HBX* detienen la expresión génica del VHB, lo que se evidencia con la disminución significativa de HBsAg, HBeAg, ARN y ADN virales^{332,338,339}. Resultados preliminares de un ensayo de fase II mostró que una sola dosis de ARC-520 (una combinación de siRNAs) dirigido contra el *HBX* en combinación con entecavir consiguió disminuir el ADN-VHB en suero en pacientes HBeAg positivo y en pacientes HBeAg negativos y disminuir el nivel de HBsAg en HBeAg-positivo, pero no en pacientes HBeAg negativos^{334,340}. Otros estudios son necesarios para determinar si la disminución de la producción de HBsAg por sí sola podría ser suficiente para restaurar la respuesta de células T específica de VHB o si es necesario combinarlos con otras terapias antivirales o inmunomoduladoras.

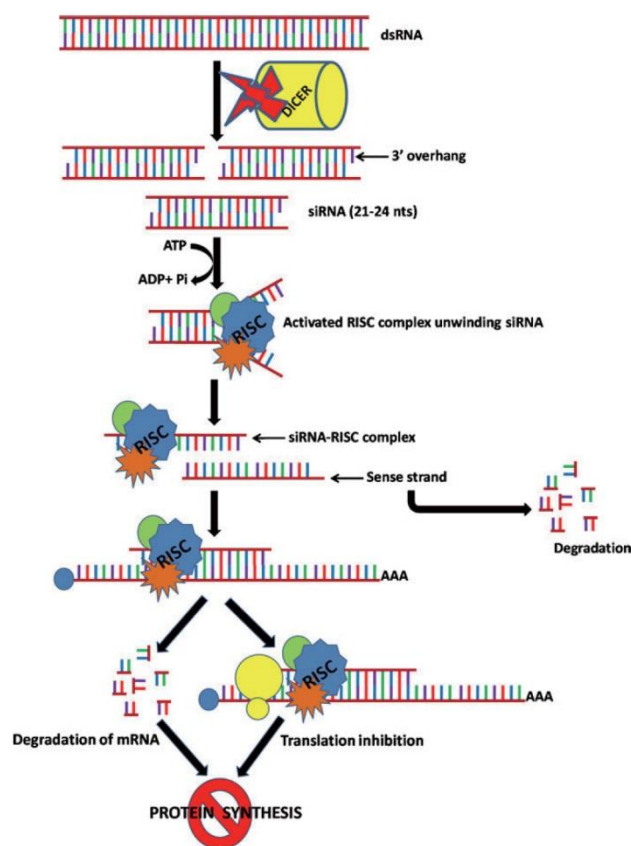


Figura 40 Esquema del funcionamiento de un siRNA. Los siRNAs promueve el silenciamiento de genes mediante su unión a una diana específica en el ARN, activando así el RISC y promoviendo la degradación de la diana del ARN.

4.4.1.5 Inhibidores de la encapsidación y el empaquetamiento del ARNpg

La formación y el empaquetamiento en la nucleocápside del ARNpg (el molde para la replicación viral) son pasos críticos del ciclo replicativo. La proteína HBcAg está involucrada en muchos aspectos del ciclo replicativo³¹⁴ (ver en el Capítulo 1 el apartado 1.4.1.1 *ORF PC/Core y las proteínas precore y core*). El HBcAg se ha postulado como un objetivo antiviral muy prometedor para inhibir o modular el empaquetamiento del ARNpg (Figura 39). Con el conocimiento de la estructura tridimensional de esta proteína, se han desarrollado varias clases de moléculas pequeñas llamadas moduladores alostéricos del ensamblaje de proteínas *core* (CpAMs), incluida la fenilpropenamida y derivados de la heteroarildihidropirimidina. Estas moléculas pueden fortalecer la interacción proteína-proteína, inhibir la encapsidación de ARNpg, y bloquear la síntesis de ADN de cadena adicional^{314,325}. Resultados del primer estudio de fase Ib de fase Ib de NVR3-778 mostró una disminución en el ADN del VHB en suero, del ARN del VHB y de la concentración de HBsAg, con un efecto más pronunciado cuando se combina con pegIFN³⁴¹.

Además, la HBcAg también desempeña un papel importante en el núcleo de las células infectadas, ya que forma parte del ADNccc y puede regular la expresión del genoma del huésped a través de mecanismos epigenéticos. Así pues, los CpAMs pueden tener funciones adicionales además de su inhibición “canónica” del ensamblaje de las nucleocápside en el citoplasma³⁴².

4.4.1.6 Inhibidores de HBsAg

Las estrategias que inhiben la expresión de los genes virales a través de la transcripción de ADNccc o de la traducción de ARNm viral (ver en el Capítulo 1 el apartado 4.4.1. *Antivirales de acción directa (DAA)*) pueden disminuir los niveles séricos de HBsAg. La eficacia de los anticuerpos monoclonales destinados a neutralizar y/o agotar el HBsAg del torrente circulatorio en personas con CHB, es reducida debido a la gran cantidad de HBsAg circulante, a menos que se utilice en combinación con otros enfoques. Se ha visto que los polímeros de ácidos nucleicos fosforilados (también conocidos como NAP) disminuyen la secreción de HBsAg, mediante la formación de interacciones hidrófobas cooperativas con la superficie hidrófoba de una hélice expuesta, evitando así cambios conformacionales o interacciones entre la hélice diana y otras hélices anfipáticas^{314,343} (Figura 39). Ensayos clínicos de los NAP REP 2055 y REP 2139 usados en monoterapia seguido de combinación con pegIFN indujeron una marcada disminución

en los niveles HBsAg circulantes, de la viremia y algunos pacientes seroconvirtieron a anti-HBs³⁴⁴. Se observaron resultados similares en estudios piloto de TNF y pegIFN en combinación con REP 2139 o REP 2165³⁴⁴. Estos resultados deben ser replicados en estudios más amplios.

4.4.1.7 Inhibidores RNAsa H

Las RNAsa H degradan el ARN cuando está unido al ADN, quitan los cebadores durante la síntesis de ADN, contribuyen a la destrucción de errores fallidos de transcripción y son necesarias para la transcripción reversa de genomas virales como el del VHB (esta función la realiza el dominio RH de su polimerasa) (Figura 39). La inhibición de esta enzima es ampliamente conocida como parte del tratamiento contra el VIH, pero poco se ha estudiado sobre su uso como tratamiento para el VHB, tan solo existen algunos estudios *in vitro* que sugieren su efectividad³⁴⁵. El objetivo de los medicamentos que comparten este mecanismo de acción es suprimir la replicación viral para detener el reaprovisionamiento de ADNccc en el hígado, lo que permite eliminar las células infectadas por el sistema inmune o por otro medicamento concomitante³⁴⁶.

4.4.2 La respuesta inmune al VHB y las implicaciones para terapias inmunomoduladoras

La inducción de la respuesta innata y la activación de citoquinas son débiles en CHB³¹⁴, por eso la producción de IFN- γ de las células natural killer se detecta en la hepatitis aguda pero es deficiente en la CHB. Las células T CD8+ específicas del VHB, necesarias para el control del VHB (que también puede desempeñar un papel en la inflamación hepática), son disfuncionales y muestran un fenotipo “agotado” en CHB^{310,314}. Por estas razones, las dianas de las terapias inmunomoduladoras son una serie de mecanismos implicados en la modulación inmune, para así poder crear o restaurar la respuesta inmune específica del VHB e inhibir su replicación y la producción de HBsAg (figura. 41). Los principales problemas de la terapia inmunomoduladora son la posible inducción de brotes de hepatitis no controlados y autoinmunidad.

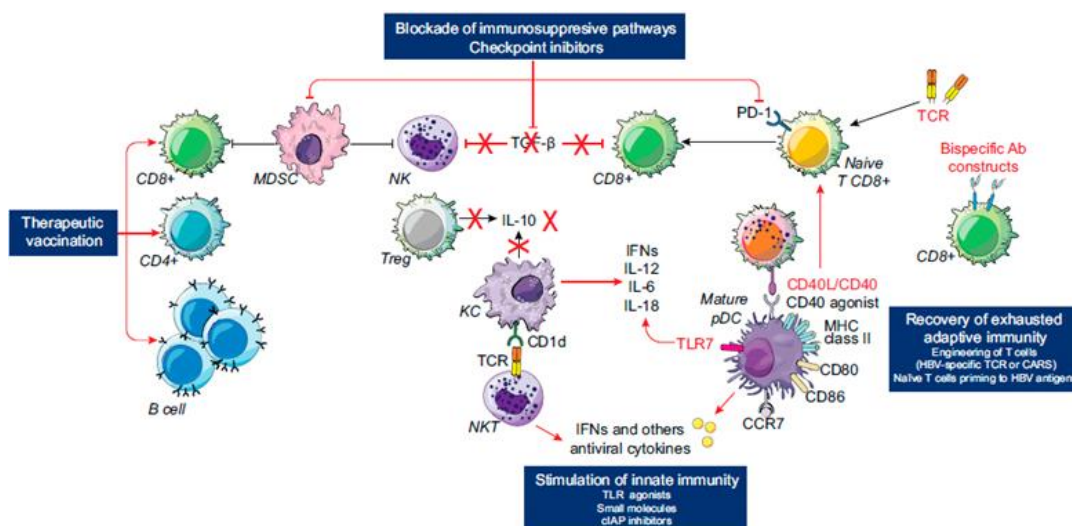


Figura 41 El microambiente inmune del hígado y las dianas inmunoterapéuticas³¹⁴

4.4.2.1 Receptores de reconocimiento de patógenos

Los receptores tipo Toll (o *Toll-like receptor* TLRs) constituyen una familia de proteínas que forman parte del sistema inmunitario innato. Estos receptores son proteínas transmembrana que reconocen patrones moleculares expresados por un amplio espectro de agentes infecciosos y estimulan una variedad de respuestas inflamatorias. Varios estudios demostraron que la expresión de TLRs en los hepatocitos reduce la presencia de varios productos virales del VHB³²⁵. Estos datos nos indican que el TLR7 mediado por la activación de la respuesta inmune puede ayudar en el control del VHB. GS-9620, un agonista oral del TLR7, indujo una disminución en el suero los niveles de ADN de HBV y HBsAg, así como de ADN-VHB en chimpancés infectados por VHB³¹⁴. Efectos similares también se observaron en marmotas³¹⁴, pero no en humanos^{347,348}. Se están llevando a cabo varios ensayos clínicos que intentaran evaluar la eficacia de los agonistas de TLRs en la eliminación de infección por VHB³⁴⁹.

4.4.2.2 Estimuladores de agonistas del gen de IFN

Los agonistas del gen IFN son proteínas adaptadoras de múltiples receptores de ADN citoplasmáticos y un receptor de reconocimiento de patógenos, por lo que pueden ser un objetivo potencial de la farmacología para la activación de la respuesta inmune innata³¹⁴. Estimuladores de estos agonistas también pueden ser usados como adyuvantes para la vacunación terapéutica. Además, estos estimuladores de los agonistas del IFN inducen una respuesta robusta de las citoquinas en los macrófagos y suprimen eficazmente la

replicación del VHB a través del bloqueo de la estructura épsilon de ARNpg³²⁵. SB 9200 es un novedoso modulador oral de inmunidad innata que se cree que actúa a través de la activación de las vías RIG-I y NOD-2, se ha visto que disminuye en suero la concentración de ADN del virus y los niveles de HBsAg en marmotas³⁵⁰. Varios ensayos clínicos están en curso³¹⁴.

4.4.2.3 Inhibidores celulares de proteínas de la apoptosis (cIAP)

Los cIAP son antagonistas naturales el efecto pro-apoptótico del TNF- α . Estudios recientes han demostrado que Birinapant, una molécula que inhibe cIAP eficaz contra numerosos cánceres³⁵¹, favoreció el aclaramiento del VHB en ratones mediante un mecanismo mediado por apoptosis asociado con la producción de TNF- α y la activación de una respuesta de células T CD4+ específica del VHB³⁵².

4.4.2.4 Recuperación de la respuesta adaptativa “agotada”

La transferencia adaptiva de células autólogas podría ser un método para aumentar el número de células T específicas para VHB. Las células T reinfundidas podrían expresar receptores de antígeno quiméricos específicos para VHB (CAR), lo que permitiría el reconocimiento independiente de HLA de hepatocitos infectados. Alternativamente se podrían transferir células T modificadas genéticamente que sobreexpresan los receptores de células T restringido por HLA (TCR)³⁵³. Otro nuevo y prometedor enfoque de terapia inmunológica consiste en redirigir las células efectoras inmunitarias hacia los hepatocitos infectados por el VHB utilizando anticuerpos similares a los receptores de células T (TCR-L)³⁵³.

4.4.2.5 Bloqueo de señales inmunoinhibidoras

El fenómeno conocido como *agotamiento* de los linfocitos T específicos del VHB, que se produce en la infección crónica por este virus y del que se ha ido hablando en los anteriores apartados se caracteriza por la escasez o ausencia de células T específicas contra el VHB. Está asociado con pobre actividad citotóxica, deterioro en la producción de citoquinas y aumento en la expresión de receptores coinhibitorios, como la proteína de muerte celular programada 1 (PD-1), antígeno 4 asociado al linfocito T citotóxico (CTLA-4), proteína 3 del gen de activación de linfocitos (LAG 3), entre otros³⁵⁴. Este fenómeno es mantenido por la presencia de un microambiente de citoquinas inmunosupresoras, como IL-10 y el TGF- β producidas por los linfocitos T reguladores

encontrados en gran cantidad en el hígado de pacientes con CHB³⁵⁵. Recientemente se ha estudiado el bloqueo de estos receptores por medio de anticuerpos específicos como diana terapéutica en el cáncer, con lo que se encontró un aumento en la supervivencia³²⁵. Tal inhibición puede revertir la disfunción inmune en la hepatitis B, como se ha demostrado en estudios con marmotas y estudios *ex vivo* en humanos³¹⁴

4.4.2.6 Vacuna terapéutica

El objetivo de la vacunación terapéutica es estimular o impulsar la respuesta inmune del huésped para restaurar el control inmune, lo que conduce a una supresión sostenida de la replicación del VHB y, en última instancia, a la pérdida del HBsAg. Las vacunas terapéuticas incluyen la vacuna convencional contra el HBsAg, con o sin potentes adyuvantes, la vacuna que estimula células T específicas, la formada por complejos inmunes HBsAg-antiHBs, la vacuna de células apoptóticas que expresan antígenos del VHB, la vacuna de ADN y la vacuna con vectores virales que expresan proteínas de VHB, todas ellas han sido evaluadas con un éxito limitado³¹⁴. El pretratamiento con AN para suprimir la replicación del VHB puede mejorar la respuesta de las células T específicas del VHB³¹⁴. Un ejemplo de vacuna terapéutica es GS-4774 que expresa regiones bien conservadas de proteínas de VHB. En el estudio con ratones se mostró que GS-4774 puede estimular las células T CD8 (+) específicas del VHB. En sujetos sanos, la inmunización con GS-4774 condujo a la respuesta del sistema inmunitario específico del VHB³²⁵. Aunque la vacunación conduce a la reducción en HBsAg y los niveles de ADN del VHB en modelos de ratón, la capacidad de aumentar la respuesta de células T para controlar la replicación del VHB en los humanos necesitan una mayor confirmación³²⁵.

4.5 Perspectivas de futuro.

Muchos programas de investigación están en curso para desarrollar nuevos tratamientos que se centran en la eliminación del HBsAg del VHB, los principales objetivos de estos programas son: suspender el tratamiento sin riesgo de recaída virológica y sin riesgo progresión de la enfermedad y disminuir el riesgo de desarrollar HCC³⁰⁵, es decir conseguir una cura funcional de la infección. Gracias a estos programas, las nuevas opciones de tratamiento pueden estar disponibles en un futuro próximo.

Se presupone que el camino hacia una cura total de la hepatitis B radicaría en la eliminación del ADNccc y como se ha comentado en el apartado anterior hay estrategias

CAPÍTULO 1. INTRODUCCIÓN

prometedoras en este sentido. Sin embargo, la curación total puede llegar a no ser factible debido a que no solo la persistencia del ADNccc en el hepatocito dificulta la eliminación del VHB, sino también a que el ADN del VHB está integrado en el genoma del huésped. Idealmente, si la curación total del VHB en las primeras etapas de la enfermedad hepática fuera posible, teóricamente tendría mayor impacto en la reducción del riesgo de HCC.

CAPÍTULO 2: HIPÓTESIS Y OBJETIVOS

CAPÍTULO 2: HIPÓTESIS Y OBJETIVOS

1. Hipótesis

La HBx del VHB, codificada por el *HBX* de este virus, es una proteína no estructural y reguladora, con múltiples funciones. Esta proteína participa en la transactivación de la expresión de los genes virales regulando la transcripción del ADNccc y por lo tanto es crucial para la replicación del VHB¹⁰⁴. Además, esta proteína interactúa con varias vías de señalización celulares y regula la expresión de múltiples genes del huésped, lo que afecta a muchas actividades celulares^{275,356,357}.

La hipótesis central de este proyecto consiste en que dado el papel esencial de la HBx deben existir en la secuencia nucleotídica del gen que la codifica, así como en su secuencia de aminoácidos, zonas altamente conservadas que podrían ser las responsables principales de su actividad reguladora. Si dichas zonas se mantienen altamente conservadas en cualquier estadio de la infección, durante la evolución natural de la misma e incluso bajo el efecto de tratamiento antiviral, esto sugeriría que estas no pueden variar para mantener la replicación viral, y por lo tanto evidenciaría su papel esencial. Estas zonas o dominios con la secuencia híper-conservada, podrían incluso ser útiles para el diseño de una terapia antiviral basada en terapia génica. Esta terapia no solo permitiría inhibir la replicación viral de manera duradera, sino que, a largo plazo, podría permitir erradicar el ADNccc, principal responsable de la persistencia de esta infección. Además, el estudio en profundidad de la QS de *HBX* (complejidad, variantes genéticas que la forman), podría conducirnos a un mejor entendimiento del comportamiento del VHB, en los diferentes estadios clínicos de su infección crónica. Estudiar en profundidad la QS viral de la región 5' del *HBX* a nivel de los pacientes con infección crónica HBeAg negativo, antes conocidos como “portadores inactivos”, podría ayudar a esclarecer porque el VHB replica a niveles más bajos en estos pacientes. A su vez, esto podría proporcionar una información muy valiosa sobre los mecanismos que permiten controlar la replicación viral.

Se ha descrito una presencia significativa de deleciones y / o inserciones (indels) en el extremo 3' de *HBX*²⁸⁶⁻²⁸⁸, lo que sugiere que difícilmente se encontrarán zonas híper-conservadas en esta región²¹⁹. Sin embargo, la conservación en el extremo 5' de *HBX* permanece todavía sin explorar por lo que hemos decidido explorar esta región en busca de estas zonas. En el extremo 5' de los transcritos del *HBX* incluye una región no codificante que también podría contener regiones híper-conservadas, tanto éstas como las regiones codificantes son potencialmente ser útiles como candidatas para contener dianas de una posible terapia génica dirigida para silenciar el *HBX* a nivel postranscripcional.

CAPÍTULO 2: HIPÓTESIS Y OBJETIVOS

Además, la secuencia del *HBX* se encuentra en el extremo 3' co-terminal de todos los ARN mensajeros (ARNm) del VHB producidos durante la infección. Por consiguiente, una terapia génica dirigida contra los transcritos del *HBX* tanto en la región codificante como en la no codificante, podría lograr la interferencia con la expresión de todas las proteínas virales.

2. Objetivos

- 1) Buscar zonas híper-conservadas en el *HBX*, potencialmente útiles como dianas para su bloqueo mediante terapia génica dirigida.
- 2) Estudiar la complejidad de las poblaciones virales (QS) del VHB, mediante diferentes estimadores (Entropía de Shannon, índice de Gini-Simpson, frecuencia de mutación, diversidad nucleotídica) para valorar su efectividad como valor pronóstico de los diferentes grados de severidad de la infección.
- 3) Estudiar las diferentes poblaciones virales con mutaciones por sustituciones, inserciones o deleciones, en la región codificante de la *HBx* del VHB, para establecer el posible valor pronóstico de dichas mutaciones como marcador de disminución de la replicación viral.

CAPÍTULO 3: PRIMER ESTUDIO

PRIMER ESTUDIO: Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy

1. Introducción

La infección por el VHB se puede controlar con los tratamientos actuales pero la eliminación completa de la infección, es decir su erradicación, no se logra debido a la persistencia del ADNccc en los núcleos de hepatocitos infectados. Estos tratamientos, mayoritariamente, actúan como inhibidores de la retrotranscriptasa viral, por lo que interfirieren en la fase de retrotranscripción del ARNpg y la producción de nueva partículas virales, pero no afectan a la actividad del ADNccc. Esta molécula de ADN viral que persiste como un minicromosoma en el núcleo de los hepatocitos virales constituyendo un reservorio de la replicación viral y permite la expresión sostenida de las proteínas del virus, incluso en presencia de una terapia antiviral eficaz. La continua expresión de antígenos virales podría ser determinante en la progresión de la enfermedad hepática, incluso en el desarrollo del HCC. Dada la dificultad para eliminar el ADNccc del núcleo de las células infectadas, hoy en día se están desarrollando nuevas estrategias terapéuticas anti-VHB con el objetivo de conseguir una “cura funcional” de la infección, que es equivalente a la resolución espontánea de una infección aguda³⁵⁸ (ver en el Capítulo 1 el apartado 5.1 *¿qué se entiende por curación de la infección del VHB?*). Estas nuevas estrategias están dirigidas o bien a potenciar la respuesta inmunitaria del huésped o bien hacia el desarrollo de terapias contra dianas específicas del virus que afecten a diferentes pasos del ciclo replicativo del VHB. A día de hoy parece que la estrategia más prometedora para conseguir una “cura funcional” de la infección, es la combinación de estos dos tipos de tratamientos³⁰⁵. Las terapias génicas que permiten la destrucción del ADNccc o el silenciamiento de su transcripción son opciones interesantes, a tener en cuenta, para conseguir la cura funcional o incluso la erradicación de la infección viral. Actualmente se están investigando terapias que incluyen: los ZFNs, TALENs, CRISPR-Cas9s, siRNAs, etc (ver en el Capítulo 1 en el apartado 5.3 *Nuevas terapias antivirales*). Los siRNAs son utilizados para silenciar genes específicos a nivel postranscripcional a través de una interacción específica de secuencia con el ARNm diana, dando como resultado la inhibición de la expresión de la proteína viral³³⁸. Recientemente Arbutus Biopharma Corporation ha publicado dos estudios en fase II, en los que se utilizó un siRNA como tratamiento para pacientes con CHB y en el estudio preliminar los datos

indicaron que la terapia fue bien tolerada y condujo a una reducción significativa en los niveles de HBsAg^{340,359}.

Entre todas las proteínas del VHB, la HBx, codificada por el *HBX*, es un factor determinante en la infección. Esta proteína regula la expresión de ADNccc e interactúa con varias vías celulares, lo que facilita la progresión de la enfermedad hepática (ver en el Capítulo 1 el apartado 1.4.1.4 *ORF X y la proteína X*). Así pues, el bloqueo de la expresión de esta proteína podría inhibir la replicación viral y frenar la progresión de la enfermedad hepática. Además, el *HBX* se encuentra cerca del extremo 3' común todos los ARNm del VHB (figura 1 del Capítulo 2, apartado 5. *Manuscrito completo: "Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy"*), en consecuencia, al dirigir la terapia génica (ej siRNA) a los transcritos del *HBX* tanto en la región codificante como en la no codificante, podría lograrse la interferencia con la expresión de todas las proteínas virales.

2. Hipótesis y objetivo

2.1. Hipótesis:

La hipótesis central de este estudio consiste en que el bloqueo de las regiones del genoma del VHB codificantes o reguladoras del *HBX*, a nivel de los ARNm, podría inhibir la replicación viral de manera duradera e incluso contribuir a erradicar el ADNccc. Sin embargo, el complejo mecanismo de replicación del genoma del VHB lo lleva a una tasa de mutación similar a las observadas en algunos virus de ARN, por lo que este virus circula como una compleja mezcla de variantes genéticas conocida como QS. Esto podría causar variabilidad en la secuencia diana y por lo tanto disminuir la efectividad de dicho bloqueo. Así pues, la identificación de regiones híper-conservadas dentro del *HBX* puede ser considerada una información valiosa para el desarrollo de una hipotética terapia de bloqueo génico. Para asegurar la efectividad de esta terapia, la conservación de dichas zonas debe demostrarse en cualquier estadio de la infección, durante la evolución natural de la misma e incluso bajo el efecto de tratamiento antiviral. Además, dado que el perfil de mutaciones del VHB está estrechamente asociado con su genotipo³⁶⁰, idealmente estas zonas también deberían estar conservadas en los diferentes genotipos virales. Se ha encontrado una presencia significativa de múltiples variantes con deleciones y/o inserciones en el extremo 3' del *HBX*²⁸⁶⁻²⁸⁸, pero la conservación en el extremo 5' del

HBX permanece sin explorar, por lo que lo consideremos un objetivo potencial para buscar dianas de una posible terapia génica dirigida.

2.2. Objetivo:

El objetivo central de este estudio ha sido identificar regiones híper-conservadas en la región codificante 5' y la región no codificante del *HBX* (incluida en todos los transcritos del VHB), independientemente del estado clínico del paciente o del genotipo del VHB.

3. Resumen del estudio

3.1 Diseño del estudio:

Se ha estudiado la región 5' del *HBX* y la región no codificante incluida en sus transcritos (nts 1255 a 1611) mediante la tecnología de NGS basada en ultra-deep pyrosequencing (UDPS), en 27 pacientes infectados por el VHB que representan un amplio espectro de situaciones clínicas y virológicas de la infección crónica por VHB. El flujo de trabajo llevado a cabo se muestra en la siguiente figura.

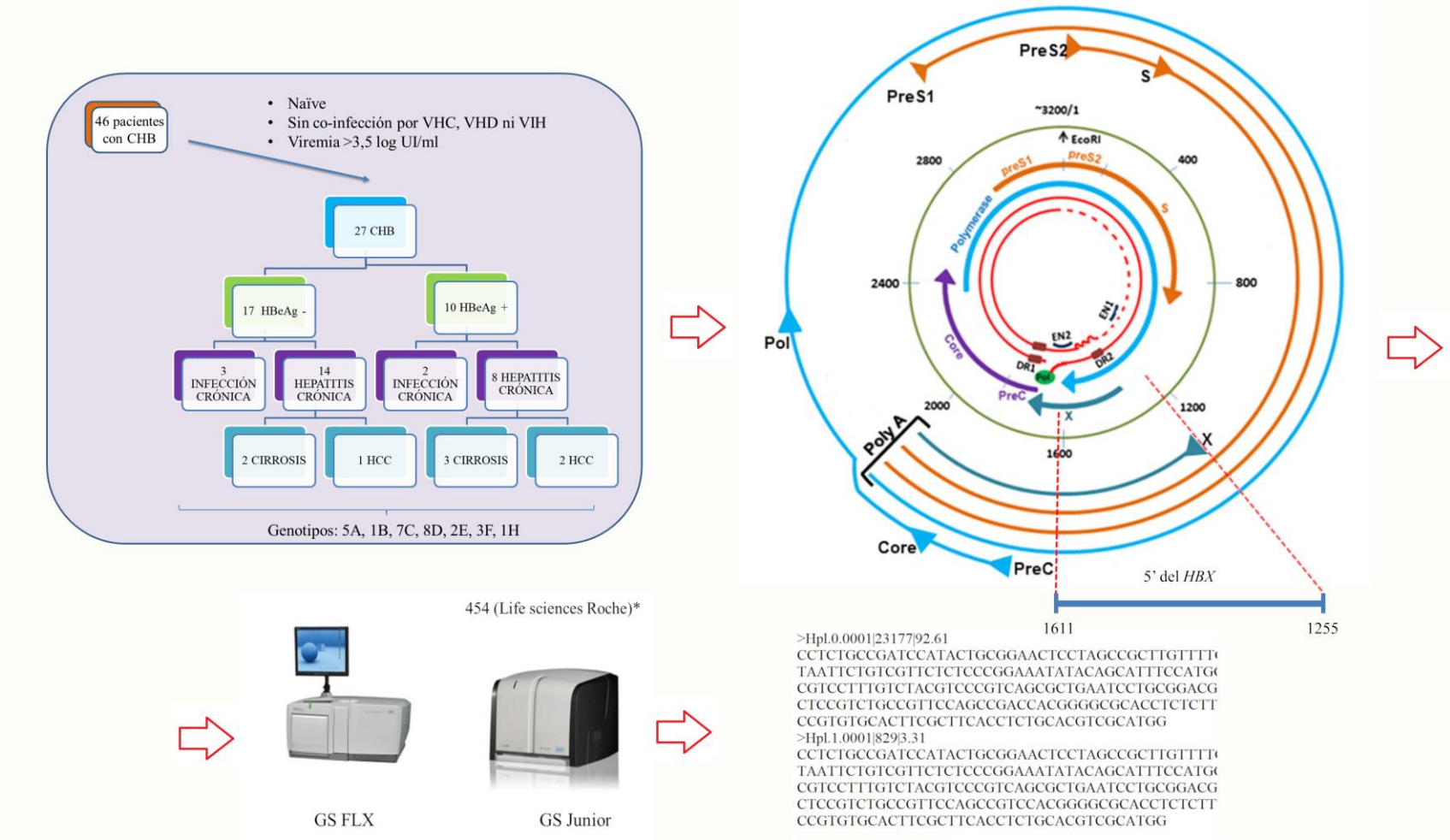


Figura 42 Flujo de trabajo del estudio. VHC virus de la hepatitis C, VHD virus de la hepatitis D, VIH virus de la inmunodeficiencia humana. * *Secuenciadores FLX y Junior (454 Life sciences-Roche, Branford, USA), ambas plataformas intercambiables*⁶⁶

3.2 Análisis bioinformático de los datos de secuenciación

3.2.1. Filtrado de los datos de secuenciación masiva:

Las lecturas obtenidas después de la secuenciación se sometieron a un procedimiento de filtrado bioinformático desarrollado por nuestro propio grupo. Los haplotipos (secuencias únicas de variantes genómicas altamente relacionadas entre sí) con una frecuencia relativa $\geq 0,25\%$ se mantuvieron en el análisis, ya que se ha descrito que para amplicones del VHB de 400-500 nt de longitud, un nivel de 0,25% abundancia, asegura, con una elevada confianza, que los haplotipos analizados no son erróneos ⁶⁶.

El sistema de filtrado de las lecturas obtenidos se ha llevado a cabo en el entorno de programación de código abierto R ³⁶¹, utilizando las librerías Bioconductor ³⁶² y Biostrings ³⁶³ para el alineamiento de las lecturas obtenidos con el patrón de referencia. Este sistema de filtrado consistió en una serie de pasos (Figura 43):

1) En primer lugar, se llevó a cabo el demultiplexado de las lecturas, para separarlos, según la muestra a la que pertenecen a partir de las secuencias MID. Se toleró un máximo de una diferencia en la identificación del MID. Después se separó según la secuencia M13 directa y reversa en la que se toleraron hasta cuatro diferencias. A continuación, se separaron las lecturas por *primers* específicos. Aquí se toleraron entre dos y tres diferencias. Estos pasos se llevaron a cabo mediante el alineamiento de cada lectura con la secuencia de cada uno de los *primers* utilizados, Finalmente se obtuvo un fichero fasta por cada combinación de MID, directa y reversa y *primer* específico, completando el demultiplexado de las muestras iniciales. A medida que se fueron demultiplexando se fue eliminando las secuencias que permitieron la identificación.

2) A continuación, se llevó a cabo un filtro de calidad de las secuencias en el que se eliminaron las lecturas que contenían indeterminaciones y los que no cubrían la longitud entera del amplicón. Las lecturas restantes que correspondieron a una misma secuencia se colapsaron en un solo haplotipo. Se asoció una frecuencia a cada haplotipo en función de la frecuencia con que fue observada su secuencia en cada QS. Cada haplotipo correspondía a una variante concreta de la QS del VHB. Los haplotipos representados por una sola secuencia se eliminaron, ya que no es posible distinguirlos de un artefacto debido a las PCRs o al propio proceso de secuenciación.

3) Posteriormente los diferentes haplotipos fueron alineados en parejas respecto del haplotipo dominante, y se eliminaron aquellos haplotipos que presentaron inserciones-delecciones (indels) respecto del haplotipo dominante.

4) Solo se aceptaron los haplotipos comunes en las cadenas directa y reversa, presentes en abundancias de al menos 0,1%. Finalmente, los haplotipos comunes en ambas cadenas se consideraron haplotipos de consenso y sus recuentos finales se calcularon como la suma de las lecturas observados en las cadenas directa y reversa. Los haplotipos con abundancias por debajo del 0,25% fueron excluidos.

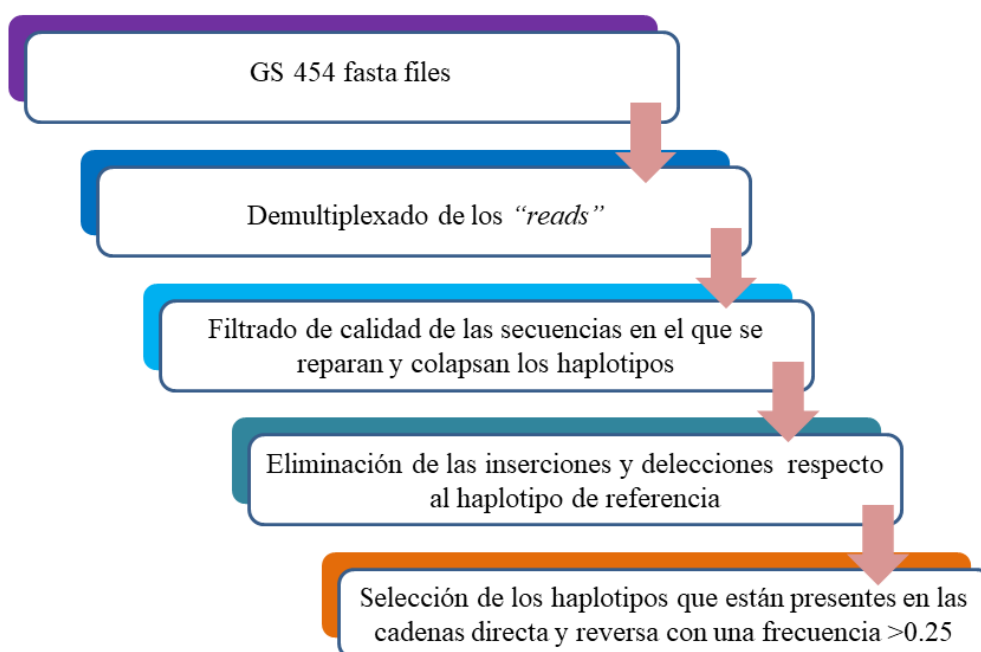


Figura 43 Flujo de trabajo del procesamiento de los datos de UDPS para obtener el muestreo de las QS virales reales del VHB.

Una vez concluido el filtrado bioinformático y obtenidos los haplotipos nucleotídicos, éstos se tradujeron a aminoácidos. Los haplotipos con la misma secuencia fueron recolapsados y sus frecuencias fueron actualizadas.

3.2.2. Genotipado de los haplotipos obtenidos:

Una vez concluido el filtrado bioinformático y obtenidos los haplotipos nucleotídicos, estos se genotiparon mediante un análisis filogenético. El genotipado mediante secuenciación masiva nos permitió poder conocer la verdadera complejidad de la mezcla de genotipos, ya que esto no es posible, mediante secuenciación por Sanger debido a sus limitaciones técnicas. Esto nos conducirá a una fotografía más real de cuantos genotipos y en qué proporción están representados en el conjunto de secuencias de todos los pacientes, que es la base sobre la cual analizaremos la conservación.

Primero, se seleccionaron 102 secuencias del genoma del VHB de longitud completa representativas de los genotipos del VHB (A a H), obtenidas de la base de datos GenBank (National Center for Biotechnology Information, USA), de los cuales se extrajo la región genómica del VHB desde el nt 1255 hasta el 1611 (Figura 44). En esta región se determinaron las distancias genéticas máximas entre secuencias del mismo genotipo y las distancias genéticas mínimas entre secuencias de diferentes genotipos, con el fin de establecer un umbral de identidad de secuencia para distinguir así entre diferentes genotipos. Dicho umbral se fijó en una coincidencia mínima entre 2 secuencias del 96% para ser consideradas del mismo genotipo. La secuencia “master” de cada grupo (el haplotipo más abundante) se incluyó en un análisis filogenético con las mismas secuencias extraídas de las 102 secuencias del genoma del VHB de longitud completa. Luego, todos los haplotipos individuales se separaron en diferentes archivos Fasta de acuerdo con el genotipo asignado a la secuencia master.

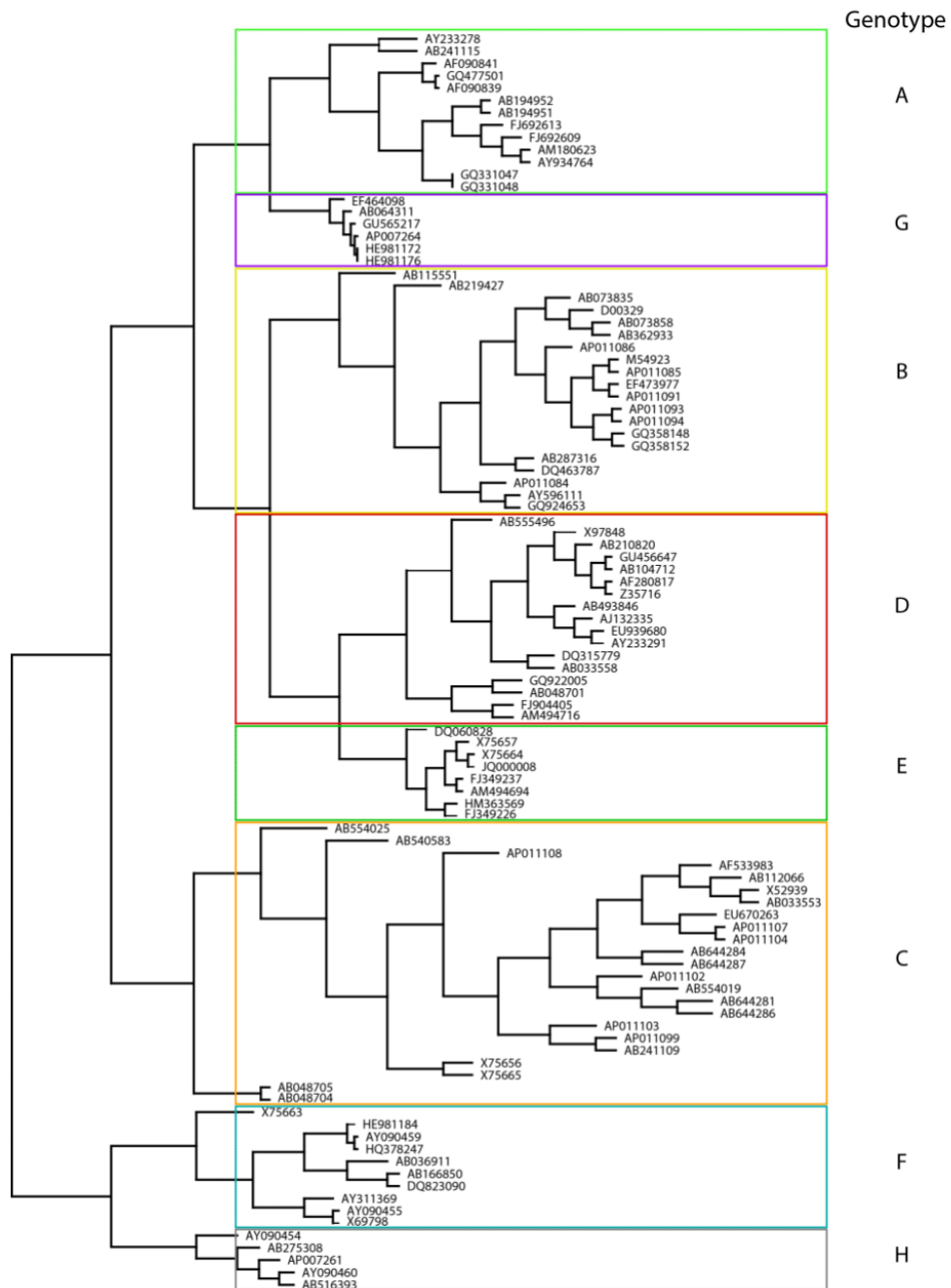


Figura 44. Árbol filogenético obtenido del análisis de la región genómica del VHB desde el nt 1255 hasta el 1611 extraída de las 102 secuencias de genoma de VHB de longitud completa representativas de los A a H genotipos del VHB, obtenidas de la base de datos GenBank (National Center for Biotechnology Information, USA).

3.2.3. Análisis de la conservación en la región analizada:

La conservación de las secuencias de QS se determinó calculando el contenido de información, basado en la incertidumbre de Shannon (ver en el Capítulo 1 el apartado

2.1.3. *Estudio de la conservación de la QS*), en cada posición en un alineamiento múltiple de todos los haplotipos diferentes obtenidos en los pacientes.

Se llevó a cabo un análisis basado en *sliding windows* para localizar los fragmentos de 25 nt (longitud mínima de un siRNA ³⁶⁴) o 10 aa con mayor contenido de información (más conservados) dentro de los alineamientos múltiples. Este análisis usa ventanas de 25 nt (o 10 aa) comenzando desde la primera posición en los alineamientos múltiples y avanzando en pasos de 1 (nt o aa). Para cada ventana, el análisis calcula la media del IC de las posiciones incluidas dentro de ella (ver en el Capítulo 1 el apartado 2.1.3. *Estudio de la conservación de la QS*). Las medias del IC de las posiciones de los fragmentos del gen *HBX* más conservados mediante el análisis de *sliding windows* se representaron como logos de secuencia (ver en el Capítulo 1 el apartado 2.1.3. *Estudio de la conservación de la QS*), creados con el paquete de lenguaje R *motifStack* ³⁶⁵.

3.3 Análisis de las secuencias de NGS obtenidas y resultados de genotipado

Después de aplicar los filtros descritos en el apartado 3.2.1. *Filtraje de los datos de secuenciación masiva*, se obtuvieron 1.333.069 secuencias de las 27 muestras de suero con una mediana [IQR] de 2817 [3541,5-6358,5] secuencias.

Los resultados del análisis filogenético de las secuencias master de cada grupo en cada paciente y los 102 patrones del GenBank se muestran en la tabla 1 (Tabla 1 del Capítulo 3 apartado 6. *Material suplementario “Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy”*). En el 52% de los casos se encontraron algunos haplotipos correspondientes a genotipos diferentes a los identificados previamente por la secuenciación de Sanger, lo que evidencia una mezcla compleja de variantes genotípicas. Esto evidencia que hemos estudiado la conservación de la región 5' del *HBX* en un grupo de variantes genéticas heterogéneo, donde se encontraban representados los principales genotipos del VHB, lo que nos ha permitido obtener una imagen más aproximada de la variabilidad y conservación globales de esta región.

3.4 Conservación de la secuencia de nucleótidos del *HBX* en la región de interés

La conservación de la región de interés se estudió en alineamientos múltiples de los haplotipos tanto a nivel de nt como de aa mediante el análisis de *sliding windows* para resaltar las regiones híper-conservadas. En dicho análisis, no se consideró la frecuencia relativa de cada haplotipo, para que los resultados de conservación no se vieran afectados

por el *fitness* del haplotipo. De esta forma se evita que las variantes minoritarias, que podrían escapar a una posible terapia génica, queden enmascaradas en este análisis. Para obtener una imagen más detallada de la conservación, los *sliding windows* se implementaron en dos escenarios: en el primero se ha utilizado el enfoque de consenso, en el que el alineamiento múltiple se crea con la secuencia consenso de cada paciente resultante de todos los haplotipos que forman su QS (N = 27 secuencias). Esta aproximación nos da una idea de la conservación de la región de interés entre los aminoácidos o nucleótidos más representados en cada posición en los diferentes pacientes. En el segundo se ha utilizado el enfoque QS en que el alineamiento múltiple se produce a partir de todos los haplotipos de las QS en proporciones > 0,25% obtenidos en el total de las muestras (N = 720 secuencias). Esta aproximación nos da una idea de la conservación de la región de interés entre todas las variantes de las QS de los diferentes pacientes, asegurando, por lo tanto, la conservación de las regiones híper-conservadas identificadas en el máximo de variantes del VHB. No se han observado diferencias a la hora de superponer los resultados de los 2 escenarios (se observaron 2 regiones altamente conservadas con una IC media cercana a 2 bits en ambos) (figura 45). Los resultados informados a continuación se refieren al análisis en el escenario QS. La primera región híper-conservada identificada fue entre nt 1255 y 1286 (23 nt de longitud) (Figura 45). La segunda región híper-conservada consistió en 3 fragmentos nt conservados (1519-1543, 1545-1573 y 1575-1603: 25, 29 y 29 nt de longitud, respectivamente) que abarcan una macro-región entre nt 1519 y 1603 (85 nt) (Figura 45).

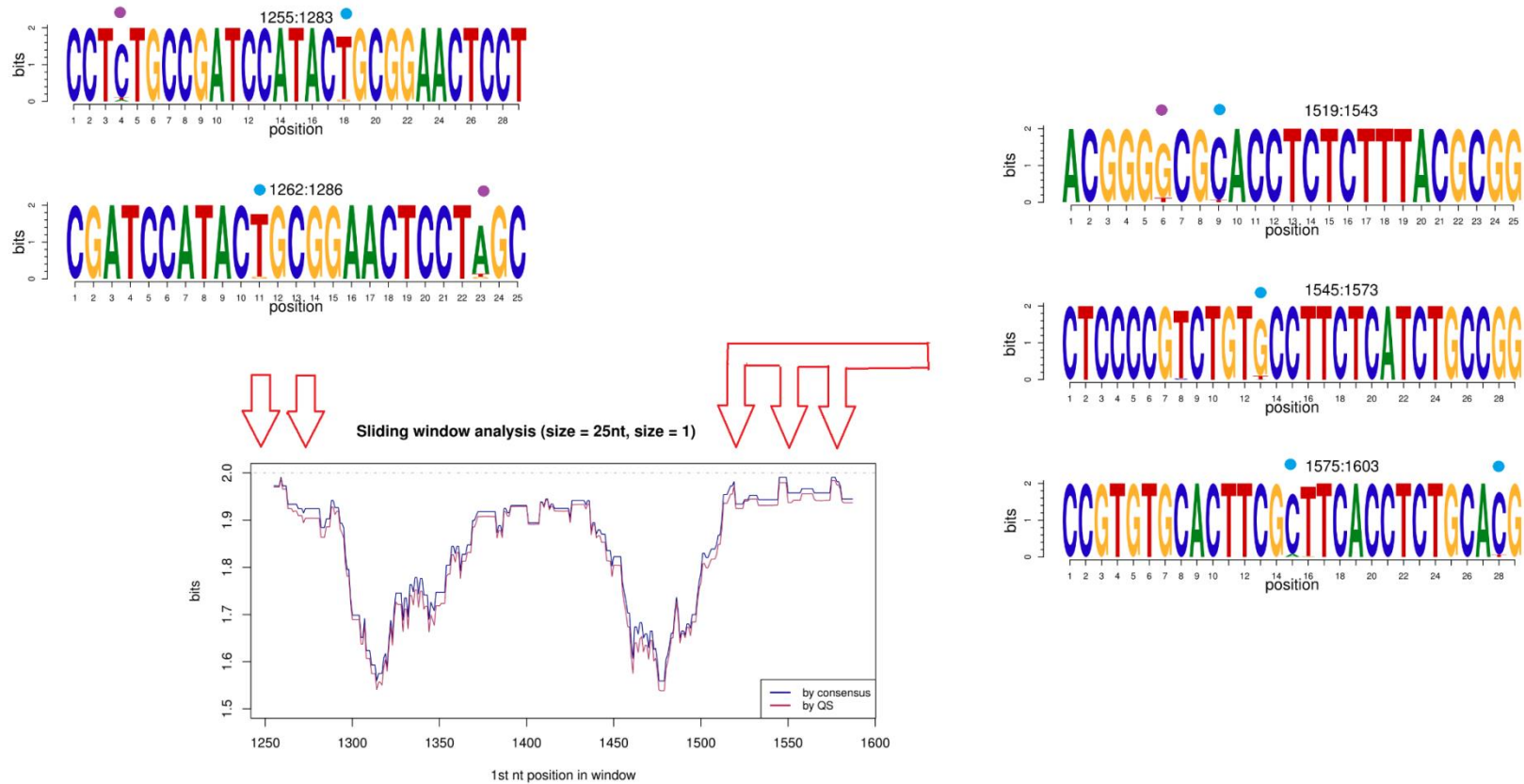


Figura 45 Análisis por *sliding windows* de la región de nucleótidos de interés en el *HBX* (nt 1255-1611). La línea violeta representa la IC media de los alineamientos múltiples de todos los haplotipos (en abundancia > 0,25%) en las QS de todos los pacientes (N = 720), mientras que la línea azul muestra la IC media obtenida a partir de los alineamientos múltiples del consenso de los diferentes haplotipos obtenidos en cada paciente (N = 27). Están señalados mediante flechas los picos de las ventanas de 25 nt que están más conservados. Las secuencias de las ventanas más conservadas, a su vez están representadas por los logos de secuencia del IC. Las posiciones de nucleótidos con un IC entre 1.6 y 1.8 (80% -90% de conservación máxima) están indicadas por círculos de color azul claro y aquellos con un IC de entre 1.4 y 1.6 bits (70% -80% de conservación máxima) por círculos violetas.

3.5 Conservación de la secuencia de aminoácidos HBx

Al fin de evaluar la conservación de la región del genoma viral estudiada también a nivel de la secuencia aminoacídica de la HBx, hemos analizado la conservación en los mismos 2 escenarios considerados para los alineamientos múltiples de haplotipos de nt (N = 27 secuencias para el consenso y N = 330 para el enfoque QS). Como se vio con las secuencias nt, no hubo diferencias cuando se superpusieron los 2 análisis (QS frente a consenso) (figura 46), en los que se resaltó una sola región altamente conservada. Una vez más, los resultados informados se refieren al análisis utilizando el enfoque QS.

Se identificó una región altamente conservada entre aa 63 y 76 (13 aa), que incluía una porción de un dominio *Kunitz-like* (ver en el Capítulo 1 el apartado 1.4.1.4. *ORF X y la proteína X*). Todos los aa mostraron una conservación cerca de 4 bits (100% de conservación máxima). Esta región en la HBx corresponde a una secuencia de nt híper-conservada entre las posiciones 1563 y 1602 (figura 45), incluida dentro de la región híper-conservada entre los nucleótidos 1519 y 1603 descrita en el apartado anterior. La primera región nt híper-conservada observada (nt 1255-1286) no se tuvo en cuenta en este análisis, ya que no corresponde a una región codificante y, por lo tanto, no se traduce en aa.

[Escriba aquí]

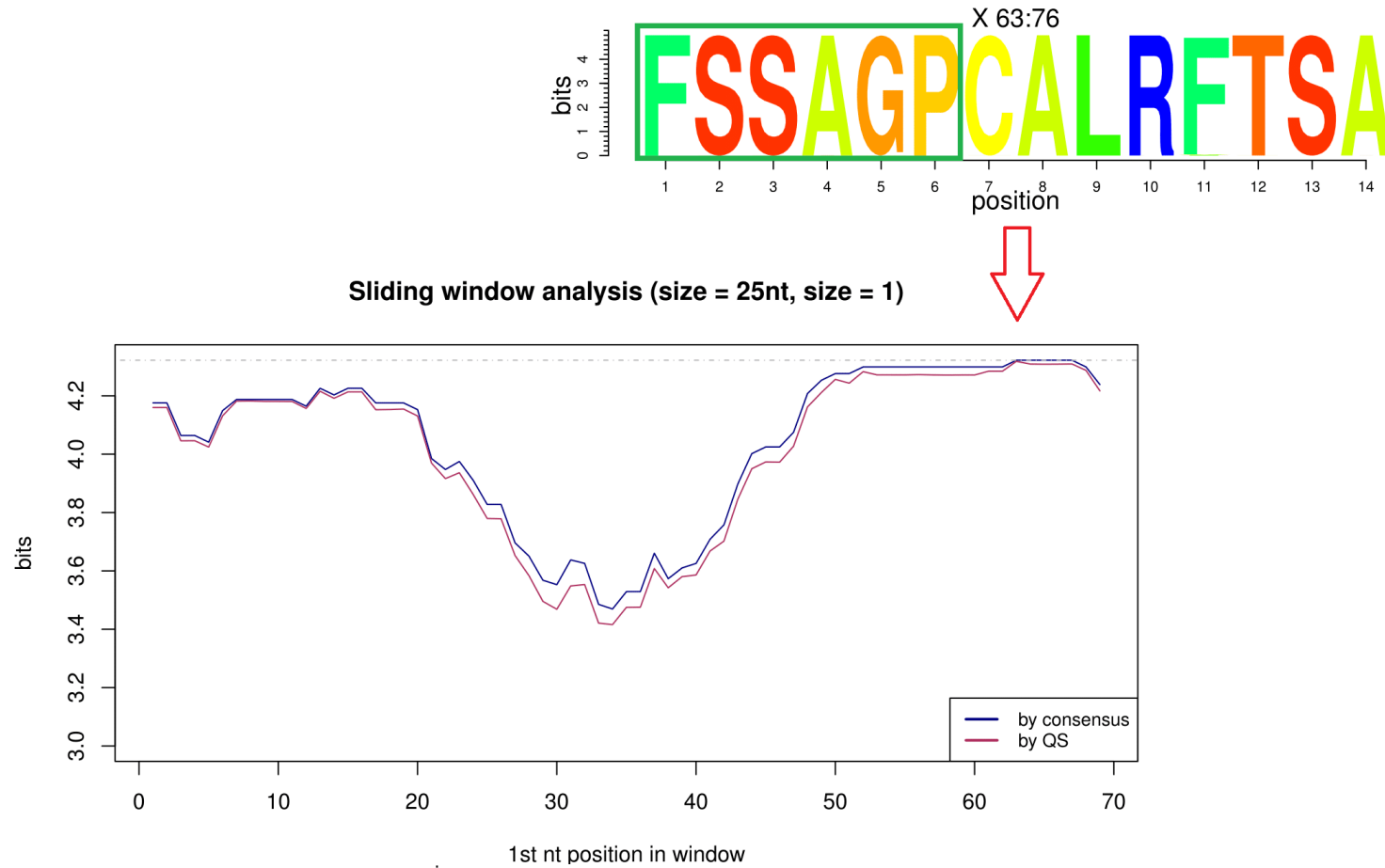


Figura 46 Sliding window de la secuencia de aminoácidos de la HBx codificada por la región del *HBX* estudiada (aa 1-79). La línea violeta representa la IC media de las alineaciones múltiples de todos los haplotipos (en abundancia > 0,25%) en las QS de todos los pacientes (N = 330), mientras que la línea azul muestra la IC media obtenida a partir de los alineamientos múltiples del consenso de los diferentes haplotipos obtenidos para cada paciente (N = 27). La ventana de 10 aa con mayor conservación está señalada mediante una flecha. La secuencia de esta ventana está representada en los logos de secuencia del IC. Los aminoácidos que pertenecen a la porción de dominio *Kunitz-like* están marcados en verde.

4. Conclusiones

1. Se encontraron 2 regiones híper-conservadas en la región 5' del *HBX*, la primera en la región no codificante de los transcritos del *HBX*, y la segunda en su región codificante, que se conservó tanto a nivel de nt como a nivel de aa.
2. Estas regiones híper-conservadas podrían ser potenciales candidatas para ser usadas como dianas en terapias génicas dirigidas.
3. Creemos que tienen particular interés para su uso en el diseño de un posible sistema de siRNA, del que se están realizando estudios preliminares de la eficacia de los posibles siRNA (ver *I. Anexo I*)³⁶⁶.

5. Manuscrito completo

Submit a Manuscript: <http://www.f6publishing.com>

World J Gastroenterol 2018 May 14; 24(18): 0000-0000

DOI: 10.3748/wjg.v24.i18.0000

ISSN 1007-9327 (print) ISSN 2219-2840 (online)

ORIGINAL ARTICLE

Basic Study

Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy

Carolina González, David Tabernero, Maria Francesca Cortese, Josep Gregori, Rosario Casillas, Mar Riveiro-Barciela, Cristina Godoy, Sara Sopena, Ariadna Rando, Marçal Yll, Rosa Lopez-Martinez, Josep Quer, Rafael Esteban, Maria Buti, Francisco Rodríguez-Frías

Carolina González, David Tabernero, Maria Francesca Cortese, Rosario Casillas, Cristina Godoy, Sara Sopena, Ariadna Rando, Marçal Yll, Rosa Lopez-Martinez, Francisco Rodríguez-Frías, Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

Maria Francesca Cortese, Josep Gregori, Rosario Casillas, Cristina Godoy, Marçal Yll, Josep Quer, Liver Unit, Liver Disease Laboratory-Viral Hepatitis, Vall d'Hebron Institut Recerca-Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

David Tabernero, Josep Gregori, Mar Riveiro-Barciela, Sara Sopena, Josep Quer, Rafael Esteban, Maria Buti, Francisco Rodríguez-Frías, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid 28029, Spain

Josep Gregori, Roche Diagnostics SL, Sant Cugat del Vallès 08174, Spain

Mar Riveiro-Barciela, Rafael Esteban and Maria Buti, Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

ORCID number: Carolina González (0000-0002-0169-5874); David Tabernero (0000-0002-1146-4084); Maria Francesca Cortese (0000-0002-4318-532X); Josep Gregori (0000-0002-4253-8015); Rosario Casillas (0000-0002-6758-6734); Mar Riveiro-Barciela (0000-0001-9309-2052); Cristina Godoy (0000-0001-5037-1916); Sara Sopena (0000-0002-3309-5486); Ariadna Rando (0000-0003-4555-7286); Marçal Yll (0000-0002-7030-3360); Rosa Lopez-Martinez (0000-0002-8450-6986); Josep Quer (0000-0003-0014-084X); Rafael Esteban (0000-0001-5280-392X); Maria Buti (0000-0002-0732-3078); Francisco Rodríguez-Frías (0000-0001-9058-4641).

Author contributions: Rodríguez-Frías F designed the research;

Cortese MF coordinated the research; González C and Tabernero D equally contributed to design the experiments; González C, Casillas R, Godoy C, Sopena S and Rando A performed the experiments; González C, Tabernero D and Gregori J analyzed data acquired during the experiments and interpreted the results; González C, Tabernero D, Cortese MF and Riveiro-Barciela M drafted the manuscript; Yll M, Lopez-Martinez R, Buti M, Quer J, Esteban R and Rodriguez-Frías F critically reviewed the manuscript;

Supported by the Instituto de Salud Carlos III, No. PI15/00856; and the European Regional Development Fund (ERDF), No. grant PI15/00856.

Institutional review board statement: The study was reviewed and approved by the Clinical Research Ethics Committee (CEIC) of Hospital Universitari Vall d'Hebron.

Conflict-of-interest statement: Josep Gregori is an employee of Roche Diagnostics, SL.

Data sharing statement: No additional data are available.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Manuscript source: Invited manuscript

Correspondence to: Maria Francesca Cortese, PhD, Research Scientist, Liver Pathology Unit, Departments of Biochemistry and Microbiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Passeig Vall d'Hebron 119-129, Barcelona 08035, Spain. maria.cortese@vhir.org
Telephone: +34-932-746000/6858

González C *et al.* hyper-conserved regions in HBX gene

Received: February 16, 2018
 Peer-review started: February 17, 2018
 First decision: March 9, 2018
 Revised: March 26, 2018
 Accepted: May 6, 2018
 Article in press: May 6, 2018
 Published online: May 14, 2018

Abstract

AIM

To detect hyper-conserved regions in the hepatitis B virus (HBV) X gene (*HBX*) 5' region that could be candidates for gene therapy.

METHODS

The study included 27 chronic hepatitis B treatment-naïve patients in various clinical stages (from chronic infection to cirrhosis and hepatocellular carcinoma, both HBeAg-negative and HBeAg-positive), and infected with HBV genotypes A-F and H. In a serum sample from each patient with viremia > 3.5 log IU/mL, the *HBX* 5' end region [nucleotide (nt) 1255-1611] was PCR-amplified and submitted to next-generation sequencing (NGS). We assessed genotype variants by phylogenetic analysis, and evaluated conservation of this region by calculating the information content of each nucleotide position in a multiple alignment of all unique sequences (haplotypes) obtained by NGS. Conservation at the HBx protein amino acid (aa) level was also analyzed.

RESULTS

NGS yielded 1333069 sequences from the 27 samples, with a median of 4578 sequences/sample (2487-9279, IQR 2817). In 14/27 patients (51.8%), phylogenetic analysis of viral nucleotide haplotypes showed a complex mixture of genotypic variants. Analysis of the information content in the haplotype multiple alignments detected 2 hyper-conserved nucleotide regions, one in the *HBX* upstream non-coding region (nt 1255-1286) and the other in the 5' end coding region (nt 1519-1603). This last region coded for a conserved amino acid region (aa 63-76) that partially overlaps a Kunitz-like domain.

CONCLUSION

Two hyper-conserved regions detected in the *HBX* 5' end may be of value for targeted gene therapy, regardless of the patients' clinical stage or HBV genotype.

Key words: Hepatitis B virus; Hepatitis B X gene; Hepatitis B X protein; Gene therapy; Next-generation sequencing; HBV conserved regions; Small interference RNA

© The Author(s) 2018. Published by Baishideng Publishing Group Inc. All rights reserved.

Core tip: Hepatitis B virus (HBV) is not cured with classic treatments, and liver disease can progress by persistence

and expression of covalently-closed circular DNA. Gene therapy with small interference RNA may be an effective approach to ensure inhibition of viral expression and disease progression, and hepatitis B virus X gene (*HBX*) transcripts could be optimal targets for this therapy. This study includes patients with different HBV genotypes and clinical stages to cover many clinical and virological situations. Using next-generation sequencing, we found two hyper-conserved *HBX* regions, candidates for small interference RNA therapy, which could enable pan-genotypic inhibition of HBV expression, regardless of the patients' disease status.

González C, Tabernero D, Cortese MF, Gregori J, Casillas R, Casillas R, Riveiro-Barciela M, Godoy C, Sopena S, Rando A, Yll M, Lopez-Martinez R, Quer J, Esteban R, Buti M, Rodríguez-Frias F. Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy. *World J Gastroenterol* 2018; 24(18): 0000-0000 Available from: URL: <http://www.wjgnet.com/1007-9327/full/v24/i18/0000.htm> DOI: <http://dx.doi.org/10.3748/wjg.v24.i18.0000>

INTRODUCTION

Despite the efficacy of preventive vaccines, an estimated 257 million people are living with chronic hepatitis B virus infection (CHB) and more than 880000 people die each year of hepatitis B virus (HBV)-related complications such as cirrhosis and hepatocellular carcinoma (HCC) (WHO report, July 2017).

HBV is an enveloped DNA virus with partially double-stranded circular DNA. HBV replication requires RNA intermediate and the activity of a reverse transcriptase. This implies a high probability that genetic mutations will occur, as the reverse transcriptase lacks 3' to 5' proofreading activity, leading to a viral mutation rate of 10^{-4} to 10^{-5} substitutions/site/year, similar to that observed for RNA viruses^[1]. Inter- and intra-genotype recombination events can further increase HBV variability^[2]. Hence, HBV circulates as a complex mixture of genetic variants, known as a quasispecies^[3], that enables the virus to escape from the host's immune system, antiviral treatment, and vaccination, thereby promoting progression to CHB. Furthermore, the mutational profile is closely associated with HBV genotype, and the genotype is associated with differing effectiveness of the treatments used and outcomes of the infection^[4,5].

The main therapeutic approach for HBV infection is based on inhibition of the viral polymerase by the action of nucleotide analogues, whose goal is to improve the patients' quality of life and prolong survival by preventing progression of the disease^[6]. However, HBV cannot be completely eradicated with these drugs because the viral intermediate known as covalently closed circular DNA (cccDNA) can persist within the nucleus of HBV-infected liver cells. cccDNA interacts

with histone and non-histone proteins, including viral proteins such as the core and X protein (HBx), and forms a minichromosome that permits transcription of HBV genes^[7], including pregenomic RNA, the precursor of *de novo* viral DNA genomes. Because cccDNA persists, it constitutes a viral reservoir that could promote reactivation of the infection after treatment interruption^[8]. Within this challenging scenario, research has been aimed at deeply investigating the host-virus interactions to better understand the mechanisms that establish persistent HBV infection and to find new therapeutic targets that can cure it.

In this line, new treatment approaches are currently under development^[9], with gene therapy being a promising option. Homing endonucleases, such as zinc-finger endonucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and RNA-guided clustered regulatory interspaced short palindromic repeats associated with the Cas endonuclease family (CRISPR/Cas), can cleave selected sequences in cccDNA, resulting in disruption of the gene due to nonspecific DNA repair with consequent elimination of the viral minichromosome^[10,11]. However, systematic random integration of the viral genome in the host genome could represent a strong limitation to this strategy. Indeed, the activity of this "molecular scissors", although sequence-specific, could entail a potential risk of damage for the human genes close to the viral site of integration.

Another promising gene therapy consists in silencing specific genes at the post-transcriptional level through a sequence-specific interaction between an mRNA target and small interfering RNA (siRNA)^[12]. With this approach, various regions of the viral mRNA sequence can be targeted, including non-coding regions, without affecting the host DNA. Although these therapies show promise, the high variability of HBV and the association between this variability and the patients' clinical outcome suggests that it may be important to find a highly conserved target to guarantee their efficacy.

A good candidate for targeted gene therapy could be the HBx protein, encoded by the HBV X gene (*HBX*). This pleiotropic and multifunctional protein transactivates the expression of the viral genes. Together with the HBV core protein (HBc), HBx attaches to the cccDNA structure and is crucial for HBV replication^[7]. In addition, this protein interacts with several cell signaling pathways and genes, thus affecting many cellular activities^[13-15]. Due to its wide range of activity, HBx plays a key role in the pathogenesis of HBV infection and disease progression, and is strongly associated with HCC. Hence, it could be an optimal target for a hypothetical curative therapy for HBV infection.

The *HBX* gene, nucleotides (nt) 1374-1838, contains important regulatory elements^[16,17]. The coded protein is comprised of 2 domains. The N-terminal domain [amino acid (aa) 1-50, encoded by the 5' end of the gene] acts as negative regulator of the HBx transactivation

function, which resides in the C-terminal domain (aa 51-154, encoded by the 3' end). Interestingly, a significant presence of multiple variants with deletions and/or insertions (indels) has been found in the 3' end of *HBX*^[18-20]. Considering this variability, the 3' coding region of the X gene would be ruled out as a possible therapeutic candidate^[21]. However, the conservation at 5' end of *HBX* and its potential for use as a gene therapy target remains unexplored. To silence *HBX* at the post-transcriptional level, the non-coding region included in *HBX* transcripts, upstream of the coding region, should also be considered. The *HBX* gene is located near the co-terminal 3' end; hence, all HBV mRNAs produced during the infection include this sequence (Figure 1). Consequently, by targeting *HBX* transcripts at the coding or non-coding level interference with expression of all the viral proteins could be achieved.

The aim of this study was to determine the conservation of a region of the HBV genome encompassing the *HBX* 5' coding region and upstream non-coding region (included in all HBV transcripts) in samples from HBV-infected patients in various clinical stages and with different viral genotypes. The ultimate objective was to find hyper-conserved regions that might be feasible targets for gene therapy, which could be used whatever the patient's clinical status or HBV genotype.

MATERIALS AND METHODS

Patients and samples

From a cohort of 46 well-characterized CHB patients attending the outpatient clinic of Vall d'Hebron University Hospital (Barcelona, Spain), we selected a group of 27 patients in various clinical stages and with different viral genotypes. The sample included 17 HBeAg-negative patients (3 with chronic infection and 14 chronic hepatitis, 2 of them with cirrhosis and 1 with HCC), and 10 HBeAg positive patients (2 with chronic infection and 8 with chronic hepatitis, 3 of them with cirrhosis and 2 with HCC, characterized according to the latest EASL guidelines^[6]), infected with several HBV genotypes: 5 A, 1 B, 7 C, 8 D, 2 E, 3 F, 1 H (Table 1).

All 27 patients were treatment-naïve, tested negative for hepatitis D virus (HDV), hepatitis C virus (HCV), and human immunodeficiency virus (HIV), and had a serum sample with viremia levels > 3.5 logIU/mL, the sensitivity limit of the PCR to amplify the study region (described below). The study was approved by the Ethics Committee of Vall d'Hebron Research Institute, and all patients signed a consent form to participate.

Serological and virological determinations

HBV serological markers (HBsAg, HBeAg, and anti-HBe) and anti-HCV antibodies were tested using commercial chemiluminescent assays on a COBAS 8000 analyzer (Roche Diagnostics, Rotkreuz, Switzerland). Antibodies against HDV were tested using the HDV Ab kit (Dia.Pro Diagnostic Bioprobes, Sesto San Giovanni, Italy), and

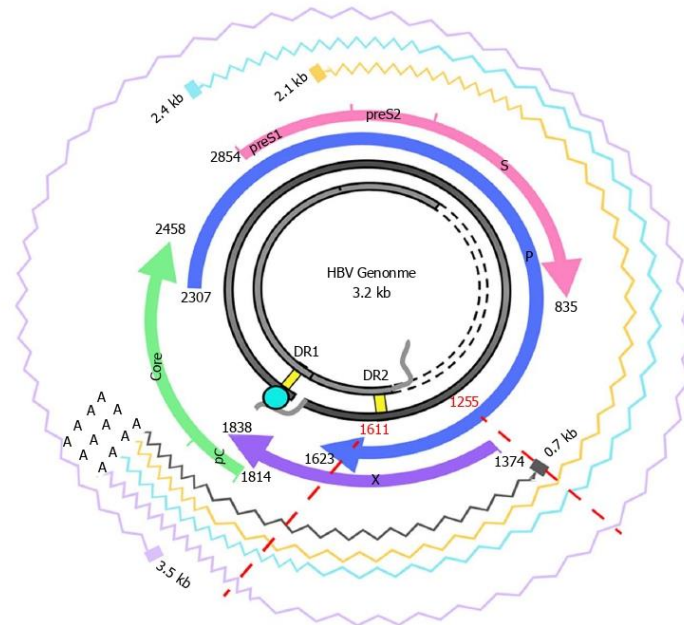


Figure 1 Hepatitis B virus genome and transcripts. The figure shows the HBV genome (in grey), with the DR1 and DR2 (direct repeat) regions, necessary for viral DNA synthesis. The viral ORFs are highlighted with colored arrows, and nucleotide positions are reported. Wavy lines show the various HBV transcripts: the 3.5-kb transcript, corresponding to the pregenomic RNA, which is translated to the core and polymerase and later subjected to reverse-transcription in the viral capsid, or to the precore/core transcript, which is translated to the precore protein; the 2.4-kb and 2.1-kb transcripts, which are translated, respectively, to large and medium/small HBsAg; and the 0.7-kb transcript, which is translated to the HBx protein. The region analyzed in this study and its corresponding nt positions are indicated by red dashed lines. Note that the region of interest is included in all the viral transcripts. HBV: Hepatitis B virus; ORF: Open reading frame.

anti-HIV antibodies were tested by the Liaison XL murex HIV Ab/Ag kit (DiaSorin, Saluggia, Italy). HBV-DNA was quantified by real-time PCR with a detection limit of 10 IU/mL (COBAS 6800, Roche Diagnostics). HBV genotypes in the region of interest were determined by Sanger sequencing and by phylogenetic analysis with the same regions extracted from 102 full-length HBV genome sequences representative of HBV genotypes A to H, obtained from GenBank (Supplementary Table 1 and Supplementary Figure 1).

Amplification of the region of interest

Analysis of the *HBX* gene encompassed nt 1255 to nt 1611, a region included in the 5' end of all the viral transcripts. It covered a non-coding upstream region (nt 1255-1373) and the 5' end of the *HBX* coding region (nts 1374-1611), encoding aa 1 to 79 of HBx.

HBV DNA was extracted from 500 µL of serum with the QIAamp UltraSens Virus Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Molecular amplification was performed by nested PCR. The first PCR round used primers carrying the universal adaptor M13 (underlined sequence) in their

5' end (forward 5'-GTTGTAAAACGACGGCCAGTAT GCGTGGAACCTTTGTGGCT-3' and reverse 5'-CACAGGAAACAGCTATGACCATGGGCGTTCACGGTGGTCT-3') using the following protocol: 95 °C for 2 min, followed by 30 cycles of 95 °C for 15 s, 60 °C for 20 s, and 72 °C for 15 s, and finally, 72 °C for 3 min. The second PCR round was performed using the primers: forward 5'-CGTATCGCTCCTCGCGCCATCAG-MID-GTTGTAAAACGACGGCCAGT-3' and reverse 5'-CTATGCGCCTTGCACGCCCGCTCAG-MID-CACAGGA AACAGCTATGACC-3'. These primers included the 2 adaptors for the ultra-deep pyrosequencing system at their 5' ends, followed by a unique identifier multiplex identifier sequence (MID), which enabled grouping the sequences for each sample/patient, and the same M13 universal adaptor sequences as those used in the first PCR in the 3' ends. This second amplification protocol comprised one denaturation step of 95 °C for 2 min, followed by 20 cycles of 95 °C for 15 s, 60 °C for 20 s, and 72 °C for 15 s, and finally, 72 °C for 3 min. All PCR steps were performed using high-fidelity Pfu Ultra II DNA polymerase (Stratagene, Agilent Technologies, Santa Clara, United States). The final PCR products (amplicons)

Table 1 Main clinical and virological characteristics of the hepatitis B virus infected patients enrolled

Patient	Age	Sex	Origin	Clinical stage	HBeAg	HBV DNA (log IU/mL)	ALT (IU/L)	Genotype ¹
1	27	M	Sub-Saharan	Chronic hepatitis	Negative	7.8	170	E
2	31	M	Asian	Chronic hepatitis	Negative	6.6	103	C
3	51	M	Caucasian	Chronic hepatitis	Negative	6.6	262	D
4	28	F	Caucasian	Chronic hepatitis	Negative	7.9	126	D
5	47	F	Caucasian	Chronic hepatitis	Negative	6.3	170	D
6	37	F	Caucasian	Chronic hepatitis	Negative	4.5	53	D
7	37	M	Caucasian	Chronic hepatitis	Negative	4.5	33	F
8	38	M	Caucasian	Chronic hepatitis	Negative	5.0	29	D
9	46	M	Caucasian	Chronic hepatitis	Negative	5.8	88	F
10	46	F	Caucasian	Chronic hepatitis	Negative	5.3	23	H
11	71	F	Caucasian	Chronic hepatitis	Negative	6.2	87	F
12	51	M	Asian	Chronic hepatitis	Negative	5.7	435	C
13	52	M	Caucasian	Chronic infection	Negative	4.4	18	A
14	40	F	Caucasian	Chronic infection	Negative	4.2	29	D
15	33	M	Asian	Chronic infection	Negative	4.3	25	D
16	63	M	Hispanic	Cirrhosis	Negative	4.0	16	A
17	53	M	Caucasian	Cirrhosis/HCC	Negative	3.7	36	A
18	35	M	Sub-Saharan	Chronic hepatitis	Positive	5.7	36	E
19	37	M	Caucasian	Chronic hepatitis	Positive	8.4	32	C
20	45	M	Caucasian	Chronic hepatitis	Positive	5.6	35	A
21	29	F	Asian	Chronic hepatitis	Positive	6.9	355	B
22	28	M	Asian	Chronic hepatitis	Positive	> 8.0	341	C
23	28	M	Asian	Chronic infection	Positive	8.7	24	C
24	28	F	Asian	Chronic infection	Positive	8.8	22	C
25	55	F	Caucasian	Cirrhosis	Positive	5.4	73.9	A
26	82	F	Caucasian	Cirrhosis/HCC	Positive	4.8	24	C
27	64	M	Caucasian	Cirrhosis/HCC	Positive	6.3	45	D

¹Genotype determined by Sanger sequencing of the X region (same region as was analyzed by next-generation sequencing). ALT: Alanine aminotransferase; HBV: Hepatitis B virus; HBeAg: Hepatitis B e antigen; M: Male; F: Female; HBV: Hepatitis B virus.

were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter, Beverly, United States). The quality of the purified products was verified with the Agilent 2200 TapeStation System using the D1000 ScreenTape kit (Agilent Technologies, Waldbronn, Germany).

Next-generation sequencing and sequence quality control

Purified DNA from each sample was quantified using the Quant-IT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific - Life Technologies, Austin, United States), and a pool was formed in which each amplicon was adequately represented in the analysis. The pool was sequenced by next-generation sequencing (NGS) based on ultra-deep pyrosequencing (UDPS) on the GS-Junior or GS FLX platforms (454 Life sciences-Roche, Branford, United States), following the manufacturer's protocol. The two platforms are reported to be interchangeable^[22].

The sequences (reads) obtained after UDPS underwent an in-house bioinformatics filtering procedure, based on scripts developed in R language^[23], as previously described by our group^[22]. Briefly, the sequences were assigned to each patient (demultiplexed) according to their specific MID, and primers were trimmed. After a general quality filter step, reads with the same nt sequence were collapsed into haplotypes (unique sequences covering the full amplicon observed on the

clean set of sequences). Only haplotypes common to the forward and reverse strands and present in abundances of at least 0.1% were accepted; their final frequencies were calculated as the sum of reads observed in each strand. Finally, haplotypes with abundances below 0.25% were excluded.

To analyze the aa sequence of HBx, all individual nt haplotypes from each patient were translated into aa sequences in the *HBX* gene open reading frame (ORF), which was translated from frame 2. In the fragment analyzed (nt 1255-1611) this ORF expanded from nt 1374 to 1611, encoding aa 1 to 79 of the HBx protein. The upstream sequence was not translated, as it corresponded to a non-coding region whose sequence is included in the *HBX* transcripts. Once translated, identical aa sequences were recollapsing into aa haplotypes and their frequencies were updated.

Genotyping of the region haplotypes

The genotype of the nt haplotypes obtained by UDPS was determined by discriminant analysis with the same regions extracted from the 102 full-length patterns used for Sanger sequencing (Supplementary Table 1 and Supplementary Figure 1). We determined the maximum genetic distances between sequences from the same HBV genotype in this region and the minimum genetic distances between sequences from different genotypes,

in order to set a sequence identity threshold: sequences with an identity above this threshold were clustered together. Genotyping of each cluster centroid was done by distance-based discriminant analysis (DB rule)^[24,25], which takes into account the inter- and intra-class variability of all genotypes. Genetic distances were computed according to the Kimura-80 model^[26].

Conservation analysis

Sequence conservation was determined by calculating the information content (IC) of each position in a multiple alignment of all the different sequences found in the patients. This analysis, based on Shannon's uncertainty, was done for a multiple alignment of nt and aa sequences, and is defined as^[27]:

$$IC_j(\text{nt}) = \log_2(4) - \sum_{i=1}^4 p_{ij} \log_2(p_{ij})$$

$$IC_j(\text{aa}) = \log_2(20) - \sum_{i=1}^{20} p_{ij} \log_2(p_{ij})$$

where j stands for the j -th position in the alignment, i runs over the 4 nucleotides (or over the 20 aa), and p_{ij} is the frequency of the i -th nucleotide (or aa) in the j -th alignment position. IC ranges from 0, indicating maximum uncertainty or variability, to $\log_2 4$ (i.e., 2 bits) for nt or $\log_2 20$ (i.e., 4.32 bits) for aa, indicating maximum information or conservation.

When considering variability in human genetics, a mutation is commonly considered fixed if it is found in at least 1% of the population^[28]. However, in viral quasispecies, variants can be present at any abundance in a patient, and the limit for defining a fixed mutation has not yet been established. Taking that into account, we considered two scenarios providing limiting values in our analysis. In the first scenario, we only included the most abundant nucleotide at each position in each patient (consensus approach). The IC values computed in this way would be the upper limit of conservation. In the second scenario, we included all variants in the haplotypes from each patient that were present at abundance greater than 0.25% (quasispecies approach). The IC values computed in this way would be the lower limit of conservation.

Sliding window analysis was then carried out to locate the fragment of at least 25 nt or 10 aa (which corresponds to the length of a possible target for siRNA therapy) with the highest IC within the multiple alignments. This analysis uses windows of 25 nt (or 10 aa) starting from the first position in the multiple alignments and moves forward in steps of 1 (nt or aa). For each window, the analysis computes the mean IC of each position within the window. In addition, the results are represented as sequence logos created using the R language package motifStack^[27].

The bioinformatics methods used in this study were reviewed by Dr. Josep Gregori from the Liver Disease-Viral Hepatitis Laboratory of Vall d'Hebron Hospital (Barcelona, Spain), CIBERehd research group, and

Roche Diagnostics SL.

RESULTS

Analysis of the NGS sequences obtained and genotyping results

After applying the quality filters, 1333069 sequences were obtained from the 27 serum samples, yielding a median (IQR) of 4578 (2478-9279) sequences per patient.

In the region from nt 1255 to 1611 extracted from the 102 full-length HBV genome sequences from GenBank, analysis of the maximum genetic distance within the same genotype (data not shown) resulted in a sequence identity threshold of 96%. Therefore, for each patient, haplotypes with a sequence identity > 96% were clustered together and were considered to belong to the same HBV genotype. Results of the phylogenetic analysis of master sequences from each cluster in each patient and the 102 GenBank patterns are shown in Table 2. Genotype D nt haplotypes were the most frequent in our patients, followed by genotypes C, A, E, F, B, and H. None of the patients included showed genotype G haplotypes. Moreover, in 14/27 cases (51.8%), some haplotypes were found corresponding to different genotypes than those previously identified by Sanger sequencing, thus yielding a complex mixture of genotypic variants.

Conservation of the HBX nucleotide sequence in the region of interest

The region of interest was studied in multiple nt alignments of the entire quasispecies in order to highlight the most highly conserved regions. Sliding windows analysis was implemented in two scenarios: using the consensus approach ($n = 27$ sequences) and using the quasispecies approach ($n = 720$ sequences). Of note, the relative frequency of each haplotype was not considered in the multiple alignments, so that the conservation results would not be influenced by haplotype fitness. As no differences were seen when the analyses by the 2 approaches were superimposed (2 highly conserved regions with a mean IC near 2 bits were observed in both; Figure 2), the results reported below all refer to the analysis in the quasispecies scenario.

The first hyper-conserved region identified was between nt 1255 and 1286 (23 nt in length) (Figure 3A). Most of the nucleotide positions showed high conservation, yielding IC values near 2 bits (100% maximum conservation), with the exception of position 1272 which showed an IC between 1.6 and 1.8 bits (80%-90% maximum conservation) and positions 1258 and 1284, with an IC between 1.4 and 1.6 bits (70%-80% maximum conservation).

The second hyper-conserved region consisted of 3 conserved nt fragments (1519-1543, 1545-1573, and 1575-1603: 25, 29, and 29 nt in length, respectively)

Table 2 Results of genotyping of nucleotide haplotypes obtained in each patient, extracted by next-generation sequencing based on ultra-deep pyrosequencing analysis %

Patient	A	B	C	D	E	F	H
1	0	0	0	0	100	0	0
2	0	0	100	0	0	0	0
3	7.1	0	0	92.9	0	0	0
4	0	0	0	100	0	0	0
5	0	0	0	100	0	0	0
6	1.7	0	0	98.3	0	0	0
7	0	0	0.3	51.3	0	48.4	0
8	7.1	0	0	92.9	0	0	0
9	0	0	0.3	7	0	92.7	0
10	0.9	0	0	50.9	0	0	48.2
11	0	0	0.3	7	0	92.7	0
12	95.1	0	4.4	0.5	0	0	0
13	0	0	100	0	0	0	0
14	46.6	0	8.2	33.2	0	12	0
15	89.7	0	4.4	8	0	1.5	0
16	0	0	0	100	0	0	0
17	0	0	0	100	0	0	0
18	0	0	0	0	100	0	0
19	0	0	95.3	3.6	0.9	0	0
20	100	0	0	0	0	0	0
21	0	99.6	0.4	0	0	0	0
22	0	0	100	0	0	0	0
23	0	0	87.8	12.2	0	0	0
24	0	0	100	0	0	0	0
25	97.9	0	0	2.1	0	0	0
26	0	0	100	0	0	0	0
27	0	0	0	100	0	0	0

%A to %H indicates the percentage of nucleotide haplotypes from each patient, classified as HBV genotype A, B, C, D, E, F or H. HBV: Hepatitis B virus.

spanning a region between nt 1519 and 1603 (85 nt). Five of these 85 nt positions (5.9%) showed an IC below 1.8 bits: positions 1527, 1557, 1589, and 1602 between 1.6 and 1.8 bits, and position 1524 between 1.4 and 1.6 bits (Figure 2B).

Conservation of the HBx amino acid sequence

To further confirm the nt conservation found, we also analyzed aa conservation in the same 2 scenarios considered for nt variants ($n = 27$ sequences for the consensus and $n = 330$ for the quasispecies approach). As was seen with the nt sequences, there were no difference when the 2 analyses (quasispecies vs consensus) were superimposed (Figure 4), which highlighted a single highly conserved region. Again, the results reported refer to the analysis using the quasispecies approach.

One highly conserved region was identified between aa 63 and 76 (13 aa), which included a portion of a Kunitz-like domain (Figure 5). All aa showed conservation near 4 bits (100% maximum conservation). This region in the HBx protein corresponded to the hyper-conserved nt sequence between positions 1563 and 1602. The first hyper-conserved nt region observed (nt 1255-1286) was not taken into account in this analysis, as it corresponded to a non-coding region and therefore, was not translated into aa.

DISCUSSION

Although classic nucleotide analogue-based therapies can effectively control HBV infection, eradication of the virus is not achieved because of persistence of the viral minichromosome, cccDNA. Furthermore, even though HBV replication can be inhibited by drug treatment, production of viral antigens may be maintained, and this could lead to progression of the disease^[29]. To overcome this challenge, new therapeutic approaches are needed, and gene therapy has emerged as an interesting option.

Ramanan *et al.*^[30] proposed a gene therapy based on CRISPR/Cas9 to specifically target a conserved region in HBV cccDNA. These authors reported an anti-HBV effect both *in vitro* and *in vivo*, together with inhibition of *de novo* infection in HepG2-hNTCP cells. However, in HBV infection, the viral genome may be inserted in the host genome. Hence, it is possible that a molecular scissors strategy, such as the CRISPR/Cas9 approach, might imply a risk of affecting the host genome in the regions of viral genome insertion.

With the siRNA approach, viral replication could be hampered and disease progression limited by direct interference with the viral messengers. As has been seen in both cell and mouse models^[12,31-33], this interfering RNA regulates the expression of specific viral genes by promoting cleavage of targeted mRNAs, thus

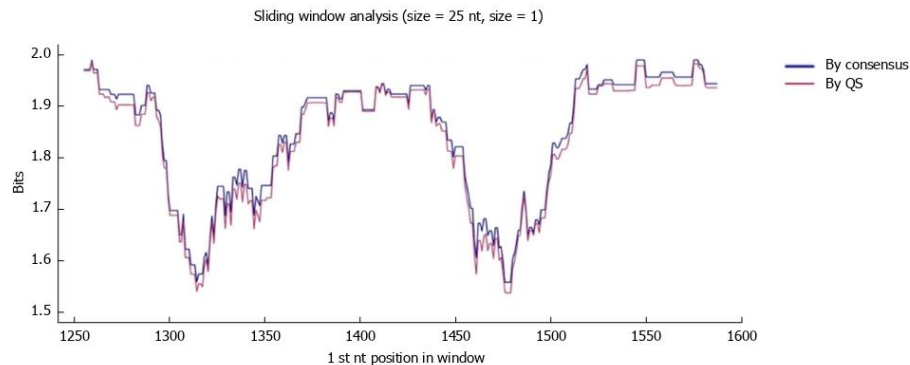


Figure 2 Sliding window analysis of the nucleotide region of interest in the hepatitis B virus X gene (nt 1255-1611). Each point on the graph is the result of the mean information content (IC, bits) of the windows 25-nt in size, with displacement between them in 1-nt steps. The purple line represents the mean IC from the multiple alignments of all haplotypes (in abundance > 0.25%) in the quasispecies (QS) of all patients ($n = 720$), whereas the blue line shows the mean IC obtained from the multiple alignments of the consensus obtained for each patient ($n = 27$).

inhibiting HBV replication. Specifically, siRNA promotes target mRNA cleavage in a sequence-specific manner through the RNA-induced silencing complex (RISC)^[34].

Definition of an extremely conserved region in an optimal HBV genomic region, such as the *HBX* gene, could be very useful for siRNA-based gene therapy strategies, and some authors have investigated this concept. In a recent study using predictive software, Thongthae *et al.*^[33] estimated potential siRNA target sites in the *HBX* gene (positions: 1317-1337, 1357-1377, and 1644-1664) from an HBV genotype A sequence. These were later tested *in vitro*, and a reduction in HBV expression was observed. In another effort, the Arbutus Biopharma Corporation recently published a phase-two study in this line. An siRNA was used as treatment for patients with chronic HBV infection, and the preliminary data indicated that the therapy was well tolerated and led to a significant reduction in HBsAg levels^[35,36].

HBX is located near the co-terminal 3' end of all the HBV mRNAs, which implies that interference at this level could abrogate the production of all the viral antigens. In addition, the *HBX* gene encodes a protein, HBx, which plays a key role in the HBV viral cycle. However, previous data reported by our group^[37] and supported in other studies^[17,38-40] have described considerable variability in the HBx transactivating C-terminal domain (encoded by the 3' end of the gene), with multiple insertions and deletions. Because of this variability, this region would not be considered an appropriate gene therapy target.

In light of the importance of the HBx protein for viral replication, it would be reasonable to posit that the gene encoding this protein would have a conserved region. On that basis and after excluding the 3' end region, we focused our study on the 5' end region of *HBX* and its upstream non-coding region (nt 1255-1611). For a gene therapy to be effective in a broad range of conditions,

the target sequence should remain conserved in a wide spectrum of clinical and virological situations. Hence, we analyzed samples from a heterogeneous group of 27 HBV-infected patients (in different clinical stages of HBV infection and with different viral genotypes) to seek a conserved target sequence over this spectrum. Two hyper-conserved regions were found. The first was located between nucleotides 1255 and 1286 in the non-coding region. Of note, *HBX* transcripts initiate at several different sites (between nt 1250-1350)^[41], which means that this conserved region might be not present in all of them, but would likely be present in the other viral transcripts. The second hyper-conserved region was located between nucleotides 1519 and 1603, within the coding region.

Conserved regions in this portion of the HBV gene have been reported previously. Karinova *et al.*^[42] observed two conserved regions in the S and X ORF of the HBV genotype A genome. These authors found that a CRISPR/Cas9 molecular scissor directed to this conserved region in *HBX* was able to modify both episomal cccDNA and chromosomally-integrated HBV DNA in reporter cell lines, thereby interfering with HBV replication and with *de novo* infection of hepatoma cell lines. In addition, with the use of predictive software, Thongthae *et al.*^[33] estimated some potential siRNA targets in the *HBX* gene (including the non-coding region identified here) in a single viral sequence, and reported the efficacy of this approach in an *in vitro* study. The value of the present study is that conservation of the regions examined was directly substantiated by sequencing analysis of patient samples, taking into account different HBV genotypes and different clinical stages of the infection. Furthermore, the nucleotide conservation documented here was supported by detection of a conserved region in the HBx protein sequence between aa 63 and 76, which is encoded by nt 1563-1602 (within the second hyper-conserved region).

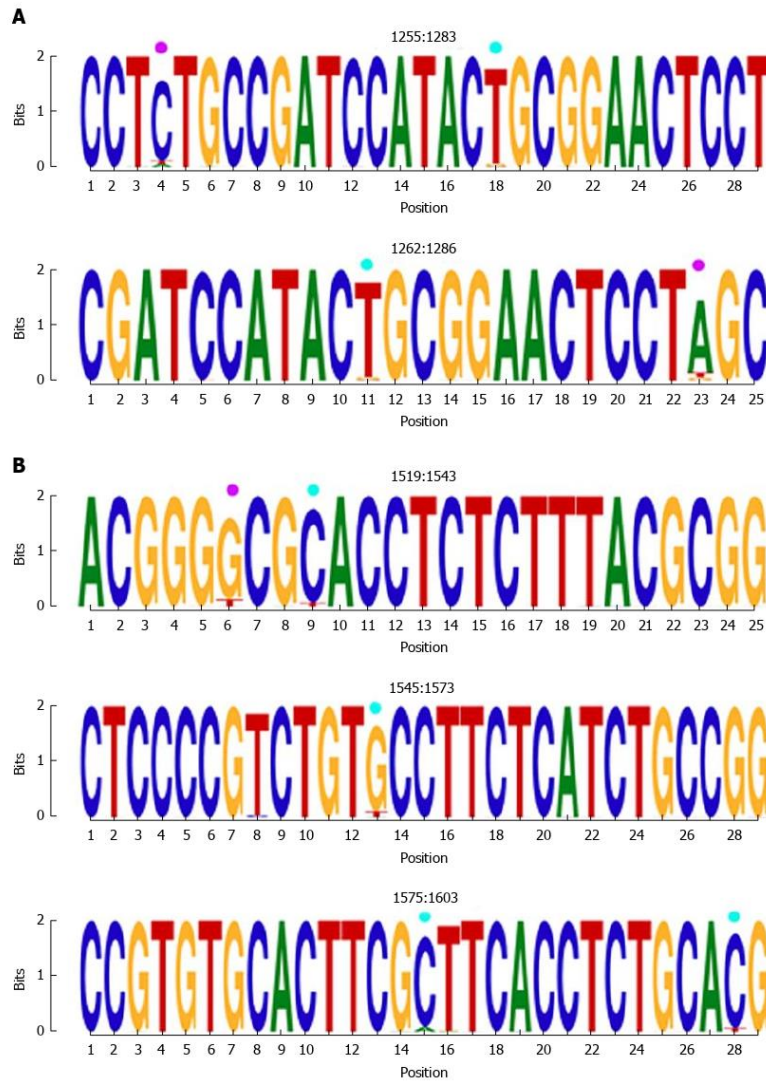


Figure 3 Representation by sequence logos of the information content of the most conserved regions detected in the multiple alignment of all nucleotide haplotypes obtained (quasispecies), nts 1255-1286 and nts 1519-1603. The relative sizes of the letters in each stack of nt sequence logos indicate their relative frequencies at each position within the multiple alignments of nt haplotypes. The total height of each stack of letters depicts the IC of each nt position, measured in bits (Y-axis): 0 bits being the minimum and 2, the maximum conservation. Nucleotide positions with an IC between 1.6 and 1.8 (80%-90% of maximum conservation) are indicated by light blue circles and those with an IC between 1.4 and 1.6 bits (70%-80% of maximum conservation) by pink circles. IC: Information content.

Of note, this fragment includes some aa from one of the HBx Kunitz-like domains (aa 58-70)^[43], which are able to inhibit the function of cellular degrading enzymes, such as proteases^[44]. This suggests that this portion of the HBx protein may be conserved to preserve the integrity of the protein, protecting it from undesired degradation.

As a limitation of the study, we should mention the relatively small sample size. From the initial group of 46 well-characterized treatment-naïve CHB patients available, only those with viremia levels high enough to amplify the HBV genome region of interest by our PCR technique could be included. Furthermore, we wished to

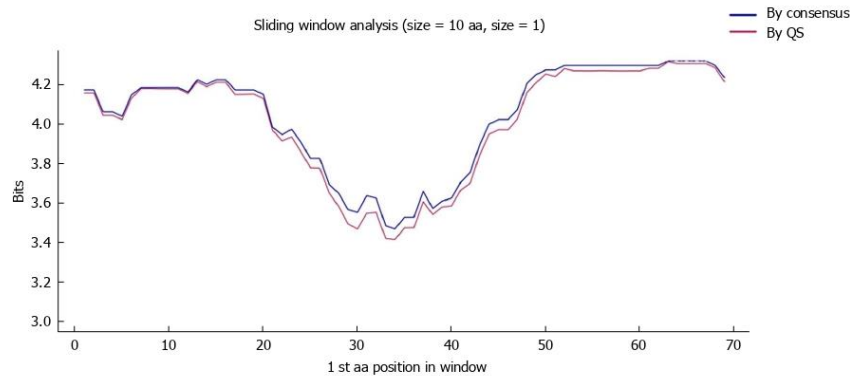


Figure 4 Sliding window of the coded core and X protein amino acid sequence (aa 1-79). Each point on the graph is the result of the mean information content (IC, bits) between windows of 10-aa size with displacement between them in 1-aa steps. The purple line represents the mean IC from the multiple alignments of all haplotypes (in abundance > 0.25%) in the quasispecies (QS) of all patients ($n = 330$), whereas the blue line shows the mean IC obtained from the multiple alignments of the consensus obtained for each patient ($n = 27$).

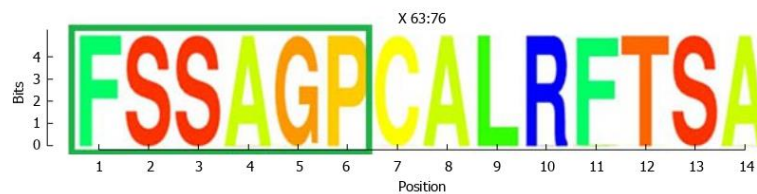


Figure 5 Representation by sequence logos of the information content of the conserved region in the core and X protein amino acid sequence (aa 63-76). The relative sizes of the letters in each stack in the sequence logos indicate their relative frequencies at each position within the multiple alignments of aa haplotypes obtained. The total height of each stack of letters depicts the IC of each aa position, measured in bits (Y-axis): 0 bits being the minimum and 4.32 bits the maximum conservation. Amino acids belonging to the Kunitz-like domain portion are framed in green. IC: Information content.

have a representation of various clinical stages of HBV infection and most HBV genotypes (A to F and H), which yielded a sample of 27 patients. Larger samples should be analyzed in future studies to confirm conservation of the regions investigated. We also have to point out that the NGS technology used in this study (GS-Junior platform, 454/Roche) has been discontinued by the supplier; nonetheless, the protocol describe here can be adapted to currently available platforms, such as the Illumina MiSeq (San Diego, United States). Finally, *in vitro* functional studies should be performed to test the potential usefulness of the 2 hyper-conserved domains described here as targets for siRNA-based antiviral gene therapy.

In summary, this study, performed in serum samples from HBV patients infected by different viral genotypes and in different clinical stages, identified regions in the *HBX* gene with high levels of conservation in all these circumstances. We found 2 hyper-conserved regions, the first in the non-coding region of *HBX* transcripts, and the second in the *HBX* coding region, which was conserved at both the nt and aa level. These hyper-conserved regions could be candidates for targeted gene therapies such as

the siRNA approach. Of particular interest, because of the co-terminal localization of the *HBX* gene, a siRNA system designed to target these regions could interfere with expression of all the HBV viral transcripts.

ARTICLE HIGHLIGHTS

Research background

Hepatitis B virus (HBV) infection can be controlled with current treatments, but cure is not achieved due to persistence of covalently closed circular DNA (cccDNA) in the nuclei of infected hepatocytes. This minichromosome forms a viral reservoir that is a source of residual viral replication and expression of viral proteins; thus, it has a key role in liver disease progression. To surmount this circumstance, new anti-HBV therapeutic approaches are under development, with gene therapy being a promising option. Among these approaches, small interference RNA (siRNA) can be used to silence specific genes at the post-transcriptional level through a sequence-specific interaction with target mRNAs, resulting in inhibition of viral protein expression. Among all the HBV proteins, Hepatitis B X protein (HBx), coded by the HBV X gene (*HBX*), is a determining factor in the infection. It regulates cccDNA expression and interacts with several cellular pathways, facilitating liver disease progression. Of particular note, because of its location near the co-terminal 3' end, all HBV transcripts include the *HBX* sequence. Hence, it could be a valuable target for a hypothetical curative treatment based on gene therapy. In this sense, identification of hyper-conserved regions within *HBX* is needed to define a new gene therapy system that would be effective whatever the patient's clinical stage or HBV genotype.

Research motivation

Although antiviral therapy can suppress viral replication, the risk of liver disease progression and development of hepatocellular carcinoma (HCC) remains due to cccDNA-related expression of viral antigens. Interference with expression of the viral proteins could be helpful to limit progression of the disease, and siRNAs would be valid tools in this sense. To design an effective siRNA, an appropriate target must be found. The *HBX* sequence is included in all the viral transcripts due to its co-terminal localization in the viral genome. siRNAs targeting hyper-conserved regions of this gene would interfere with expression of all the viral proteins. Furthermore, as these regions are conserved in the spectrum of clinical disease phases and viral genotypes, it would be a valid therapeutic approach for a wide range of situations. This could profoundly limit the risk of HCC, particularly in patients with low viremia due to antiretroviral efficacy.

Research objectives

Considering the essential role of HBx in viral infection and its potential utility as target for gene therapy, the aim of this study was to identify hyper-conserved regions within the HBV genome encompassing the *HBX* 5' coding region and the upstream non-coding region (included in all HBV transcripts) in samples from HBV-infected patients in various clinical stages and with different viral genotypes. The regions identified might be feasible targets for a gene therapy able to inhibit viral protein expression in a wide spectrum of clinical and virological circumstances, thus limiting liver disease progression and the risk of HCC.

Research methods

The study included 27 treatment-naïve chronic hepatitis B mono-infected patients in different clinical stages and with several HBV genotypes (from A-F and H). A serum sample from each patient with viremia > 3.5 log IU/mL was analyzed. The *HBX* 5' end region [nucleotide (nt) 1255-1611] was PCR-amplified and later analyzed using next-generation sequencing (NGS). The sequences (reads) obtained after sequencing underwent an in-house bioinformatics filtering procedure, and haplotypes with a relative frequency $\geq 0.25\%$ were maintained in the analysis. Haplotypes were genotyped by discriminant analysis with the same regions extracted from the 102 full-length patterns. Conservation of the quasispecies sequences was determined by calculating the information content (IC), based on Shannon's uncertainty, of each position in a multiple alignment of all different sequences found in the patients. Sliding window analysis was then carried out to locate the fragment of at least 25 nt or 10 aa (which corresponds to the length of a possible target for siRNA therapy) with the highest IC within the multiple alignments, moving forward in steps of 1 (nt or aa). This method enables detection of conserved regions within the 5' *HBX* gene by directly analyzing the viral quasispecies obtained with NGS.

Research results

After applying the quality filter, 1333069 haplotype sequences were obtained. Genotyping analysis highlighted a complex mixture of HBV genotypes. By studying the nt conservation, we identified two hyper-conserved nucleotide regions in *HBX*. The first one, between nt 1255 and 1286, corresponded to a non-coding region, whereas the second one, consisting of 3 conserved fragments (spanning an overall portion between 1519 and 1603), coincided with a coding region. Of note, the fragment between nt 1563 and 1602 was also conserved at the amino acid level, identifying a region between residues 63 and 76, which included a portion of a Kunitz-like domain. These results highlight new potential targets for gene therapy, mainly based on siRNA. Of note, *in vitro* and *in vivo* functional studies of the specific siRNAs should be performed to test their potential usefulness for therapy.

Research conclusions

Gene therapy represents a highly promising therapeutic tool to achieve a cure against HBV infection. Several sequence-specific treatment systems are currently in development, and identification of conserved sequences would provide useful therapeutic targets. Detection of a target present in all the clinical disease stages and HBV genotypes could lead to development of a therapy that would be effective in a wide range of situations. Considering the key role of HBx in viral infection and disease progression, we focused the study

on analyzing conservation of the *HBX* gene. Of note, considering the high variability previously observed in the 3' end of *HBX*, we speculated that the 5' end could be a better subject for study. Moreover, thanks to the co-terminality of this viral gene, a siRNA targeting this gene could interfere with all the viral transcripts. Here, we investigated conservation of a portion of the HBV genome encompassing the *HBX* 5' coding region and upstream non-coding region, both of which are included in all HBV transcripts. By NGS analysis, we identified two hyper-conserved regions in our region of interest in serum samples from HBV patients with different clinical and virological characteristics. This new therapeutic tool could have relevant applicability in clinical practice. Together with inhibition of the expression of one of the main viral proteins involved in HBV replication and disease progression, it could block the expression of the other viral antigens, thus profoundly interfering with disease evolution and the appearance of HCC. Furthermore, the NGS method developed here could be used to find other hyper-conserved regions within the HBV genome that could be potential targets for gene therapy based on siRNA.

Research perspectives

This study describes a method that can be used to find other conserved sequences in the HBV genome, making it a starting point in the search for other possible targets for gene therapy. Here, the hyper-conserved regions were found by directly analyzing the viral quasispecies sequences obtained using NGS. These regions can then be used to produce siRNA molecules for *in vitro* and *in vivo* testing of antiviral activity.

ACKNOWLEDGMENTS

The statistical and bioinformatics methods used in this study were reviewed by Dr. Josep Gregori from the liver disease-viral hepatitis laboratory (Vall d'Hebron Institut Recerca-Hospital Universitari Vall d'Hebron), CIBERehd and Roche Diagnostics SL. The authors thank Celine Cavallo for English language support and helpful editing suggestions.

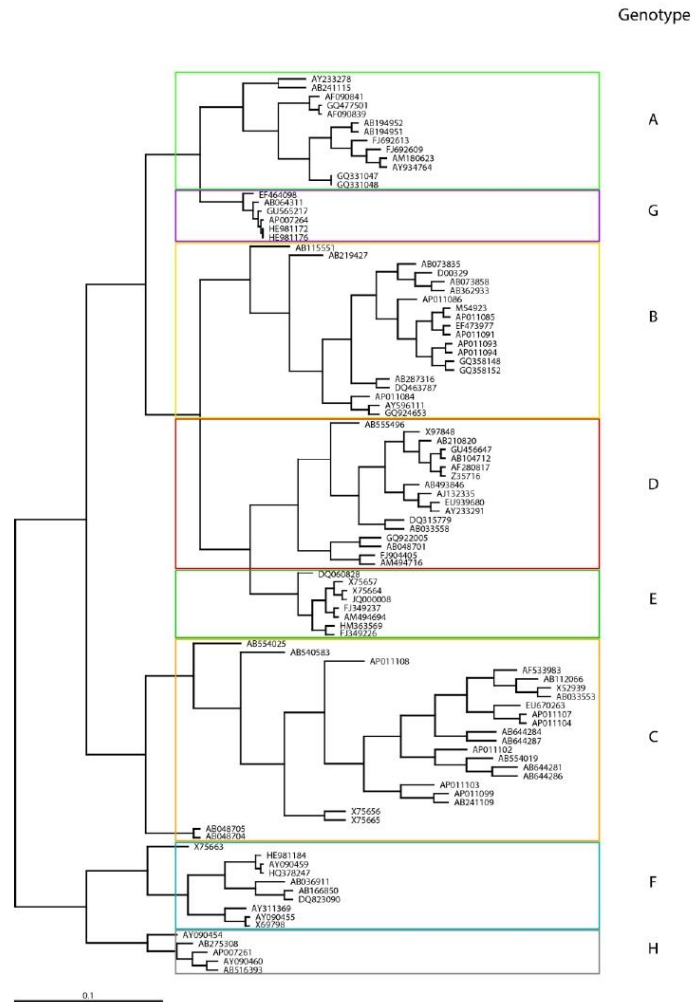
REFERENCES

1. Lin YY, Liu C, Chien WH, Wu LL, Tao Y, Wu D, Lu X, Hsieh CH, Chen PJ, Wang HY, Kao JH, Chen DS. New insights into the evolutionary rate of hepatitis B virus at different biological scales. *J Virol* 2015; **89**: 3512-3522 [PMID: 25589664 DOI: 10.1128/JVI.03131-14]
2. Kay A, Zoulim F. Hepatitis B virus genetic variability and evolution. *Virus Res* 2007; **127**: 164-176 [PMID: 17383765 DOI: 10.1016/j.virusres.2007.02.021]
3. Locarnini S, Zoulim F. Molecular genetics of HBV infection. *Antivir Ther* 2010; **15** Suppl 3: 3-14 [PMID: 21041899 DOI: 10.3851/IMP1619]
4. Sunbul M. Hepatitis B virus genotypes: global distribution and clinical importance. *World J Gastroenterol* 2014; **20**: 5427-5434 [PMID: 24833873 DOI: 10.3748/wjg.v20.i18.5427]
5. Pujol FH, Navas MC, Hainaut P, Chemin I. Worldwide genetic diversity of HBV genotypes and risk of hepatocellular carcinoma. *Cancer Lett* 2009; **286**: 80-88 [PMID: 19683385 DOI: 10.1016/j.canlet.2009.07.013]
6. European Association for the Study of the Liver. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol* 2017; **67**: 370-398 [PMID: 28427875 DOI: 10.1016/j.jhep.2017.03.021]
7. Belloni L, Pollicino T, De Nicola F, Guerrieri F, Raffa G, Fanciulli M, Raimondo G, Leviero M. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc Natl Acad Sci USA* 2009; **106**: 19975-19979 [PMID: 19906987 DOI: 10.1073/pnas.0908365106]
8. Yang HC, Kao JH. Persistence of hepatitis B virus covalently

- closed circular DNA in hepatocytes: molecular mechanisms and clinical significance. *Emerg Microbes Infect* 2014; **3**: e64 [PMID: 26038757 DOI: 10.1038/emi.2014.64]
- 9 **Bloom K**, Ely A, Arbutnot P. Recent advances in use of gene therapy to treat hepatitis B virus infection. *Adv Exp Med Biol* 2015; **848**: 31-49 [PMID: 25757614 DOI: 10.1007/978-1-4939-2432-5_2]
 - 10 **Dong C**, Qu L, Wang H, Wei L, Dong Y, Xiong S. Targeting hepatitis B virus cccDNA by CRISPR/Cas9 nuclease efficiently inhibits viral replication. *Antiviral Res* 2015; **118**: 110-117 [PMID: 25843425 DOI: 10.1016/j.antiviral.2015.03.015]
 - 11 **Weber ND**, Stone D, Sedlak RH, De Silva Feelixge HS, Roychoudhury P, Schiffer JT, Aubert M, Jerome KR. AAV-mediated delivery of zinc finger nucleases targeting hepatitis B virus inhibits active replication. *PLoS One* 2014; **9**: e97579 [PMID: 24827459 DOI: 10.1371/journal.pone.0097579]
 - 12 **Gish RG**, Yuen MF, Chan HL, Given BD, Lai CL, Locarnini SA, Lau JY, Wooddell CI, Schlupp T, Lewis DL. Synthetic RNAi triggers and their use in chronic hepatitis B therapies with curative intent. *Antiviral Res* 2015; **121**: 97-108 [PMID: 26129970 DOI: 10.1016/j.antiviral.2015.06.019]
 - 13 **Poch O**, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* 1989; **8**: 3867-3874 [PMID: 2555175 DOI: 10.1093/emboj/16.6.1248]
 - 14 **Tu H**, Bonura C, Giannini C, Mouly H, Soussan P, Kew M, Paterlini-Bréchet P, Bréchet C, Kremsdorf D. Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Res* 2001; **61**: 7803-7810 [PMID: 11691796]
 - 15 **Wang Q**, Zhang WY, Ye LH, Zhang XD. A mutant of HBx (HBxDelta127) promotes hepatoma cell growth via sterol regulatory element binding protein 1c involving 5-lipoxygenase. *Acta Pharmacol Sin* 2010; **31**: 367-374 [PMID: 20173757 DOI: 10.1038/aps.2010.5]
 - 16 **Panjaworayan N**, Roessner SK, Firth AE, Brown CM. HBVRegDB: annotation, comparison, detection and visualization of regulatory elements in hepatitis B virus sequences. *Virology* 2007; **4**: 136 [PMID: 18086305 DOI: 10.1186/1743-422X-4-136]
 - 17 **Kim H**, Lee SA, Kim BJ. X region mutations of hepatitis B virus related to clinical severity. *World J Gastroenterol* 2016; **22**: 5467-5478 [PMID: 27350725 DOI: 10.3748/wjg.v22.i24.5467]
 - 18 **Ali A**, Abdel-Hafiz H, Suhal M, Al-Mars A, Zakaria MK, Fatima K, Ahmad S, Azhar E, Chaudhary A, Qadri I. Hepatitis B virus, HBx mutants and their role in hepatocellular carcinoma. *World J Gastroenterol* 2014; **20**: 10238-10248 [PMID: 25132741 DOI: 10.3748/wjg.v20.i30.10238]
 - 19 **Lee SA**, Mun HS, Kim H, Lee HK, Kim BJ, Hwang ES, Kook YH, Kim BJ. Naturally occurring hepatitis B virus X deletions and insertions among Korean chronic patients. *J Med Virol* 2011; **83**: 65-70 [PMID: 21108340 DOI: 10.1002/jmv.21938]
 - 20 **Peng Y**, Liu B, Hou J, Sun J, Hao R, Xiang K, Yan L, Zhang J, Zhuang H, Li T. Naturally occurring deletions/insertions in HBV core promoter tend to decrease in hepatitis B e antigen-positive chronic hepatitis B patients during antiviral therapy. *Antivir Ther* 2015; **20**: 623-632 [PMID: 25838313 DOI: 10.3851/IMP2955]
 - 21 **Homs M**, Caballero A, Gregori J, Tabernero D, Quer J, Nieto L, Esteban R, Buti M, Rodriguez-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 [PMID: 25393280 DOI: 10.1371/journal.pone.0112306]
 - 22 **Ramirez C**, Gregori J, Buti M, Tabernero D, Camós S, Casillas R, Quer J, Esteban R, Homs M, Rodriguez-Frias F. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res* 2013; **98**: 273-283 [PMID: 23523552 DOI: 10.1016/j.antiviral.2013.03.007]
 - 23 **Team RC**. A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. 2016; Available from: URL: <https://www.r-project.org/>
 - 24 **Cuadras C**. A distance approach to discriminant analysis and its properties. In: *Mathematics preprint series*. Barcelona: 1991.
 - 25 **Cuadras C**. Distance analysis in discrimination and classification using both continuous and categorical variables. In: *Statistical AData analysis and Interference*. Amsterdam, 1989; 459-473
 - 26 **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 [PMID: 7463489]
 - 27 **Schneider TD**. Information content of individual genetic sequences. *J Theor Biol* 1997; **189**: 427-441 [PMID: 9446751 DOI: 10.1006/jtbi.1997.0540]
 - 28 **Griffiths AJF**, Miller JH, Suzuki DT *et al*. *An Introduction to Genetic Analysis*. 7th edition. New York, WH. Freeman: 2000
 - 29 **Kwon H**, Lok AS. Hepatitis B therapy. *Nat Rev Gastroenterol Hepatol* 2011; **8**: 275-284 [PMID: 21423260 DOI: 10.1038/ngastro.2011.33]
 - 30 **Ramanan V**, Shlomai A, Cox DB, Schwartz RE, Michailidis E, Bhatta A, Scott DA, Zhang F, Rice CM, Bhatia SN. CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. *Sci Rep* 2015; **5**: 10833 [PMID: 26035283 DOI: 10.1038/srep10833]
 - 31 **Klein C**, Bock CT, Wedemeyer H, Wüstefeld T, Locarnini S, Dienes HP, Kubicka S, Manns MP, Trautwein C. Inhibition of hepatitis B virus replication in vivo by nucleoside analogues and siRNA. *Gastroenterology* 2003; **125**: 9-18 [PMID: 12851866 DOI: 10.1016/S0016-5085(03)00720-0]
 - 32 **Shlomai A**, Shaul Y. Inhibition of hepatitis B virus expression and replication by RNA interference. *Hepatology* 2003; **37**: 764-770 [PMID: 12668968 DOI: 10.1053/jhep.2003.50146]
 - 33 **Thongthae N**, Payungporn S, Poovorawan Y, T-Thienprasert NP. A rational study for identification of highly effective siRNAs against hepatitis B virus. *Exp Mol Pathol* 2014; **97**: 120-127 [PMID: 24953337 DOI: 10.1016/j.yexmp.2014.06.006]
 - 34 **Tan FL**, Yin JQ. RNAi, a new therapeutic strategy against viral infection. *Cell Res* 2004; **14**: 460-466 [PMID: 15625012 DOI: 10.1038/sj.cr.7290248]
 - 35 **Agarwal K**, Gane E, Cheng W, Sievert W, Roberts S, Ahn SH, Kim YJ, Streinu-cercel A, Denning J, Symonds W, Mendez P. HBcrAg, HBV-RNA Declines in A Phase 2a Study Evaluating the Multi-Dose Activity of ARB-1467 in HBeAg-Positive and Negative Virally Suppressed Patients With Hepatitis B. *2017*; 1-16
 - 36 **Eley T**, Russ R, Streinu-cercel A, Gane EJ, Roberts SK, Ahn SH, Kim YJ, Symonds W, Mendez P. Pharmacokinetics and exploratory exposure-response of siRNAs administered monthly as ARB-001467 (ARB-1467) in a Phase 2a study in HBeAg positive and negative virally suppressed subjects with chronic hepatitis B. *2017*
 - 37 **Caballero A**, Gregori J, Buti M, Tabernero D, Quer J, Blasi M, Rodriguez-Algara F, Casillas R, González C, Belmonte I, Nieto L, Costa X, Esteban R, Homs M, Rodriguez-Frias F. Insertions and/or deletions in the main regulatory region of hepatitis B virus suggest multicoding of the X protein. *J Hepatol* 2015; **62**: S523 [DOI: 10.1016/S0168-8278(15)30762-5]
 - 38 **Zhang ZH**, Wu CC, Chen XW, Li X, Li J, Lu MJ. Genetic variation of hepatitis B virus and its significance for pathogenesis. *World J Gastroenterol* 2016; **22**: 126-144 [PMID: 26755865 DOI: 10.3748/wjg.v22.i1.126]
 - 39 **Lazarevic I**. Clinical implications of hepatitis B virus mutations: recent advances. *World J Gastroenterol* 2014; **20**: 7653-7664 [PMID: 24976703 DOI: 10.3748/wjg.v20.i24.7653]
 - 40 **Salarnia F**, Besharat S, Zhand S, Javid N, Khodabakhshi B, Moradi A. Mutations in Hepatitis-B X-Gene Region: Chronic Hepatitis-B versus Cirrhosis. *J Clin Diagn Res* 2017; **11**: OC31-OC34 [PMID: 28511432 DOI: 10.7860/JCDR/2017/22570.9498]
 - 41 **Treinin M**, Laub O. Identification of a promoter element located upstream from the hepatitis B virus X gene. *Mol Cell Biol* 1987; **7**: 545-548 [PMID: 3494193]
 - 42 **Karimova M**, Beschornor N, Dammermann W, Chemnitz J, Indenbirken D, Bockmann JH, Grundhoff A, Lüth S, Buchholz F,

- Schulze zur Wiesch J, Hauber J. CRISPR/Cas9 nickase-mediated disruption of hepatitis B virus open reading frame S and X. *Sci Rep* 2015; **5**: 13734 [PMID: 26334116 DOI: 10.1038/srep13734]
- 43 **Datta S**, Banerjee A, Chandra PK, Biswas A, Panigrahi R, Mahapatra PK, Panda CK, Chakrabarti S, Bhattacharya SK, 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 [PMID: 18952249 DOI: 10.1016/j.virol.2008.09.007]
- 44 **Lee J**, Chung JH, Kim HM, Kim DW, Kim H. Designed nucleases for targeted genome editing. *Plant Biotechnol J* 2016; **14**: 448-462 [PMID: 26369767 DOI: 10.1111/pbi.12465]
- P-Reviewer:** Enomoto H **S-Editor:** Wang XJ **L-Editor:** A
E-Editor: Huang Y

6. Material suplementario



Supplementary Figure 1. Phylogenetic analysis of the region between nucleotides 1255 and 1611, extracted from 102 full-length HBV genome patterns, used for genotyping.

The representation of the phylogenetic analysis shows that the different patterns clustered according to their HBV genotype, which coincided with that previously determined using the full HBV genome sequence. Each color box highlights the sequences that correspond to a specific genotype (GenBank accession number reported on the right).

Supplementary table 1. Reference sequences used for genotyping

NCBI GenBank accession numbers of the 105 reference sequences used in HBV genotyping:

HBV genotype (N sequences)	GenBank accession numbers
A (13)	AY233278, AB241115, GQ477501, AF090841, AF090839, AB194952, AB194951, AM180623, AY934764, FJ692609, FJ692613, GQ331047, GQ331048
B (20)	AB073858, AB362933, D00329, AY596111, AP011084, GQ924653, M54923, AP011085, AB073835, AB115551, AB219427, AP011086, AB287316, DQ463787, EF473977, AP011091, AP011093, AP011094, GQ358148, GQ358152
C (23)	AB112066, X52939, AF533983, AB033553, X75656, X75665, AB048705, AB048704, AP011099, AB241109, AP011102, AP011103, EU670263, AP011107, AP011104, AP011108, AB540583, AB554019, AB554025, AB644281, AB644284, AB644286, AB644287
D (17)	AB555496, GU456647, AB104712, AF280817, AB210820, X97848, Z35716, EU939680, AY233291, AJ132335, GQ922005, AB048701, DQ315779, AB033558, AB493846, FJ904405, AM494716
E (8)	X75664, X75657, FJ349237, AM494694, HM363569, FJ349226, DQ060828, JQ000008
F (10)	AY090459, HQ378247, HE981184, AY090455, AY311369, X69798, X75663, AB036911, AB166850, DQ823090
G (6)	EF464098, HE981172, HE981176, AP007264, GU565217, AB064311
H (5)	AP007261, AB275308, AY090460, AY090454, AB516393

CAPÍTULO 4: SEGUNDO ESTUDIO

SEGUNDO ESTUDIO: HBeAg-ve chronically infected patients: sophisticated viral quasispecies with a genotype-related pattern of mutations in the hepatitis B X gene.

1. Introducción

Los pacientes con infección crónica HBeAg negativo (-)(IC), conocidos anteriormente como "portadores inactivos", constituyen un grupo clínico característico que se define por HBeAg (-), anti-HBeAg positivo, una viremia ≤ 2.000 UI/ml, un nivel bajo de HBsAg, ALT dentro de los valores normales (< 40 UI / L) y sin evidencia de daño hepático³⁰⁵. Debido a su baja viremia, el estudio de la quasispecies (QS) circulante de este grupo de pacientes es difícil y el mecanismo que conduce a esta tasa de replicación tan baja permanece desconocido. Además, como estos pacientes tienen un buen pronóstico, no se le suele prestar especial atención. Sin embargo, el riesgo de desarrollar HCC y / o reactivación de la hepatitis después de la terapia inmunosupresora aún persiste, por lo que es aconsejable un seguimiento continuo^{189,367,368}. La HBx, codificada por el *HBX*, tiene un papel clave en la infección por el VHB y en la progresión de la enfermedad ya que es esencial para iniciar una infección productiva^{105,106} y permite la expresión de genes virales, al actuar sobre el control epigenético del ADNccc^{105,369}.

Además, existen un grupo de pacientes que son difíciles de clasificar, que se denominan pacientes en "zona gris". Estos pacientes pueden tener niveles de ADN del VHB de entre 2.000 IU / ml y 20.000 UI / ml acompañados de una ALT dentro de los rangos de normalidad o ligeramente elevada y sin la presencia de daño hepático³⁷⁰. Debido a estas características resulta difícil clasificarlos como hepatitis o infecciones crónicas, lo cual es esencial para hacer un abordaje terapéutico acertado de estos pacientes.

2. Hipótesis y objetivo

2.1. Hipótesis

En el estudio anterior se identificaron 2 regiones híper-conservadas en la extremo 5' del *HBX*, independientemente de la situación clínica y virológica de los pacientes estudiados. Sin embargo, se pudieron incluir pocos pacientes con infección crónica HBeAg (-) por lo

que no se pudieron comparar con pacientes con hepatitis crónica (CHB). La conservación/variabilidad de la secuencia del extremo 5' del *HBX* puede ser un factor determinante para la etapa clínica (infección o hepatitis crónica) en la infección crónica por el VHB HBeAg (-). Asimismo, estas 2 etapas clínicas también podrían diferenciarse en la complejidad de la QS y en las variantes específicas en esta región. Por lo tanto, el estudio en profundidad de la conservación/variabilidad, complejidad y variantes específicas de la QS del extremo 5' del *HBX* en pacientes en diferentes etapas clínicas de la infección crónica por el VHB HBeAg (-), podría proporcionar una valiosa información para despejar los mecanismos que permiten controlar la replicación viral. Dicha información podría ser muy útil para el desarrollo de nuevos enfoques terapéuticos. Asimismo, la identificación de características exclusivas de los diferentes grupos clínicos podría ayudar la correcta clasificación de pacientes de la zona gris.

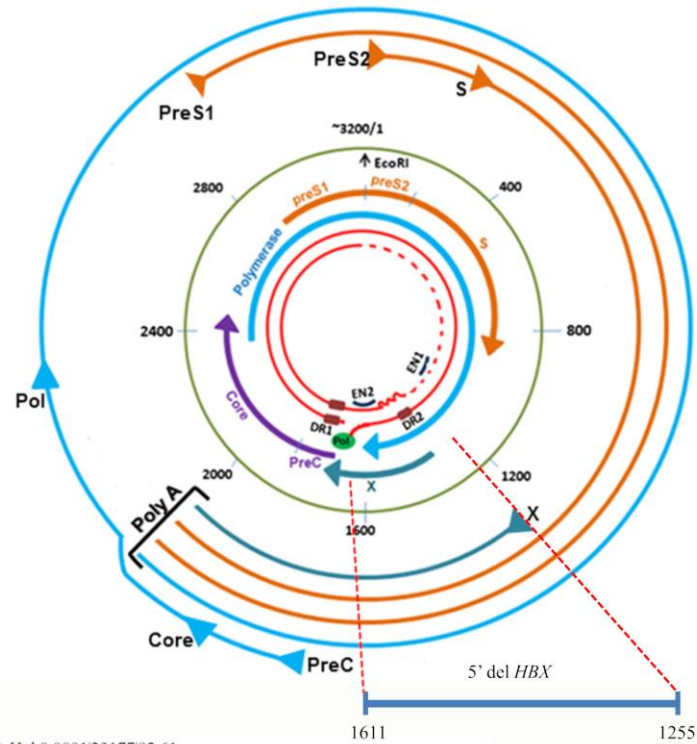
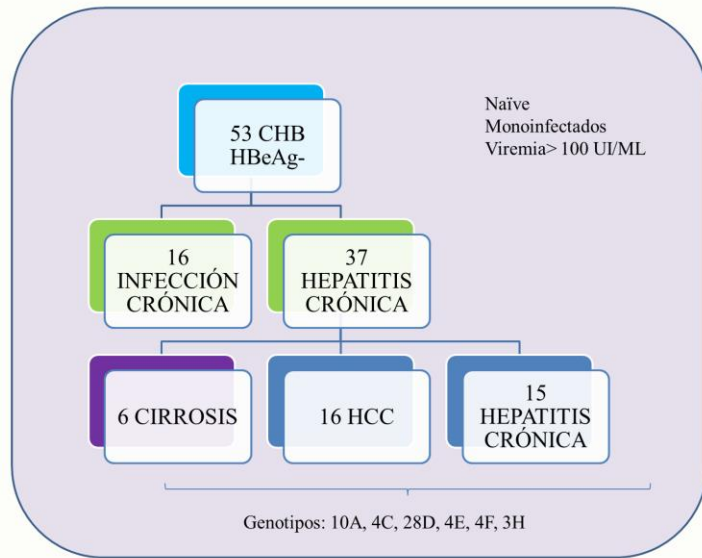
2.2. Objetivo

El objetivo de este estudio fue analizar por secuenciación masiva (NGS) la conservación/variabilidad, complejidad y variantes específicas de la QS de la región 5' de *HBX* en los pacientes IC. Estos parámetros fueron comparados con los de QS de pacientes con CHB (en presencia o no de lesiones hepáticas como la cirrosis y HCC) para identificar así los factores genéticos virales que podrían caracterizar a los pacientes con IC, con el fin de establecer la posible relación entre éstos y los mecanismos de control de la replicación viral.

3. Resumen de estudio

3.1. Diseño del estudio

En el presente estudio se ha analizado el mismo fragmento del *HBX* (nts 1255 a 1611) que, en la primera parte de este trabajo de tesis doctoral, en esta ocasión mediante la tecnología de NGS basada en la tecnología de Sequencing By Synthesis (SBS) de Illumina. En él se han incluido 53 pacientes con infección crónica por el VHB HBeAg (-), los cuales fueron previamente bien clasificados en infección o hepatitis crónicas y dentro de estos últimos en función de su progresión: en CHB, en cirróticos sin HCC o en HCC. El flujo de trabajo llevado a cabo se muestra en la siguiente figura.



Illumina Miseq platform*



```
>Hpl.0.0001|23177|92.61
CCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTT
TAATTCGTGCTCTCTCCCGAAATATACAGCATTCCATG
CGTCCCTTGTCTACGTCCTCCGTCAGCGCTGAATCCTGCGGACG
CTCCGTCTGCCGTTCCAGCCGACCACGGGGCGCACCTCTCTT
CCGTGTGCACCTTCGCTTACCTCTGCACGTCCGATGG
>Hpl.1.0001|829|3.31
CCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTT
TAATTCGTGCTCTCTCCCGAAATATACAGCATTCCATG
CGTCCCTTGTCTACGTCCTCCGTCAGCGCTGAATCCTGCGGACG
CTCCGTCTGCCGTTCCAGCCGTCACGGGGCGCACCTCTCTT
CCGTGTGCACCTTCGCTTACCTCTGCACGTCCGATGG
```

Figura 47. Flujo de trabajo del estudio. * Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA)

3.2. Análisis bioinformático de los datos de secuenciación

Las lecturas obtenidas después de la secuenciación se sometieron a un procedimiento de filtrado bioinformático desarrollado por nuestro propio grupo. El sistema de filtrado de las lecturas obtenidas se ha llevado a cabo igual que en el primer estudio en el entorno de programación de código abierto R³⁶¹, utilizando las librerías Bioconductor³⁶² y Biostrings³⁶³ para el alineamiento de las lecturas obtenidas con el patrón de referencia. Hay que destacar que la tecnología de NGS utilizada en este estudio permite obtener lecturas de un máximo de 300 nts. Por lo tanto, de un mismo amplicón se han obtenido lecturas en sentido *forward* y *reverse* que se solaparon en el centro de la secuencia del amplicón completo (de 458 pares de bases o pb) pero no en sus extremos. Este sistema de filtrado consistió en una serie de pasos (Figura 48):

- 1) Filtrado de calidad de las secuencias: Las lecturas que contenían indeterminaciones y las que no tenían la longitud suficiente, fueron eliminadas. Además, se evaluaron los parámetros generales del instrumento relacionados con la calidad.
- 2) Colapso de los haplotipos. Para ello utilizamos el entorno de programación FLASH³⁷¹, con lo que conseguimos imponer un mínimo de 20 pb superpuestas entre las lecturas obtenidos en ambos sentidos, con un máximo de 10% de desapareamientos (rendimiento 60% -80% para amplicones de 450 a 500 pb).
- 3) Eliminación de lecturas. Las lecturas fueron descartadas en el caso de que más del 5% de las bases estuvieran por debajo de 30 de la puntuación de Phred³⁷², lo que se corresponde con una precisión estimada de 99,9% (rendimiento 75% -85%).
- 4) Demultiplexado de las lecturas. La demultiplexación se realizó mediante la identificación de secuencias de oligonucleótidos (MID) en ambos extremos dentro de las ventanas en las posiciones esperadas de las lecturas (rendimiento 70% -85%). Primero, los MID individuales (10 secuencias de oligonucleótidos) fueron utilizados para distinguir entre muestras de diferentes pacientes, en los que solo se permitió un desapareamiento. En segundo lugar, se utilizaron cebadores específicos (oligos de 20 a 30 pb) para distinguir entre las diferentes regiones en el genoma, y entre las dos cadenas, aquí se permitieron hasta tres desapareamientos. Finalmente, fueron recortados los MID y los cebadores y se obtuvo un archivo fasta para cada combinación de MID, cebador y hebra, en los cuales las lecturas se colapsaron en haplotipos con sus frecuencias correspondientes.

5) Eliminación de las inserciones y deleciones respecto al haplotipo de referencia. En cada archivo fasta, los haplotipos fueron alineados con la secuencia de referencia o *master* (el haplotipo más abundante en el archivo) y se aplicó un filtro de calidad (rendimiento > 90%). Este filtro de calidad consistió en descartar haplotipos que no cubrieron el amplicón completo y aquellos que tuvieron más de dos indeterminaciones, tres “huecos” (*gaps*) o más de 99 diferencias con respecto a la de referencia. Finalmente, las indeterminaciones y *gaps* aceptados se reparan según la secuencia de referencia.

6) Intersección de haplotipos. Para esta etapa, fueron seleccionados los haplotipos que presentaron una abundancia igual o superior al 0,1% en ambas cadenas (rendimiento 50% -60%), mientras que los haplotipos únicos para una cadena se descartaron. La cobertura de los haplotipos que pasaron el filtro de calidad se tomó como la suma de lecturas en ambas cadenas.

7) Resultado final. Todos los haplotipos con abundancias no inferiores al 0,25% se mantuvieron. Los haplotipos finales fueron llamados haplotipos consenso, y estos fueron las bases para el análisis posterior en este estudio (rendimiento total final 15% -25%).

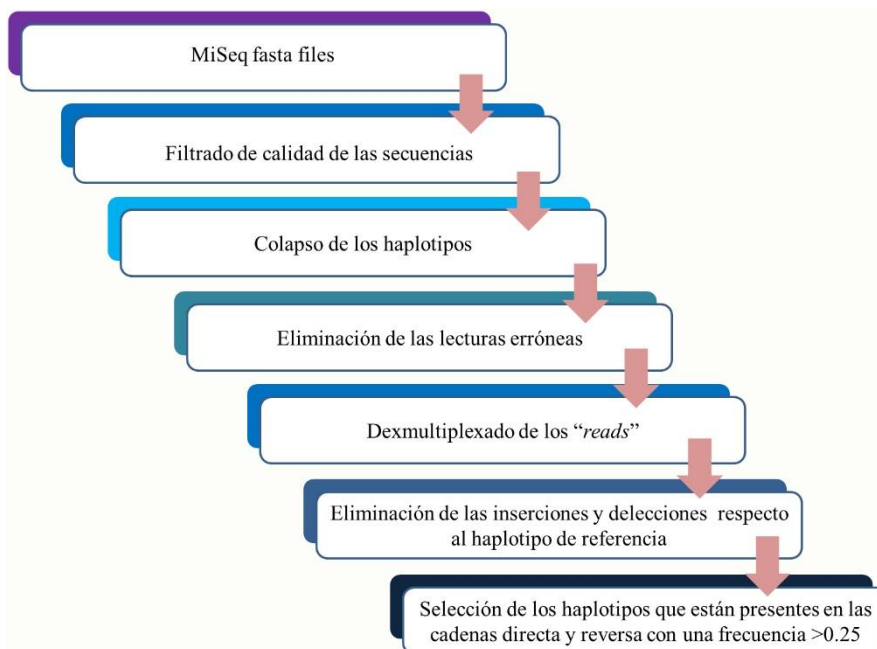


Figura 48 Flujo de trabajo del procesamiento de los datos de MiSeq para obtener el muestreo de las quasiespecies virales reales del VHB.

3.3. Análisis de conservación de las quasiespecies

El estudio de la conservación de los haplotipos (variantes, es decir poblaciones virales) que forman la QS se determinó calculando el contenido de información en cada posición en un alineamiento múltiple, de todas las secuencias de nt y aminoácidos (aa) diferentes, encontradas en los pacientes. Posteriormente se realizó un Sliding window (ver en el Capítulo 1 el apartado 2.1.3. *Estudio de la conservación de la quasiespecie*) para localizar los fragmentos de al menos 25 nt (longitud mínima de un siRNA ³⁶⁴) o 10 aa con el contenido de información más alto dentro de los alineamientos múltiples. Las secuencias de estos fragmentos fueron representadas como logos de secuencia (ver en el Capítulo 1 el apartado 2.1.3. *estudio de la conservación de la quasiespecie*), creados con el paquete de lenguaje R motiveStack.

3.4. Análisis de la diversidad y complejidad de la QS

Los índices más comúnmente utilizados para comparar la diversidad y la complejidad de las QS virales son la entropía de Shannon (H_{SN}), el índice de Gini-Simpson (H_{GS}), la frecuencia de mutación (Mf) y el índice de diversidad de nucleótidos (π)¹⁸⁹ (ver en el Capítulo 1 el apartado 2.1.2. *Índices para el estudio de la diversidad de la quasiespecie*). La entropía de Shannon normalizada y el índice de Gini-Simpson miden la diversidad de haplotipos teniendo en cuenta el número de haplotipos y su frecuencia relativa (índices de abundancia).

$$H_{SN} = - \sum_{i=1}^H p_i \log p_i / \log(H) *$$

* p_i = frecuencia del haplotipo i en la población de QS. H = número de Haplotipos en la población de QS

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

* p_i = frecuencia del haplotipo i en la población de QS.

Mf mide la diversidad genética con respecto al haplotipo más prevalente, mientras que π mide la diversidad genética de la población viral como el número promedio de mutaciones por sitio, entre cada par de haplotipos en la población viral. Estos índices,

definidos como índices funcionales, dan una idea de la complejidad de la QS en términos de mutaciones nucleotídicas (diferencias entre la secuencia de los haplotipos que forman la QS), pero no tienen en cuenta la frecuencia relativa de los haplotipos.

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

* d = Proporción de mutaciones relacionadas con la secuencia master

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

* p_i = frecuencia del haplotipo i en la QS, p_j = frecuencia del haplotipo j en la QS, d_{ij} = distancia de Hamming (número de mutaciones que diferencia el haplotipo i del j).

La diferencia intergrupar en la complejidad de la QS se evaluó comparando las medianas de cada índice entre los distintos grupos de pacientes mediante una prueba de Kruskal-Wallis, adaptada para una comparación múltiple mediante la aplicación de una prueba posthoc (Dunn test) utilizando el software R 3.0.2 ⁶³⁶¹. Un p-valor de <0,05 se consideró estadísticamente relevante (intervalo de confianza del 95%). Los p-valores se ajustaron con la corrección de Bonferroni y solo aquellos <0,05 se consideraron estadísticamente significativos.

3.5. Análisis de las mutaciones

Las secuencias de todos los haplotipos fueron traducidas a aa y recolapsadas de manera que solo quedaron las secuencias diferentes entre ellas. Estas secuencias se alinearon contra una secuencia de referencia de su mismo genotipo, para observar así posibles cambios de aa. Las secuencias de referencia de cada genotipo se generaron creando un consenso a partir del alineamiento de la región de interés (nt 1255 a 1611) extraída de secuencias de genoma del VHB obtenidas de GenBank (13 secuencias para genotipo A, 23 para genotipo C, 16 para genotipo D, 8 para genotipo E, 10 para el genotipo F, 6 para el genotipo G y 5 para el genotipo H) (Ver Table 2 del Capítulo 4 en el apartado 6. *Material suplementario*). La frecuencia de los cambios de aa se calculó como la suma de

las frecuencias relativas de cada haplotipo con dicho cambio en la QS. La frecuencia de las mutaciones observadas en el grupo de IC se comparó con su frecuencia en los otros grupos aplicando una prueba de Kruskal-Wallis y un test de Dunn como prueba posthoc. También se analizó la agrupación de estos cambios en patrones determinados, mediante una prueba de correlación de Spearman y un análisis de agrupamiento jerárquico (software R 3.0.2.6³⁶¹). En todas las pruebas estadísticas se aplicó una corrección de Bonferroni a los p-valores y solo aquellos $<0,05$ se consideraron estadísticamente significativos.

3.6. Análisis de la complejidad de la quasiespecies. Comparación entre pacientes IC y los otros grupos

Después de aplicar los filtros de calidad bioinformáticos descritos en el apartado 3.2. Análisis bioinformático de los datos de secuenciación de este mismo capítulo, se obtuvieron 1.156.140 secuencias, lo que produjo una mediana [IQR] de 17.507 [13.718-24.901] secuencias por paciente.

Al comparar la complejidad de la QS en términos de abundancia no observamos ninguna diferencia entre el grupo de IC y los otros grupos, observándose la misma tendencia para H_{SN} (mediana [IQR] de 22.660 [17.490-26.200], 19.885 [11.815-24.755], 29.315 [23.867-31.630], 17.020 [3.456-21.303] en IC, CHB, LC y HCC respectivamente, Figura 49, Panel A) y H_{GS} (0,8 [0,8-0,9] para CI, 0,7 [0,5-0,8] para CHB, 0,9 [0,8-0,9] para LC y 0,7 [0,4-0,8] para HCC, Figura 49, Panel B).

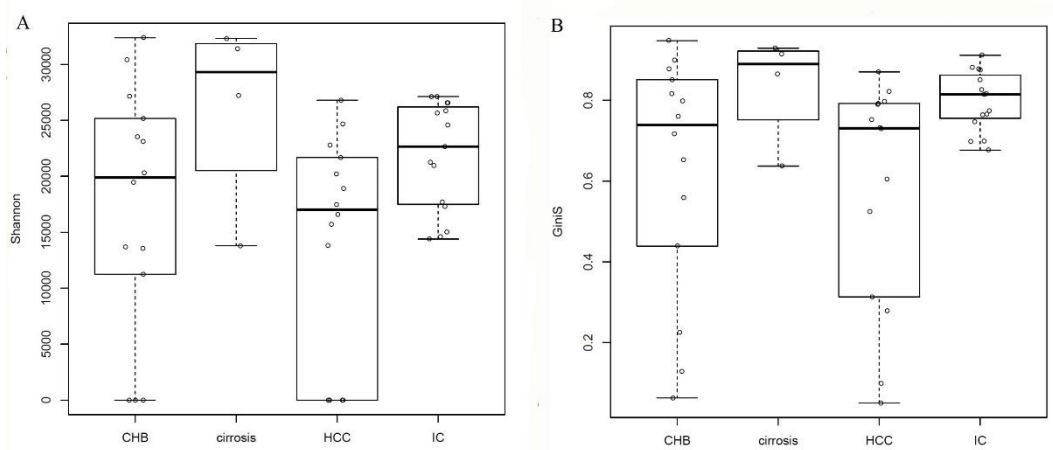


Figura 49. Comparación de los índices de abundancia [Shannon (H_{SN}) y Gini-Simpson (H_{GS})] entre los distintos grupos. Los diagramas de caja muestran la comparación de los valores H_{SN} (panel A) y H_{GS} (panel B) entre todos los grupos de pacientes. Cada punto muestra un paciente único. No se muestran diferencias significativas entre los grupos.

De manera diferente, al comparar los índices de funcionalidad entre los distintos grupos, se observa que el grupo de IC presentó valores más elevados que el de CHB y HCC. En el caso de Mf entre 4,9 y 5,7 veces más que CHB y HCC (mediana de CI [IQR] de Mf = 17,5 [9,5-35,6], para CHB = 3,1 [2,7-11,6] y para HCC = 3,6 [1,6-8,0], p-valor relacionado con HCC = 0,006) (Figura 50, panel A). El grupo de los LC mostró la Mf más alta (mediana [IQR] = 30,4 [12,1-51,9]), sin embargo, se incluyeron pocos pacientes en este estudio, por lo que estos resultados deben confirmarse en una cohorte más grande. Se detectó una tendencia similar al considerar el índice π , donde el IC mostró un valor alrededor de 5 veces más alto que los grupos CHB y HCC (mediana [IQR] de 0.03 [0.02-0.04], 0.005 [0.004-0.02] y 0.005 [0.003-0.01] en CI, CHB y HCC respectivamente), p-valor, 0,03 para CHB y 0,006 para HCC) (Figura 50, panel B).

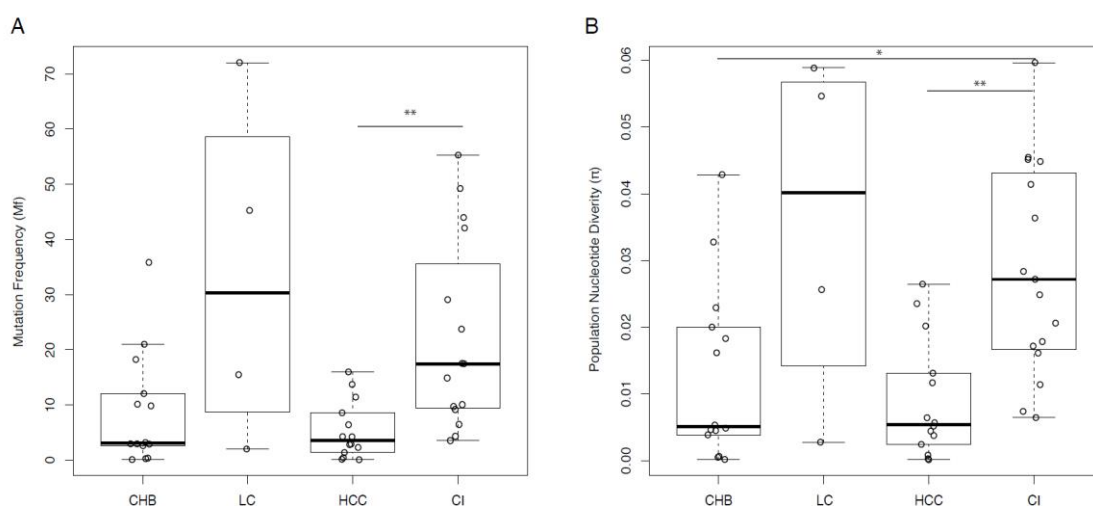


Figura 50. Comparación de los índices de funcionalidad [Frecuencia de mutación (Mf) y diversidad nucleotídica (π)]. Los diagramas de caja muestran la comparación de los valores Mf (panel A) y π (panel B) entre todos los grupos de pacientes. Cada punto muestra un paciente único. Los p-valores se reportan como asteriscos donde el número de asterisco identifica un intervalo específico de p-valores (* $p < 0.05$, ** $p < 0.01$)

Nuestros resultados mostraron que los pacientes con IC presentaron una mayor complejidad de la QS en términos de índices funcionales (Mf y π) en comparación con los pacientes con CHB y HCC. Esto sugiere que en estos dos últimos grupos las variantes altamente replicativas y / o carcinogénicas pueden haber sido seleccionadas debido a su mayor *fitness*, desplazando a las otras variantes con una eficiencia menor para garantizar la replicación y persistencia del VHB. En cambio, la mayor complejidad en los pacientes con IC podría deberse a la presencia en la quasiespecies de variantes altamente mutadas y

probablemente con una baja aptitud para la replicación, pero que a pesar de ello no han sido desplazadas por otras variantes con mayor *fitness*.

3.7. Estudio de la conservación de la secuencia

Al calcular el contenido de información y aplicar el *Sliding Windows*, observamos cuatro regiones de nucleótidos altamente conservadas (Figura 51, panel A y B). Los primeros segmentos híper-conservados abarcan desde los nt 1255 a 1286. La segunda porción estaba entre los nt 1411 y 1435. El tercer y cuarto fragmento abarcaba los nt 1519-1543 y 1575-1605 (Figura 52, panel B). El primer segmento, el tercero y el cuarto coinciden con los resultados obtenidos en el primer estudio, lo cual confirma los resultados anteriormente publicados³⁶⁸. Es de destacar que se observa respecto al primer estudio una nueva región hiper-conservada entre los nt 1411 y 1435.

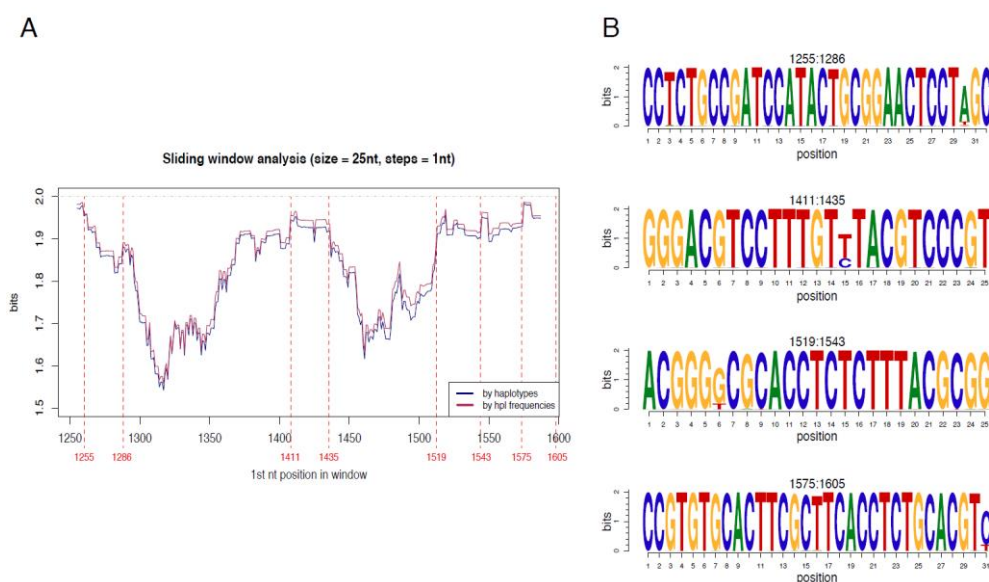


Figura 51. *Sliding window* del fragmento del *HBX* entre los nts 1255 y 1611, en ventanas de 25 nt y con desplazamiento de 1 nt entre ellas (A) y representación del contenido de información de cada posición de nt en los alineamientos de las regiones híper-conservadas identificadas en forma de logo de secuencias (B). Los bits representan el grado de conservación, de 0 (no conservado) a 2 bits (100% de conservación).

Al analizar alineamientos múltiples de las secuencias de aa, se observó una región híper-conservada entre los aa 63 y 77 (Figura 52, panel A y B). De nuevo se confirman los resultados obtenidos en el primer estudio³⁶⁸.

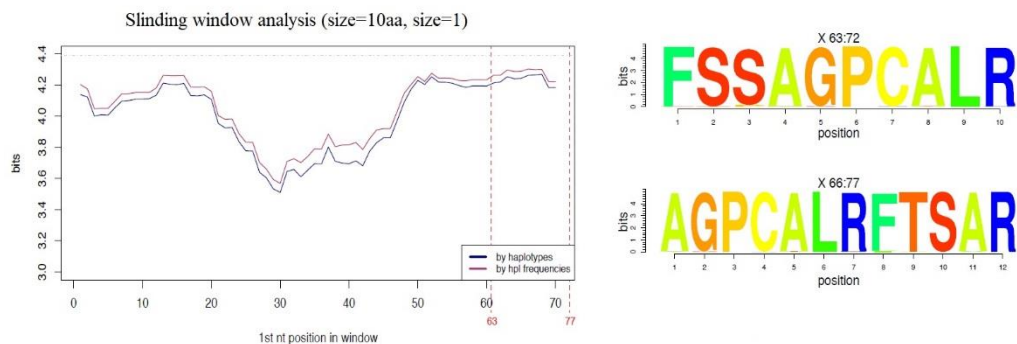


Figura 52 *Sliding window* del fragmento de la HBx entre los aa 1 y 79, en ventanas de 10 aa y con desplazamiento de 1 aa entre ellas (A) y representación del contenido de información de cada posición de aa de las regiones híper-conservadas identificadas en forma de logo de secuencias (B). Los bits representan el grado de conservación, de 0 (no conservado) a 4 bits (100% de conservación).

Al analizar los *Sliding Windows* de cada grupo por separado y compararlos entre sí, se observó que estos grupos presentaban una tendencia similar de conservación tanto a nivel nt como aa (Figura 53, panel A y C para nt y aa respectivamente). A continuación, se comparó esta conservación entre los distintos grupos, mediante el cálculo de la desviación estándar de cada grupo respecto a la media general del contenido de información de todos los grupos. El estudio de la variabilidad intergrupar en la conservación mostró que los IC junto con los LC eran los grupos de pacientes más conservados, principalmente en comparación con el grupo CHB. Al observar el *Sliding Windows* se identificaron dos regiones donde la diferencia entre la conservación del grupo de los IC y los CHB era más significativa. La primera a nivel de nt, situada en la región no codificante del HBx entre nt1250 y nt 1350 ($p = 0.005$ al comparar los IC con CHB) (Figura 53, panel B), donde se han descrito diferentes sitios de iniciación de transcritos de *HBX*. La segunda a nivel aminoacídico en la región de HBx que codifica el dominio putativo de dimerización de HBx³⁷³ (aa 20-50) ($p = 0.019$) (Figura 53, panel D). Una mayor variabilidad a este nivel podría conferir un estado diferente de configuración tridimensional de la HBx que podría influir en su función de transactivación. La menor conservación de los haplotipos de CHB podría sugerir que la variabilidad en estos niveles podría ayudar a que el VHB se adapte mejor al entorno externo, garantizando su replicación. En el caso del grupo de los IC, la mayor conservación de la QS podría indicar una selección de los haplotipos más conservados y probablemente menos replicativos, lo que conduciría a la persistencia viral, ya que esta replicación a bajo nivel no estimularía la respuesta inmune en la célula ni tampoco produciría un daño hepático inmediato. Sin embargo, esta hipótesis debería ser confirmada en futuros estudios.

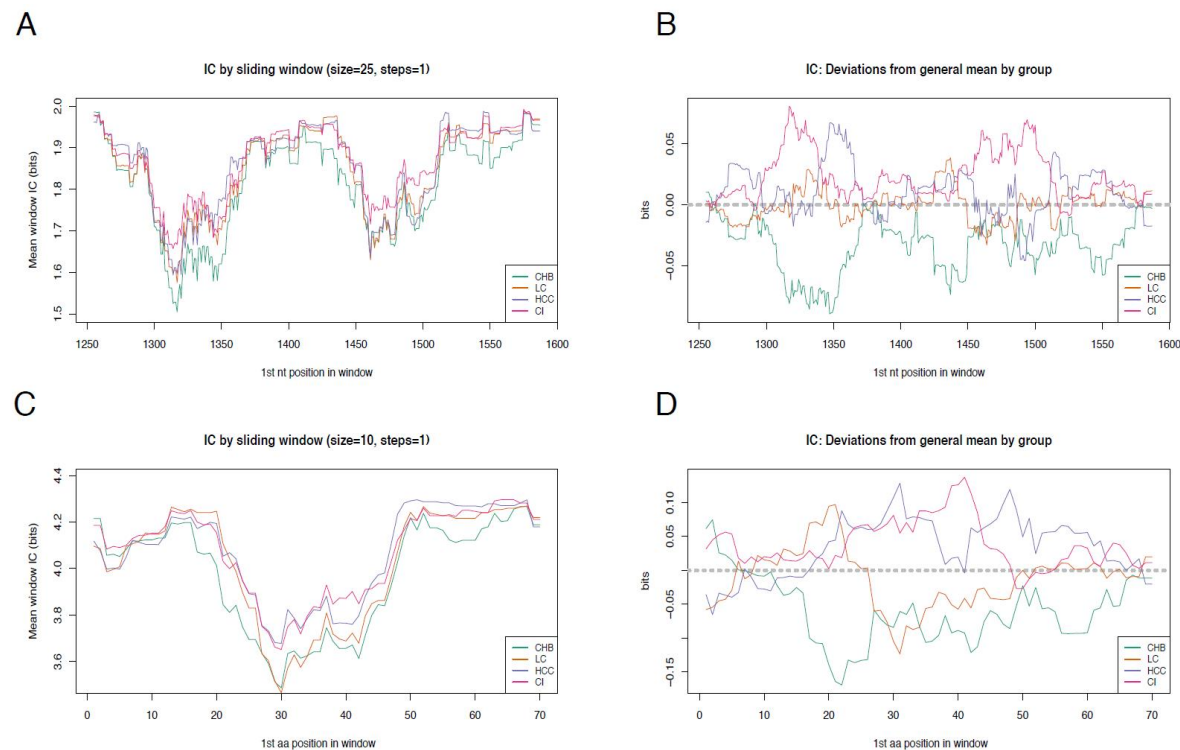


Figura 53 Variabilidad intergrupar de conservación de la secuencia. A. Análisis de *sliding window* de las secuencias nt (1255 – 1611) de los diferentes grupos clínicos (CHB en verde, LC en naranja, HCC en púrpura y IC en rosa). B. Representación de la desviación en relación con la media general del contenido de información de las diferentes ventanas de los *sliding windows* de cada grupo para estudiar la variabilidad intergrupar de la conservación de nt. La media general del contenido de la información se resalta con una línea de puntos grises. C. Análisis de *sliding window* de las secuencias de aa (1 – 79) de los diferentes grupos clínicos. D. Desviación en relación con la media general del contenido de información de las diferentes ventanas de los *sliding windows* de cada grupo para estudiar la variabilidad intergrupar de la conservación de aa.

Estos resultados pueden parecer contradictorios con los mostrados en el análisis de la complejidad del QS (mayor complejidad de IC respecto CHB y HCC), pero representan una observación adicional de la gran sofisticación de la QS del VHB en los pacientes IC. Nuestros datos sugieren que la quasispecies de los IC está compuesta por variantes altamente mutadas a baja frecuencia que, aunque no afectan a la conservación de la población principal, aumentan la complejidad de las QS a nivel funcional (M_f y π). La posible interferencia en la replicación del VHB de estos haplotipos menores debe analizarse más a fondo.

3.8. Identificación de mutaciones específicas de los pacientes IC

Para explicar la replicación viral limitada en los pacientes con IC, se analizó la presencia de mutaciones a nivel aminoacídico en la región estudiada. Encontrar mutaciones en este grupo podría ser útil para clasificar mejor a los pacientes con infección crónica por el VHB HBeAg (-) incluidos en la zona gris, cuyo manejo todavía es difícil debido a la falta de factores que poder clasificarlos adecuadamente como hepatitis o infecciones crónicas³⁷⁰. Al comparar los diferentes grupos no detectamos mutaciones específicas asociadas a los pacientes IC. Sin embargo, al separarlos por genotipos identificamos en los haplotipos de genotipo D las mutaciones genotípico-específicas: A12S, P33S, P46S y T36A / D / G. Estas mutaciones se correlacionaron entre ellas con un coeficiente de correlación o rho (r) ≥ 0.68 y un valor $p \leq 0.00001$ (Figura 54, panel A), formando un patrón de mutaciones (Figura 54, panel B).

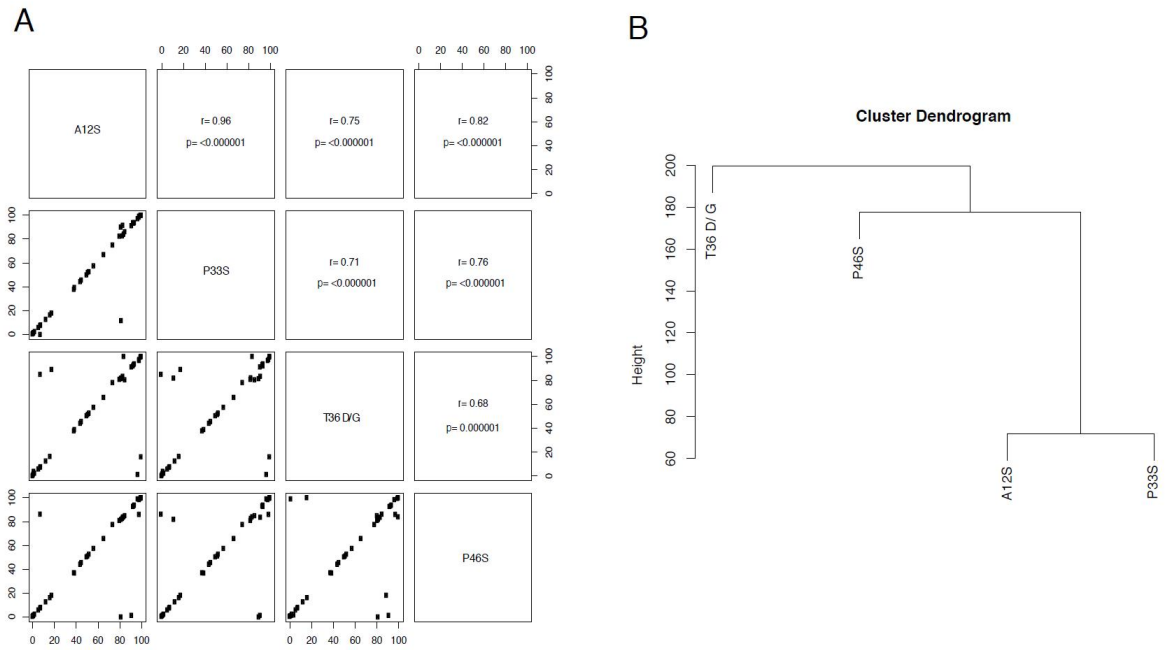


Figura 54 Patrón de mutaciones en pacientes IC de genotipo D. A. Correlación de Spearman entre la frecuencia de estas mutaciones. Se informan los valores de rho y p-valor. B. Dendrograma que representa los clústers jerárquicos de las mutaciones evidenciadas en los haplotipos de genotipo D. El eje y muestra la medida de la proximidad de cada mutación o clúster.

Al estudiar la frecuencia de este patrón de mutaciones entre los diferente grupos clínicos, se evidenció que en los pacientes IC este patrón presentaba una frecuencia mediana alrededor de 16 veces más alta que el grupo CHB (mediana de IC [IQR] = 81.7 [51.7-93.3] vs CHB [IQR]=5.1 [1.4-31.0], $p = 0.004$) (Figura 55).

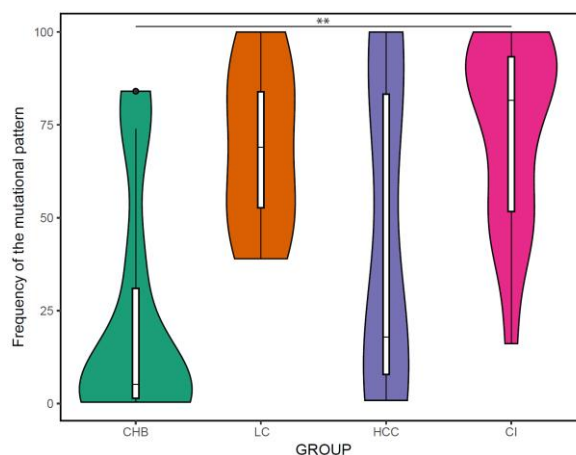


Figura 55 El *violin plot* muestra la comparación de las frecuencias del patrón de mutaciones (A12S/P33S/P46S /T36D-G) entre los distintos grupos clínicos. El rango intercuartil, la mediana y el intervalo de confianza están reportados en el diagrama de cajas interno. La forma de violín muestra la densidad de probabilidad de que el patrón mutacional esté presente en una determinada frecuencia (las secciones más anchas indican la mayor probabilidad de cada grupo clínico de presentar el patrón de

mutaciones con una determinada frecuencia). El p-valor entre CI y CHB (<0.01) está indicado por el doble asterisco.

Curiosamente, el patrón (A12S / P33S / P46S / T36G-D) se caracteriza principalmente por la sustitución de un aa hidrófobo (alanina o prolina) por un aa polar (serina). Esto es un dato relevante, ya que dentro del dominio de dimerización rico en Ser / Pro hay algunos aa polares altamente conservados (como Ser25 y Ser41) que son susceptibles de modificaciones post-traduccionales⁸⁹. La adición de esta serina podría proporcionar una nueva diana de fosforilación y O-β-glicosilación, lo que podría alterar la tridimensionalidad de la proteína HBx y, por consiguiente, modificar su actividad transactivadora. Es de destacar que se necesitan más estudios para determinar los efectos de las mutaciones en la estructura tridimensional de HBx y su relación con la expresión del VHB.

A raíz de los resultados obtenidos en el punto anterior y como un estudio derivado de esta tesis, se ha comprobado *in vitro* la eficiencia replicativa (*fitness*) del VHB (ver *Anexo 2: Estudios in vitro sobre el efecto de las mutaciones sobre la replicación del VHB*)

4. Conclusiones

1. Los segmentos híper-conservados identificados en el primer estudio se confirman en el presente estudio, en un grupo de pacientes más grande y con los diferentes estadios clínicos de la infección crónica por el VHB HBeAg (-) mejor representados. Esto reafirma su potencial utilidad como dianas para una terapia génica dirigida. Además, se ha descrito una nueva porción híper-conservada entre los nt 1411 y 1435.
2. Se ha evidenciado una mayor complejidad y conservación de la QS en los pacientes IC, probablemente debido a la presencia de un grupo de variantes altamente mutadas que se encuentran en baja frecuencia en la QS, por lo que afectan poco a la conservación.
3. La presencia de un patrón de mutaciones en las secuencias de los pacientes con IC de genotipo D posiblemente podría relacionarse con la baja replicación viral observada y podría sugerir una evolución de la población viral de forma genotipo-específica.
4. La presencia de mutaciones aa específicas refuerza la necesidad del genotipado adecuado del VHB durante el seguimiento del paciente

5. Manuscrito completo

En revisión para su posible aceptación en una revista de alto impacto (Gut, factor de impacto en el 2017 de 17,016).

HBeAg-ve chronically infected patients: sophisticated viral quasispecies with a genotype-related pattern of mutations in the hepatitis B X gene.

Cortese Maria Francesca^{a,1,2}, González Carolina^{a,2} , Gregori Josep^{1,3}, Riveiro Barciela Mar^{4,5}, Casillas Rosario^{1,2}, Godoy Cristina^{1,2}, Sopena Sara^{2,4}, Yll Marçall^{1,2}, Quer Josep^{1,4}, Rando Ariadna², Lopez-Martinez Rosa², Esteban-Mur Rafael^{4,5}, Buti Maria^{4,5}, Tabernero David⁴, Rodríguez-Frías Francisco^{2,4}

^a Cortese MF and González C contributed equally to this paper.

¹Vall D'hebron Research Institute, Liver Unit, Barcelona, Spain; ²Vall D'hebron University Hospital, Biochemistry and Microbiology/Liver Pathology Unit, Barcelona, Spain; ³Roche Diagnostics SL, Sant Cugat del Vallès, Spain; ⁴Ciberehd, Instituto De Salud Carlos III, Madrid, Spain; ⁵Vall D'hebron University Hospital, Department of Internal Medicine/Liver Unit, Barcelona, Spain.

ORCID number: Maria Francesca Cortese (0000-0002-4318-532X); Carolina González (0000-0002-0169-5874); Josep Gregori (0000-0002-4253-8015); Mar Riveiro-Barciela (0000-0001-9309-2052); Rosario Casillas (0000-0002-6758-6734); Marçal Yll (0000-0002-7030-3360); Cristina Godoy(0000-0001-5037-1916); Sara Sopena (0000-0002-3309- 5486); Josep Quer (0000-0003-0014-084X); Ariadna Rando (0000-0003-4555-7286); Rosa Lopez-Martinez (0000-0002-8450-6986); Rafael Esteban (0000-0001-5280-392X); Maria Buti (0000-0002-0732-3078); David Tabernero (0000-0002-1146-4084);Francisco Rodríguez-Frías (0000-0001-9058-4641).

ABSTRACT

Background:

Patients with HBeAg-negative chronic infection (CI) are a generally overlooked clinical group due to their low viremia. The HBx protein, encoded by *HBX*, has a key role in viral replication. This study aimed to analyze the viral quasispecies at the 5' end of *HBX* in HBeAg-negative CI patients and compare it with that of patients in other clinical stages.

Methods:

Fifty-three HBeAg-negative patients included: 16 with CI, 15 chronic hepatitis B (CHB), 6 liver cirrhosis (LC), and 16 hepatocellular carcinoma (HCC). The region between nucleotides (nt) 1255 and 1611 was analyzed by next-generation sequencing to determine quasispecies complexity and conservation. Amino acid (aa) mutations, detected by aligning haplotypes to the genotype consensus, were tested *in vitro*.

Results:

CI patients showed higher complexity in terms of mutation frequency and nucleotide diversity (respectively, 5- and 6-fold higher than CHB and HCC; $p < 0.05$). Four hyper-conserved regions were highlighted at the nt level (1255-1286, 1411-1435, 1519-1543, 1575-1605), and one at the aa level (63-72). CI patients showed higher quasispecies conservation than CHB, mainly between nt 1300-1375 and aa 20-50 ($p < 0.05$). A genotype D-specific pattern of mutations (A12S/P33S/P46S/T36D-G) was identified in CI

(median frequency, 81.7%). A reduction of up to 1.5log in HBV release was observed *in vitro* in presence of this pattern.

Conclusion:

CI patients showed a more complex and conserved viral quasispecies than the other groups. The genotype-specific pattern of mutations could partially explain the low viremia observed in these patients. The hyper-conserved regions detected could be valuable targets for gene therapy.

What is already known about this subject?

- Chronically infected HBeAg-negative patients usually have a favorable prognosis, but a non-negligible percentage is at risk of HCC or infection reactivation, making this group of great scientific interest.
- The HBx protein, encoded by the homonym *HBX* gene, covers a key role in hepatitis B virus (HBV) replication, being essential for initiating and trans-activating viral expression.
- Little is known about the viral quasispecies populations in chronically infected patients, particularly in a region of HBX gene known to be highly conserved.

What are the new findings?

- A more sophisticated viral quasispecies showing greater complexity and conservation than those of the other patient groups was observed in CI patients. This was likely due to the presence of

highly mutated haplotypes at low frequency, which did not affect the overall nt and aa conservation of the quasispecies, but did increase its complexity.

- Regardless of the clinical stage or viral genotype, some hyper-conserved regions were observed, and these could be promising targets for gene therapy.
- A genotype D-specific pattern of mutations that reduced HBV expression *in vitro* was mainly seen in CI patients; this could partially explain the low viremia state observed in this population.

How might it impact on clinical practice in the foreseeable future?

- Selection of a mutation pattern associated with decreased viral expression could be used as a factor to improve the classification of chronically infected HBeAg-ve patients.
- The presence of genotype-specific mutations and the co-existence, in the same viral population, of different genotypes, highlight the importance to adequately classifying the hepatitis B virus genotype.
- The hyper-conserved regions identified in the present study regardless of the clinical stage or viral genotype could be feasible targets for new therapeutic approaches based on gene therapy.

INTRODUCTION

Currently, around 257 million people worldwide are chronically infected by the hepatitis B virus (HBV), and HBV infection remains the main cause of death due to viral hepatitis (WHO report, July 2018). HBV has a small (3.2 kb), partially double-stranded DNA genome with 4 highly overlapping open reading frames (ORFs): polymerase, surface, Core, and X genes[1]. HBV has been classified into 8 genotypes (A-H) and further divided into sub-genotypes. Replication of this virus requires a reverse transcription step, and because the viral reverse transcriptase lacks proofreading activity, the HBV genome acquires spontaneous mutations, reaching a mutational rate resembling that of RNA viruses[2]. As result of this variability, HBV circulates as quasispecies, a complex mixture of closely related viral genetic variants referred to as haplotypes.

The HBx protein, encoded by *HBX* gene, covers a key role in HBV infection and disease progression, and it is an essential element to initiate a productive infection[3, 4]. Moreover, it allows the expression of covalently closed-circular DNA (cccDNA) by acting on its epigenetic control [3, 5] and it interacts with cellular protein, such as DDB1 and cyclic-AMP-response element binding protein (CREB) involved in HBV replication[3, 6]. Furthermore, it interferes with other cellular pathways, which makes it a multifunctional and pleiotropic protein with a key role in the development of hepatocellular carcinoma (HCC)[7].

HBV patients are classified as having different clinical stages of the disease based on determination of several viral and biochemical markers (eg, HBeAg, viremia, alanine aminotransferase [ALT])[8]. Those with HBeAg-negative (HBeAg-ve) chronic HBV infection, previously known as “inactive carriers”, constitute a unique clinical group characterized by viremia ≤ 2000 IU/mL, low HBsAg levels[8], normal ALT concentrations, and minimal or absent hepatic necroinflammation. These patients have a favorable prognosis, but the risk of HCC and hepatitis reactivation after immunosuppressive therapy persists, and continuous periodic follow-up testing is advisable[9, 10].

By studying these patients, valuable information about the mechanisms that enable control of viral replication could emerge and suggest potential targets for new therapeutic approaches. However, due to their low HBV viral load, this patient group is usually overlooked and, to date, the mechanism behind their low replication rate is currently unknown. Could the variability and complexity of the viral genome be determinant factors in this interesting clinical stage? To explore this question, we analyzed samples from a group of well-characterized HBeAg-ve chronically infected (CI) patients, and compared the results with those of a chronic hepatitis group, including patients with liver cirrhosis and HCC. Considering the role of HBx in viral replication and disease progression, we focused our study on a region that includes the 5' end of HBX gene

and the upstream non-coding region. In a previous study, we observed some hyper-conserved regions in this portion of HBX[11], but we did not examine the conservation profile related to different clinical groups at that time.

The aim of this study was to identify genetic factors in the HBV quasispecies that characterize CI patients, in order to elucidate the control mechanisms affecting their viral replication.

MATERIALS AND METHODS

Patient and public involvement.

The study was approved by the Ethics Committee of Vall d'Hebron Research Institute. Patients were enrolled during the inclusion stage from the population attending the outpatient clinic of Vall d'Hebron Hospital (Barcelona, Spain). All patients were informed about the aims of the project, and signed an informed consent form. After inclusion, patients were not further involved in this work.

Patients and samples.

Plasma samples were collected from 16 patients with chronic infection (CI), and 15 chronic hepatitis (CHB), 6 liver cirrhosis (LC) and 16 hepatocellular carcinoma (HCC) patients. Patients were all HBV-monoinfected, HBeAg-ve, and treatment-naïve. Viral load was >100

IU/mL, the limit of PCR sensitivity. Demographic, virologic, and serologic characteristics are reported in Table 1.

HBsAg and HBeAg were tested using commercial enzyme immunoassays (COBAS 8000 analyzer, Roche Diagnostic). HBV-DNA was quantified by real-time PCR (COBAS 6800, Roche Diagnostics) with a detection limit of 10 IU/mL.

Table 1. Main virologic and serologic characteristics of the clinical groups enrolled in the study

	CHB (n=15)	LC (n=6)	HCC (n=16)	CI (n=16)
Male (%)	60 (n=9)	100 (n=6)	6.2 (n=1)	43.7 (n=7)
Age Median [Q1-Q3]	37 [30.5-51.5]	46 [42-56.8]	60 [54.5-68.3]	50 [35.7-57.2]
Genotype ⁽¹⁾ (%)				
A	20	0	43.75	0
C	20	0	12.5	6.25
D	6.7	83.3	43.75	81.25
E	26.7	0	0	0
F	13.3	0	0	12.5
H	13.3	16.7	0	0
HBV DNA (logIU/mL) Median [Q1-Q3]	4.6 [3.9-5.4]	4.6 [4.0-6.0]	5.6 [5.0-6.3]	3.1 [2.2-3.3]
ALT (IU/mL) Median [Q1-Q3]	45.5 [32.3-56.0]	45.5 [35.9-69.0]	61 [48.5-112.5]	28 [19.3-31.3]

CHB, chronic hepatitis B; LC, liver cirrhosis; HCC, hepatocellular carcinoma; CI, Chronic infection; Q1, 25th percentile; Q3, 75th percentile; HBV, hepatitis B virus; ALT, alanine aminotransferase.

⁽¹⁾ Genotype was evaluated by NGS sequencing, as described [11].

HBV X gene amplification and sequencing.

The region of interest encompassed nucleotide (nt) 1255 to 1611, which includes the 5' end of the HBX coding region (nt 1374-1611, corresponding to aa 1-76 in the coded protein) and the upstream non-coding region.

HBV DNA was extracted from 500 µL of plasma sample with the QIAamp UltraSens Virus Kit (QIAGEN), according to the manufacturer's instructions. The region under study was amplified using a 3-round nested PCR protocol that enabled amplification of samples with viremia >100 IU/mL. Briefly, the first-round PCR was performed using external primers (forward 5'-TGTATTCCCATCCCATCATC at position nt 599 and reverse 5'-AGWAGCTCCAAATTCTTTATAAGG, at position nt 1936) with the following protocol: 95°C for 5 minutes followed by 35 cycles of 95°C for 20 s, 53°C for 20 s, 72°C for 15 s, and finally, 72°C for 3 minutes. The second-round PCR (using primers carrying the M13 universal adaptor) and third-round PCR (using primers including a unique multiplex identifier sequence [MID] for each sample/patient) were performed as described [11]. To ensure that the 3-round PCR did not change the viral populations, a control run with a 2-round PCR (excluding the external PCR) was performed in samples from 5 of the patients included. All PCR steps were performed using the high-fidelity Pfu Ultra II DNA polymerase (Stratagene, Agilent Technologies). PCR products were purified using the QIAquick Gel Extraction Kit (QIAGEN) following

the manufacturer's instructions, and DNA quality was evaluated using the Agilent 2200 TapeStation (Agilent Technologies). Purified PCR products were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) to equilibrate the representation of each sample in the pool, and then sequenced by next generation sequencing (NGS) on the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA).

The sequences obtained were bioinformatically filtered, as described[12], which resulted in unique sequences covering the full amplicon (haplotypes) that form the viral quasispecies (QS). The haplotypes included in the following analysis had common reverse and forward sequences, and an abundance of $\geq 0.1\%$ in the QS complexity analysis and $\geq 0.25\%$ in the study of QS conservation. Haplotypes were genotyped as previously reported[11] (Supplementary Table 1).

Quasispecies complexity.

QS complexity was determined by applying four parameters: Shannon entropy, the Gini-Simpson index, mutation frequency (Mf), and the nucleotide diversity index (π)[13]. Shannon entropy and the Gini-Simpson index are abundance indices that measure haplotype diversity based on the number of haplotypes and their relative frequencies. The functionality indices include the mutation frequency, which measures the genetic diversity of the viral population with respect to the most prevalent haplotype, and the nucleotide diversity index, which measures the genetic

diversity as the average number of mutations per site between each pair of haplotypes in the viral population[13] (Supplementary Table 2).

Intergroup differences in QS complexity were evaluated using the Kruskal-Wallis test and the Dunn posthoc test with R language software (3.2.3). P-values were adjusted with the Bonferroni correction and those <0.05 were considered statistically significant.

Analysis of quasispecies conservation.

Sequence conservation was determined by calculating the information content of each position (both nt and aa) in a multiple alignment of all haplotypes obtained with NGS, followed by sliding window analysis, as previously described by our group[11].

To determine the intergroup variability in sequence conservation, the standard deviation of the mean overall information content was calculated for each group of patients, and the Wilcoxon test was applied for the comparisons (R software 3.2.3).

Mutation analysis.

The sequences of all haplotypes were aligned with a reference consensus sequence of the same genotype to detect aa mutations. Genotype consensus was generated by aligning the region of interest (nt 1255 to

1611) extracted from full-length HBV genome sequences obtained from GeneBank (13 sequences for genotype A, 23 for genotype C, 17 for genotype D, 8 for genotype E, 10 for genotype F, and 5 for genotype H; Supplementary Table 2). Nucleotide variations that determine a change in the coded aa were considered to be mutations and their frequency per patient was calculated as the sum of the relative frequencies of each mutated haplotype in the patient's QS population. Spearman correlation testing and hierarchical cluster analysis were performed (R software 3.2.3). The frequency of mutations in the CI group was compared to that of the other groups using the Kruskal-Wallis test plus the posthoc Dunn test. Bonferroni correction was applied to the p-values and those <0.05 were considered statistically significant.

***In vitro* HBV expression in the presence of mutations**

Mutations that were differentially represented between the patient groups were tested *in vitro* by HBV monomer linear transfection to allow the formation of cccDNA[3].

The HBV monomer was obtained from the pCRII.HBV.ayw plasmid (kindly donated by Prof. Massimo Levrero and Dr. Laura Belloni), which contains a full-length HBV genotype D genome (subtype ayw). Mutations were introduced by site-directed mutagenesis (QuikChange Lightning Site-Directed Mutagenesis kit, Agilent) using the manufacturer's procedure, and mutated plasmids were isolated using the Plasmid Midi

kit (Qiagen) according to the manufacturer's instructions. Linear HBV genomes, both wild-type (wt) and mutated, were obtained by digestion using EcoRI and PvuI, extracted from the gels using the QIAquick Gel Extraction Kit (Qiagen), and later quantified with Qubit Fluorimeter (ThermoFisher).

HepG2-hNTCP cells were cultured with Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin (100 U/mL), streptomycin (100 µg/mL), Glutamax (2 mM), and Puromycin (5 µg/mL). To synchronize the cells, they were treated with DMSO 2.5% at least 14 days before plating [14-16]. The day before transfection, cells were plated in 24-well plates at a density of 60,000 cells/mL in DMEM with 10% FBS (complete medium). The next day, cells were transfected with 250 ng/well of wt or mutated linear HBV monomer using the TransIT-X2 Dynamic Delivery System (Mirus). The pmaxFP-green plasmid (Amara Biosystem) was added to each well for control of transfection (1:10 relative to HBV DNA). The next day, transfected cells were extensively washed with PBS and the complete medium was replaced. The medium was changed every 2 days, and supernatants were collected at 5 days post-transfection.

Supernatants were treated with 1 mg/mL of DNaseI (Sigma) in the presence of 25 mM of MgCl₂ to remove residual linear DNA. After 1h of digestion at 37°C, the reaction was stopped using 25 mM of EDTA.

HBV DNA was quantified by real-time PCR (COBAS 6800, Roche

Diagnostics). The Kruskal plus Dunn test (R software 3.2.3) was used to determine differences in HBV DNA production. P-values <0.05 were considered statistically significant.

RESULTS.

CI: an inactive state of infection with higher functional QS complexity.

The 3-round PCR protocol did not affect the composition of the HBV quasispecies, as confirmed by comparing the 3-round haplotype distribution with that obtained by 2-round PCR (Supplementary Figure 1).

Only patients with $\geq 10,000$ reads (sequencing data) were included in the analysis; that is, 16/16 CI patients, 15/15 CHB, 4/6 LC, and 16/16 HCC, 51/53 patients in total. After applying the quality filters, 1,156,140 sequences were obtained, which yielded a median [Q1-Q3] of 17,507 [13,718-24,901] sequences per patient. NGS data have been submitted to the GenBank SRA database (BioProject accession number PRJNA437055, BioSample accession numbers in Supplementary Table 3).

In the comparison of viral quasispecies complexity, no differences were found between CI patients and the other groups for Shannon entropy, (median [Q1-Q3] 22660 [17490-26200], 19885 [11815-24755], 29315 [23867-31630], and 17020 [3456-21303], respectively, for CI, CHB, LC and HCC)

(Table 2). Similar findings were observed for the Gini-Simpson index with a median [Q1-Q3] of 0.8 [0.8-0.9] for CI, 0.7 [0.5-0.8] for CHB, 0.9 [0.8-0.9] for LC, and 0.7 [0.4-0.8] for HCC (Table 2).

Table 2. Viral quasispecies complexity in the different clinical groups.

	Sn Median [Q1-Q3]	G Median [Q1-Q3]	Mf Median [Q1-Q3]	π Median [Q1-Q3]
CHB	19885 [11815-24755]	0.7 [0.5-0.8]	3.1 [2.7-11.6]	0.005 [0.004-0.02]
LC	29315 [23867-31630]	0.9 [0.8-0.9]	30.4 [12.1-51.9]	0.04 [0.02-0.06]
HCC	17020 [3456-21303]	0.7 [0.4-0.8]	3.57 [1.60-8.03]	0.005 [0.003-0.01]
CI	22660 [17490-26200]	0.8 [0.8-0.9]	17.5 [9.5-35.6]	0.03 [0.02-0.04]

CHB: chronic hepatitis; LC: liver cirrhosis; HCC: hepatocellular carcinoma; CI: chronic infection; IQR: interquartile range; Q1: 25th percentile; Q3: 75th percentiles; Sn: Shannon entropy; G: Gini-Simpson Index; Mf: mutation frequency; π : nucleotide diversity.

By contrast, in terms of the functional parameters, the CI group showed respectively 4.9- and 5.7-fold higher mutation frequency values than CHB and HCC (median [Q1-Q3] in CI = 17.5 [9.5-35.6], CHB = 3.1 [2.7-11.6], and HCC = 3.6 [1.6-8.0], p-value related to HCC = 0.006) (Table 2 and Figure 1, panel A). LC showed the highest mutation frequency (median [Q1-Q3] = 30.4 [12.1-51.9]); however, as few patients were included in this

study, these results should be confirmed in a larger cohort. Consistent findings were recorded for the nucleotide diversity index, in which CI showed a value around 5-fold higher than the CHB or HCC groups (median [Q1-Q3] 0.03 [0.02-0.04], 0.005 [0.004-0.02], and 0.005 [0.003-0.01], respectively, for CI, CHB ($p=0.03$), and HCC ($p=0.006$) (Table 2 and Figure 1, panel B).

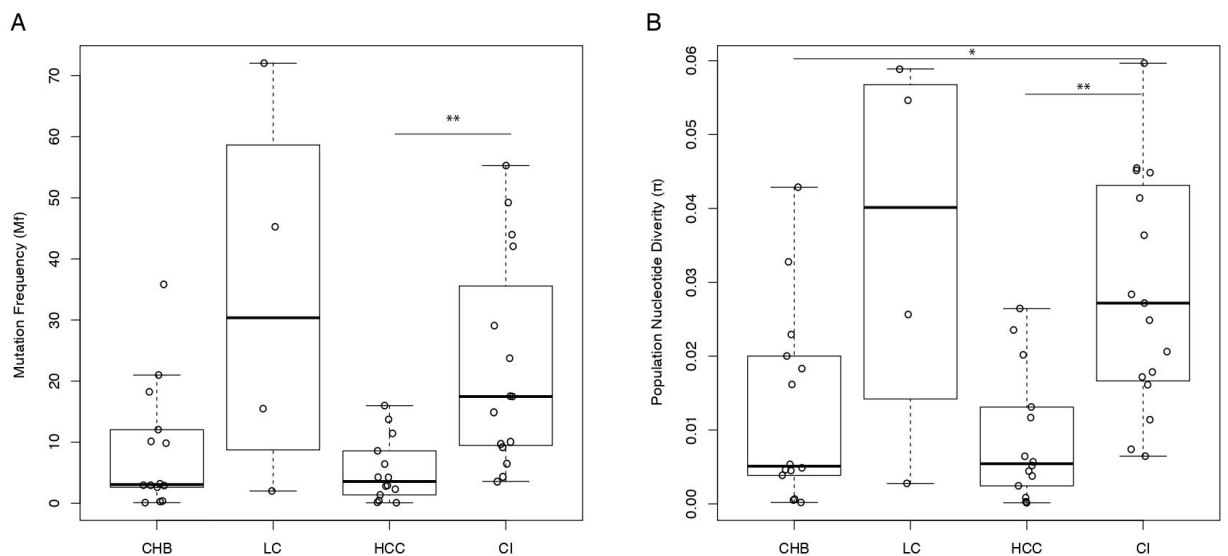


Figure 1. Mutation frequency (Mf) and population nucleotide diversity (π) comparisons. Boxplots show the comparisons of Mf (panel A) and π (panel B) values between the clinical groups. Each dot represents a single patient. Asterisks above the boxes denote specific p-values (1 asterisk <0.05 ; 2 asterisks <0.01). CHB, chronic hepatitis B; CI, chronic infection; HCC, hepatocellular carcinoma; LC, liver cirrhosis.

Similar quasispecies conservation between patient groups: information content study.

After calculating the information content and applying sliding window analysis, four highly conserved nucleotide regions were found (Figure 2, panels A and B), in keeping with our previous report[11]. The conservation results did not change when haplotype frequencies were analyzed. The first hyper-conserved segments encompassed nt 1255 to 1296, with a single nt in position 1294 showing a conservation percentage of around 70% (Figure 2, panel B). The second region was between nt 1411 and 1435, with position 1425 showing two possible nucleotides with low-intermediate conservation. The third and fourth fragments encompassed nt 1519 to 1543 and nt 1575 to 1605.

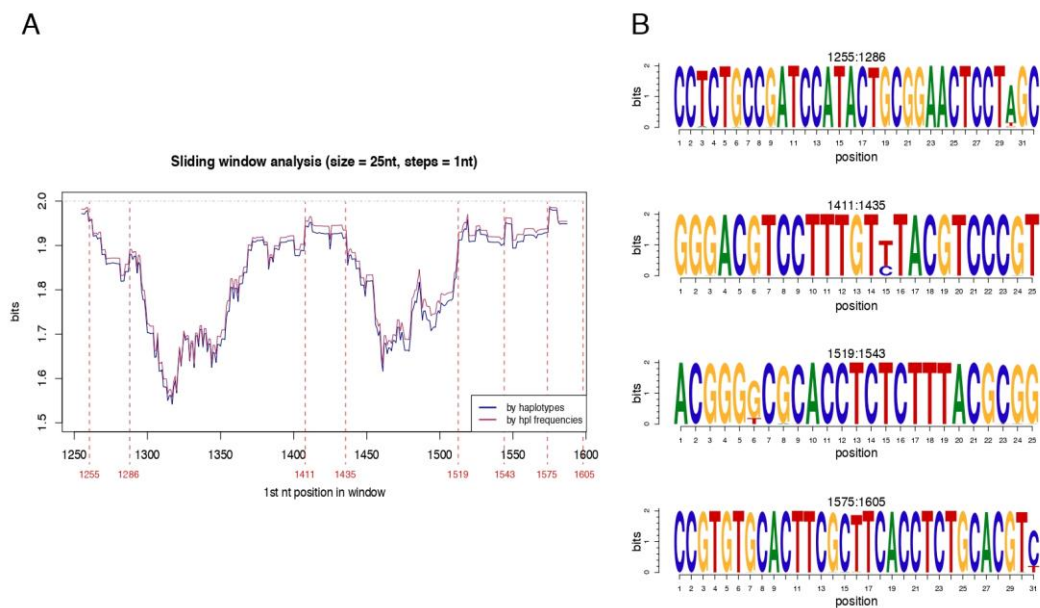


Figure 2. Conservation in QS multiple nucleotide sequences alignment. A. The sliding window analysis is the result of the mean information content (IC, bits) of the 25-nt long

windows with a displacement between them of 1-nt obtained by the multiple alignment of all the haplotypes (blue line) and taking into account their frequencies (red line). B. Representation by sequences logos of the information content of the most conserved regions detected in sliding window analysis. The total height of each stack of letters depicts the IC of each nt position, measured in bits (Y-axis): 0 bits being the minimum and 2, the maximum conservation. The hyperconserved region extremes are reported.

Analysis of multiple alignments of the aa sequences showed a hyper-conserved region between aa 63 and 77 (Figure 3, panel A and B).

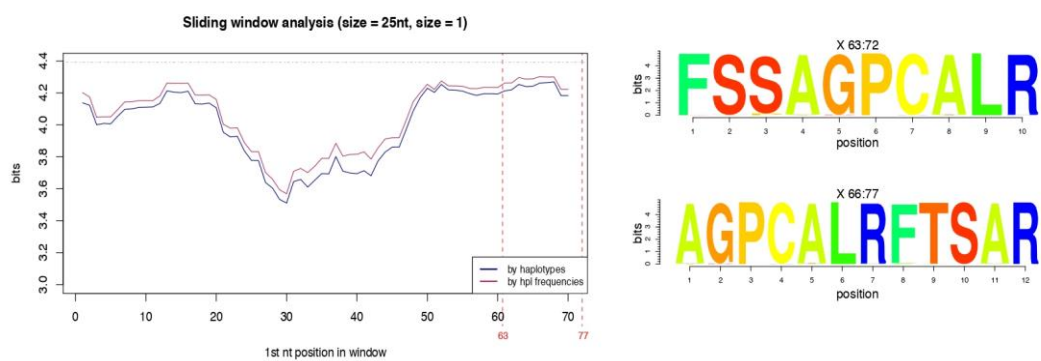


Figure 3. Conservation in QS multiple amino acid sequences alignment. A. The sliding window analysis of the 10-aa long windows with a displacement between them of 1-aa obtained by the multiple alignments of all the aa haplotypes (blue line) and taking into account their frequencies (red line). B. Representation by sequences logos of the information content of the most conserved aa regions detected in sliding window analysis. The height of the letters indicates the grade of conservation, from a maximum of 4 bits (100% of conservation) to a minimum of 0. The positions of hyper-conserved region are reported.

Comparison of the information content standard deviation of each group relative to the overall mean found that conservation at both the nt and aa level was similar but not identical in all groups (Figure 4, panels A and C). Examination of the intergroup variability in nt conservation showed that CI and LC sequences were the most highly conserved, mainly in the non-coding portion encompassing nt 1300 to 1375 ($p=0.005$ between the CI and

CHB information content deviation) (Figure 4, panel B). The aa comparison yielded similar results: CI showed higher conservation than CHB, especially in the region between aa 20 and 50 ($p=0.019$) (Figure 4, panel D).

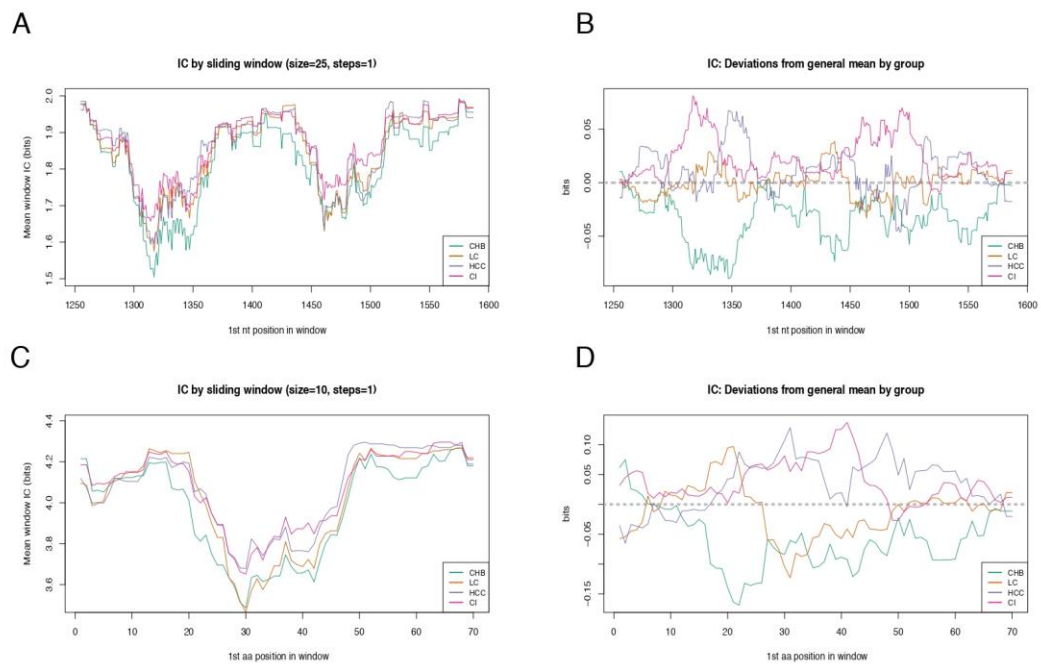


Figure 4. Information content intergroup variability. A. Sliding window analysis of the nt sequences among the different clinical groups. Each line displays a specific clinical group (CHB in green, LC in orange, HCC in purple and CI in pink). B. Representation of the intergroup variability in nt-conservation. Each line indicates a group of patients. The information content general mean is highlighted with grey dot line (positive deviation indicates more conservation, whereas negative deviation indicates less conservation). C. Sliding window analysis of the aa sequences among the different clinical groups. D. Intergroup variability in aa-conservation. Grey dot line displays the general mean of information content.

Genotype-specific pattern of mutations in CI patients.

To determine whether the presence of certain aa mutations might enable differentiation between the clinical groups, all haplotypes were aligned with each specific genotype consensus, and some genotype-specific changes were observed (Supplementary Figure 2). None of the mutations in haplotypes from genotypes A, C, E, F or H were differentially represented between the four clinical stages. However, in the case of genotype D, the A12S, P33S, P46S, and T36A/D/G mutations were more highly represented in one specific group, CI patients. These mutations correlated with each other ($\rho \geq 0.68$, $p \leq 0.00001$) (Figure 5, panel A), thereby forming a pattern of mutations (Figure 5, panel B). This pattern was found in CI patients at a median [Q1-Q3] frequency around 16-fold higher than in the CHB group (CI 81.7 [51.7-93.3] vs CHB 5.1 [1.4-31.0], $p=0.004$) (Figure 6).

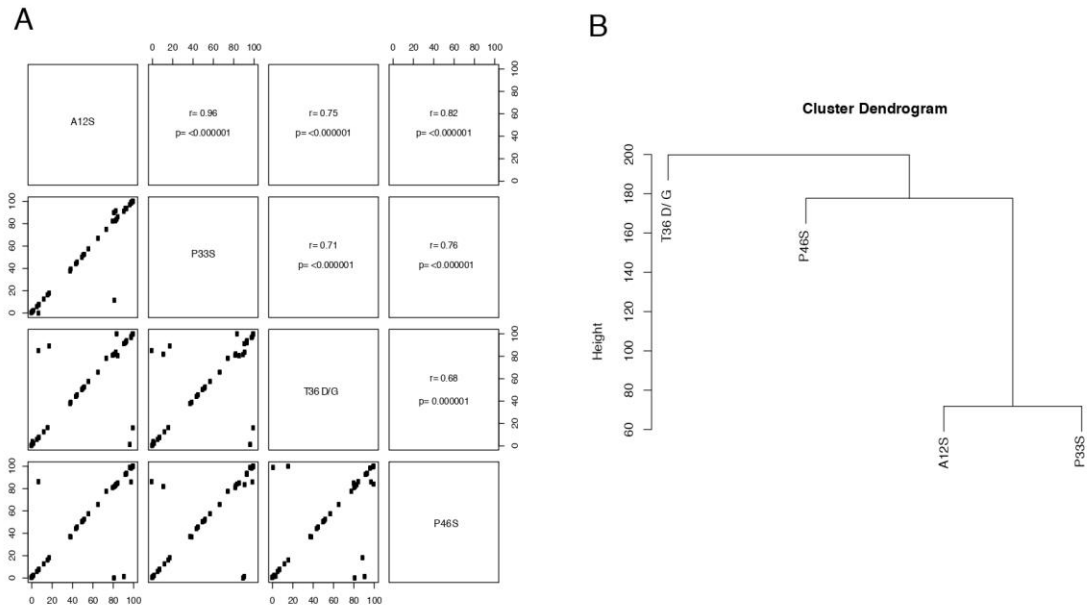


Figure 5. Pattern of mutations in genotype D haplotypes. A. Spearman correlation between the amino acidic mutations frequency highlighted in genotype D haplotypes. Each dot represents the mutation frequency in each patient. Rho and p-values are reported. B. Hierarchical cluster dendrogram of the observed mutations. The y-axis is a measure of closeness of either individual mutation or cluster.

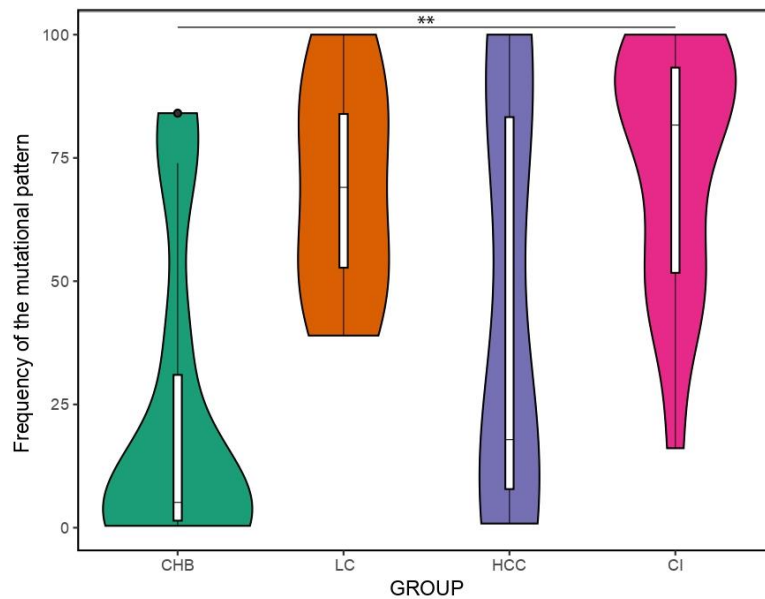


Figure 6. Frequency of the mutational pattern. The violin plot shows the comparison of the frequency of the mutational pattern between the different clinical groups. The internal white bar represents the interquartile range with median in black horizontal line,

whereas the vertical line shows the 95% confidential interval. The violin shape indicates the probability density; where wider sections display higher probability that patient will take a given frequency of mutational pattern. P-values are reported as asterisks where the number of asterisks identifies a specific interval of p-value (1 asterisk for respectively <0.05 and two for <0.01).

Lower *in vitro* HBV expression in the presence of the mutation pattern.

To investigate the effects of HBx aa changes on HBV expression, mutations were tested *in vitro* following the order of hierarchical clustering. All mutations were found to reduce viral particle release at 5 days post-transfection in cell supernatants relative to wt HBV (Figure 7). Specifically, HBV DNA showed a >6.2-fold decrease in the presence of the double mutant A12S/P33S and the triple mutant A12S/P33S/P46S (median [Q1-Q3] 4.83E+04 IU/mL [1.96E+04-4.84E+04], and 3.57E+04 IU/mL [3.12E+04-8.03E+04], respectively, versus 2.20E+05 [1.37E+05-6.46E+05] for wt HBV). Furthermore a 9.8-fold (2.26E+04 IU/mL [1.52E+04-4.56 E+04]) reduction occurred in the presence of the quadruple mutant A12S/P33S/P46S/T36G (p-value <0.05). Overall, the decrease in HBV viremia was between 0.8 and 1 log. Finally, in the presence of the quadruple mutant A12S/P33S/P46S/T36D, there was a 23.4-fold reduction in HBV DNA (9.94E+03[5.10E+03-9.40E+03] IU/mL), which accounted for a 1.4-log drop in HBV viremia relative to the wt virus (p-value=0.00012).

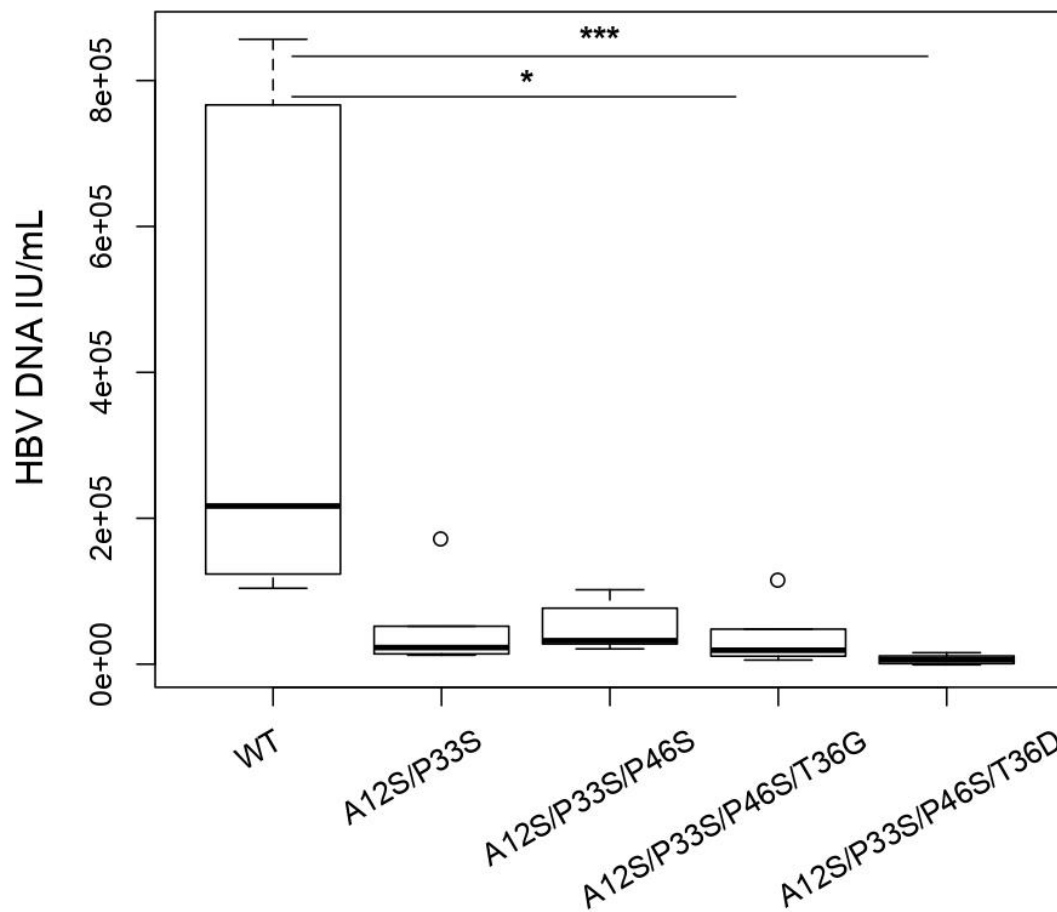


Figure 7. *In vitro* HBV expression in presence of the HBx mutations. The boxplot represents the HBV particles release (HBV DNA in IU/mL) in cells supernatant at 5dpt in presence of wild type (wt) or mutated HBV monomers. A Dunn's multiple comparison test was implemented and the obtained Bonferroni-corrected p-values are reported as asterisks, where the number of asterisk identify a specific interval of p-value (1 asterisk for <0.05 and 2 asterisks between 0.001 and 0.01 and 3 for <0.001).

DISCUSSION.

Because of the low viral load in chronically infected HBeAg-ve patients, it has been difficult to amplify and sequence the HBV genome in this population. Therefore, little is known about the virological basis of their specific clinical characteristics. To investigate the viral QS during the

HBeAg-ve chronic infection, samples from 16 CI patients were analyzed by NGS, and the results obtained were compared with those of chronic hepatitis patients in different clinical stages.

Study of the QS composition at the functional level showed that the viral population was more complex in CI than in chronic hepatitis patients. In agreement with these findings, high complexity and spontaneous reverse transcriptase mutations (in the A-B interdomain, overlapping with the HBsAg “a-determinant”) have been reported in the viral QS related to the immune-tolerant and immune-active states[17]. Furthermore, an increase in haplotype number in a region including the 5'-end of HBX was detected in a HBeAg-positive woman (previously defined as immune-tolerant) who seroconverted to HBeAg-negative (CI state)[18]. Our results suggest that the accumulation of nt mutations in the QS of CI patients could influence HBV replication, thus providing a possible explanation for the low viral replication rate in this group.

Clinical groups such as CHB and HCC patients are generally found to have highly replicative and/or carcinogenic variants, but curiously, the QS composition was less complex in these patients, suggesting that some variants had been selected to guarantee HBV replication and persistence. The QS of LC patients was found to be the most complex, but very few of these patients were included in the study, and these results may be less reliable.

The portion of the *HBX* gene examined here comprised the 5' terminal end

and the upstream non-coding region, where some hyper-conserved regions have been reported by our group[11]. The current analysis of QS conservation in different, well-characterized, clinical groups and across genotypes confirms these previous results and supports the idea that these regions could be useful targets for gene therapies, which could be effective regardless of the clinical and virological conditions.

In light of the higher mutation rate observed in CI patients, and taking into account that the nucleotide and amino acid conservation findings were similar but not identical between the different clinical groups, we examined the intergroup conservation variability. Surprisingly, sequences from CI patients were the most highly conserved at both the nt and aa level, particularly relative to the CHB group. The difference between CI and CHB conservation was higher in the upstream non-coding region of *HBX* (between nt 1250 and nt 1350 where several sites of *HBX* transcript initiation have been reported,[19]) and in the protein dimerization site (aa 20-50,[20]). The lower conservation of CHB haplotypes may suggest that nt and aa variability help HBV to re-adapt to the external environment and guarantee replication. Conversely, in the CI QS, higher conservation may indicate selection of more highly conserved and probably less replicative haplotypes, which could presumably promote viral persistence. These complexity and conservation results seem to be indicative of a more complicated HBV QS in CI patients. The high functional complexity would be due to the presence of a large number of mutated haplotypes at

a low frequency, which do not, however, affect the nt and aa conservation of the main population.

To explain the limited viral replication in CI patients, we inspected the presence of aa mutations. Determination of aa mutations in these patients could be helpful to better classify those falling into the “grey zone” (viremia between 2,000 and 20,000 IU/mL, and/or marginally elevated ALT) whose management is still difficult due to the lack of factors that distinguish this intermediate state from chronic HBeAg-ve infection or hepatitis[21]. *HBX* and its encoded protein have a key role in HBV replication and disease progression, and may be determinant for the low replication activity seen in CI. A recent study reported a high *HBX* mutation rate in CI patients compared to that of “active” chronic hepatitis patients[22]. Nonetheless, the study investigated only dominant mutations determined by Sanger sequencing (at a frequency of 15%-20%)[23]. Deletions in the 3'-terminal end of *HBX* were detected by NGS in the above-mentioned woman experiencing HBeAg seroconversion[18]. In the present study, the presence of mutations was analyzed by aligning haplotype sequences obtained by NGS with their corresponding genotype consensus. As genotyping was performed considering haplotype sequences, we were able to detect subtle mixtures of genotypes, usually not identified by Sanger sequencing. Notably, a genotype-specific viral evolution was observed. Effectively, the HBV QS may evolve differently depending on the genotype, and thereby, differentially influence disease

progression and therapy outcome[24, 25]. Specific mutations in the HBsAg C-terminal domain of genotype D HBV have been associated with viremia <2000 IU/mL[26] and certain HBx mutations highly associated with HCC have been reported specifically for genotypes C and D[27, 28]. Here, we detected a pattern of aa changes (A12S/P33S/P46S/T36G-D) that was highly represented in CI genotype D haplotypes. These observations emphasize the importance of using NGS to accurately identify the viral genotype, and suggest that it may also be useful for following up the disease.

The mutation pattern identified partially involves the HBx Ser/Pro-rich dimerization site[20]. The pattern was mainly characterized by replacement of a hydrophobic aa (alanine or proline) by a polar aa (serine). This could be relevant, as some highly conserved polar aa (eg, Ser25 and Ser41) within the Ser/Pro-rich domain are targets of post-translational changes[29]. Potential addition of new phosphorylation and O- β -glycosylation sites could interfere with the tridimensional structure of the HBx protein and thereby limit its trans-activating activity.

In vitro investigation of HBV expression in the presence of the A12S/P33S/P46S/T36G mutation pattern showed an approximately 1 log reduction in HBV expression. This relatively small decrease suggests that other accessory mutations may further contribute to controlling HBV replication in CI. A more marked inhibition was detected in relation to the A12S/P33S/P46S/T36D quadruple mutant, which could indicate that the

specific threonine to aspartate change may further alter the transactivating activity of HBx.

The results of this study provide insight into the composition of the HBV QS in CI. However, further work is required to characterize the mechanisms responsible for the complex viral population found and investigate the role of the highly mutated haplotypes in viral replication. A genotype-specific pattern of mutations was detected in the CI QS. Genotype D was the most prevalent genotype in CI patients and this may have contributed to the significant presence of the mutation pattern in this group. However, our *in vitro* experiments showed that the pattern was related to a reduction in viral replication, a typical situation in CI patients. Studies in larger samples with various genotypes are needed to confirm the association between mutations and the CI state and to detect other genotype-specific aa changes that may enable differentiation between the clinical groups. Additional *in vitro* and *in silico* studies are required to further investigate whether and how these mutations interfere with the activity and tridimensional structure of HBx.

In summary, the HBV viral population at the 5' end of HBX was investigated by NGS analysis in a group of HBeAg-ve chronically infected patients. The sophisticated (both conserved and complex) QS characterized may explain the limited viral replication in this patient population. The finding of hyper-conserved regions common to all the clinical groups analyzed warrants further study to determine their value

as gene therapy targets. The presence of aa mutations specific to a determined genotype underscores the need to accurately genotype the HBV virus during follow-up. The pattern of mutations observed in this study could help to better classify chronically infected HBeAg-ve patients and the low viral replication rate.

FOOTNOTES.

Presented at: This study was partially presented as poster presentation at the Liver Meeting 2018 of the American Association for the Study of Liver Disease in San Francisco, California, in November 2018, and at the International Liver Congress 2019 for the European for the European Association for the Study of the Liver 2019, in Vienna, Austria in April 2019.

Contributors: Rodríguez-Frías F designed the research; Cortese MF and González C equally contributed to design the experiments; Cortese MF coordinated the research; González C, Sopena S, Godoy C and Rando A performed the molecular experiments; Cortese MF and Casillas R performed *in vitro* experiments; Cortese MF, Gregori J and Riveiro-Barciela M analyzed data acquired during the experiments and interpreted the results; González C, Cortese MF drafted the manuscript; Yll M, Lopez-Martinez R, Godoy C, Sopena S, Buti M, Quer J, Esteban R and Rodríguez-Frías F critically reviewed the manuscript.

Acknowledging contributors: The authors thank Celine Cavallo for English language support and helpful editing suggestions.

Funding: This study was supported by grants from the Institute of Health Carlos III (grant PI15/00856) and co-financed by the European Regional Development Fund (ERDF).

Competing interests: Josep Gregori is an employee of Roche Diagnostics, SL.

Patient consent: Obtained.

Ethics approval: Ethics Committee of Vall d'Hebron Research Institute.

References.

- 1 Okamoto H, Tsuda F, Sakugawa H, *et al.* Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol* 1988;**69** (Pt 10):2575–83. doi:10.1099/0022-1317-69-10-2575
- 2 Lin Y-Y, Liu C, Chien W-H, *et al.* New Insights into the Evolutionary Rate of Hepatitis B Virus at Different Biological Scales. *J Virol* 2015;**89**:3512–22. doi:10.1128/JVI.03131-14
- 3 Belloni L, Pollicino T, De Nicola F, *et al.* Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc Natl Acad Sci* 2009;**106**:19975–9. doi:10.1073/pnas.0908365106
- 4 Lucifora J, Arzberger S, Durantel D, *et al.* Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. *J Hepatol* 2011;**55**:996–1003. doi:10.1016/j.jhep.2011.02.015
- 5 Pollicino T, Belloni L, Raffa G, *et al.* Hepatitis B virus replication is regulated by the acetylation status of hepatitis B virus cccDNA-bound H3 and H4 histones. *Gastroenterology* 2006;**130**:823–37. doi:10.1053/j.gastro.2006.01.001
- 6 Quasdorff M, Protzer U. Control of hepatitis B virus at the level of transcription. *J Viral Hepat* 2010;**17**:527–36. doi:10.1111/j.1365-2893.2010.01315.x
- 7 Ringelhan M, O'Connor T, Protzer U, *et al.* The direct and indirect roles of HBV in liver cancer: prospective markers for HCC screening and potential therapeutic targets. *J Pathol* 2015;**235**:355–67. doi:10.1002/path.4434
- 8 European Association for the Study of the Liver. 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
- 9 Taida T, Arai M, Kanda T, *et al.* The prognosis of hepatitis B inactive carriers in Japan: a multicenter prospective study. *J Gastroenterol* 2017;**52**:113–22. doi:10.1007/s00535-016-1229-6
- 10 Slagle BL, Bouchard MJ. Role of HBx in hepatitis B virus persistence and its therapeutic implications. *Curr Opin Virol* 2018;**30**:32–8. doi:10.1016/j.coviro.2018.01.007
- 11 González C, Tabernero D, Cortese MF, *et al.* Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy.

- World J Gastroenterol* 2018;**24**:2095–107. doi:10.3748/wjg.v24.i19.2095
- 12 Soria ME, Gregori J, Chen Q, *et al.* Pipeline for specific subtype amplification and drug resistance detection in hepatitis C virus. *BMC Infect Dis* 2018;**18**:446. doi:10.1186/s12879-018-3356-6
- 13 Gregori J, Perales C, Rodriguez-Frias F, *et al.* Viral quasispecies complexity measures. *Virology* 2016;**493**:227–37. doi:10.1016/j.virol.2016.03.017
- 14 Nikolaou N, Green CJ, Gunn PJ, *et al.* Optimizing human hepatocyte models for metabolic phenotype and function: effects of treatment with dimethyl sulfoxide (DMSO). 2016;**4**:1–11. doi:10.14814/phy2.12944
- 15 Cells H, Hepatocytes PH, Ni Y, *et al.* Hepatitis B Virus Infection of HepaRG Cells, HepaRG-hNTCP Cells, and Primary Human Hepatocytes. In: Press H, ed. *Hepatitis B Virus. Methods in Molecular Biology*. New York, NY: 2017. 15–25. doi:10.1007/978-1-4939-6700-1
- 16 S. Choi, B. Sainz, Jr., P. Corcoran, S. Uprichard and HJ. Characterization of increased drug metabolism activity in dimethyl sulfoxide (DMSO)-treated Huh7 hepatoma cells. 2010;**39**:205–17. doi:10.1080/00498250802613620
- 17 Li H, Song X-F, Hu T-T, *et al.* A strong conservative tendency in HBV transcriptase (RT): a majority of natural RT mutations derived from the S gene. *Liver Int* 2016;**36**:963–70. doi:10.1111/liv.13051
- 18 Jones LR, Sede M, Manrique JM, *et al.* Virus evolution during chronic hepatitis B virus infection as revealed by ultradeep sequencing data. 2019;**4**:35–44. doi:10.1099/jgv.0.000344
- 19 Treinin M, Laub O. Identification of a promoter element located upstream from the hepatitis B virus X gene. *Mol Cell Biol* 1987;**7**:545–8. <http://www.ncbi.nlm.nih.gov/pubmed/3494193>
- 20 Datta S, Chatterjee S, Veer V. Recent advances in molecular diagnostics of hepatitis B virus. 2014;**20**:14615–25. doi:10.3748/wjg.v20.i40.14615
- 21 Papatheodoridis G V. Hepatitis B virus treatment: Which patients can have treatment deferred? *Clin Liver Dis* Published Online First: 2013. doi:10.1002/cld.160
- 22 Al-Qahtani AA, Al-Anazi MR, Nazir N, *et al.* Hepatitis B virus (HBV) X gene mutations and their association with liver disease progression in HBV-infected patients. *Oncotarget* 2017;**8**:105115–25. doi:10.18632/oncotarget.22428

- 23 Tsiatis AC, Norris-kirby A, Rich RG, *et al.* Comparison of Sanger Sequencing , Pyrosequencing , and Melting Curve Analysis for the Detection of KRAS Mutations Diagnostic and Clinical Implications. *J Mol Diagnostics* 2010;**12**:425–32. doi:10.2353/jmoldx.2010.090188
- 24 Li S, Wang Z, Li Y, *et al.* Adaptive evolution of proteins in hepatitis B virus during divergence of genotypes. *Sci Rep* 2017;**7**:1–11. doi:10.1038/s41598-017-02012-8
- 25 Rajoriya N, Combet C, Zoulim F, *et al.* Review How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B . Time for an individualised approach ? *J Hepatol* 2017;**67**:1281–97. doi:10.1016/j.jhep.2017.07.011
- 26 Mirabelli C, Surdo M, Van Hemert F, *et al.* Specific mutations in the C-terminus domain of HBV surface antigen significantly correlate with low level of serum HBV-DNA in patients with chronic HBV infection. *J Infect* 2015;**70**:288–98. doi:10.1016/j.jinf.2014.10.015
- 27 Salpini R, Surdo M, Cortese MF, *et al.* The novel HBx mutation F30V correlates with hepatocellular carcinoma in vivo, reduces hepatitis B virus replicative efficiency and enhances anti-apoptotic activity of HBx N terminus in vitro. *Clin Microbiol Infect* Published Online First: 23 November 2018. doi:10.1016/j.cmi.2018.11.017
- 28 Ali A, Abdel-Hafiz H, Suhail M, *et al.* Hepatitis B virus, HBx mutants and their role in hepatocellular carcinoma. *World J Gastroenterol* 2014;**20**:10238–48. doi:10.3748/wjg.v20.i30.10238
- 29 Hernández S, Venegas M, Brahm J, *et al.* The viral transactivator HBx protein exhibits a high potential for regulation via phosphorylation through an evolutionarily conserved mechanism. *Infect Agent Cancer* 2012;**7**:27. doi:10.1186/1750-9378-7-27

6. Material suplementario

Supplementary Material.

Supplementary Table 1.

NCBI GenBank accession numbers of the 102 reference sequences used in HBV genotyping. The consensus reference sequence used for the analysis of the aa mutations was obtained by aligning these sequences per each genotype.

HBV genotype (N sequences)	GenBank accession numbers
A (13)	AY233278, AB241115, GQ477501, AF090841, AF090839, AB194952, AB194951, AM180623, AY934764, FJ692609, FJ692613, GQ331047, GQ331048
B (20)	AB073858, AB362933, D00329, AY596111, AP011084, GQ924653, M54923, AP011085, AB073835, AB115551, AB219427, AP011086, AB287316, DQ463787, EF473977, AP011091, AP011093, AP011094, GQ358148, GQ358152
C (23)	AB112066, X52939, AF533983, AB033553, X75656, X75665, AB048705, AB048704, AP011099, AB241109, AP011102, AP011103, EU670263, AP011107, AP011104, AP011108, AB540583, AB554019, AB554025, AB644281, AB644284, AB644286, AB644287
D (17)	AB555496, GU456647, AB104712, AF280817, AB210820, X97848, Z35716, EU939680, AY233291, AJ132335, GQ922005, AB048701, DQ315779, AB033558, AB493846, FJ904405, AM494716
E (8)	X75664, X75657, FJ349237, AM494694, HM363569, FJ349226, DQ060828, JQ000008
F (10)	AY090459, HQ378247, HE981184, AY090455, AY311369, X69798, X75663, AB036911, AB166850, DQ823090
G (6)	EF464098, HE981172, HE981176, AP007264, GU565217, AB064311
H (5)	AP007261, AB275308, AY090460, AY090454, AB516393

Supplementary Table 2.

Definition of each studied complexity index.

Indices	Definition	Formula	
Shannon Entropy Index (H_{SH}) [†]	The grade of uncertainty in predict sequences identity when taken on random. Quasispecies population with more haplotypes at low frequencies will present higher value of H _{SH}	$-\sum_{i=1}^H p_i \log(p_i) / \log(H)$	p_i = frequency of the haplotype i in quasispecies population H = number of Haplotypes in quasispecies population
Gini Simpson Index (G) [†]	The probability that two genomes taken at random belong to a different haplotype. In quasispecies with higher number of haplotypes at lower frequency the HGS will be presented bigger value of H _{GS} .	$1 - \sum_{i=1}^H p_i^2$	p_i = frequency of the haplotype i in quasispecies population
Mutation Frequency (MF) [¶]	The fraction of mutated residues in the multiple alignment with respect to the dominant haplotype.	$\frac{1}{H} \sum_{i=0}^H d_{1i}$	d = proportion of mutations related to the master sequence
Nucleotide Diversity (π) [¶]	average number of substitutions among pairs of haplotypes in the multiple alignment	$\sum_{i=1}^H \sum_{j=1}^H p_i d_{ij} p_j$	p_i = frequency of the haplotype i in the viral quasispecies p_j = frequency of the haplotype j in the viral quasispecies d_{ij} = Hamming distance (number of mutations that differences the haplotype i and j).

[†] Abundance Indices = take into account the haplotypes presented in the quasispecies and their abundance (in term of relative frequency)
[¶] Functional Indices = based on the genetic differences between haplotypes (known as genetic distance)

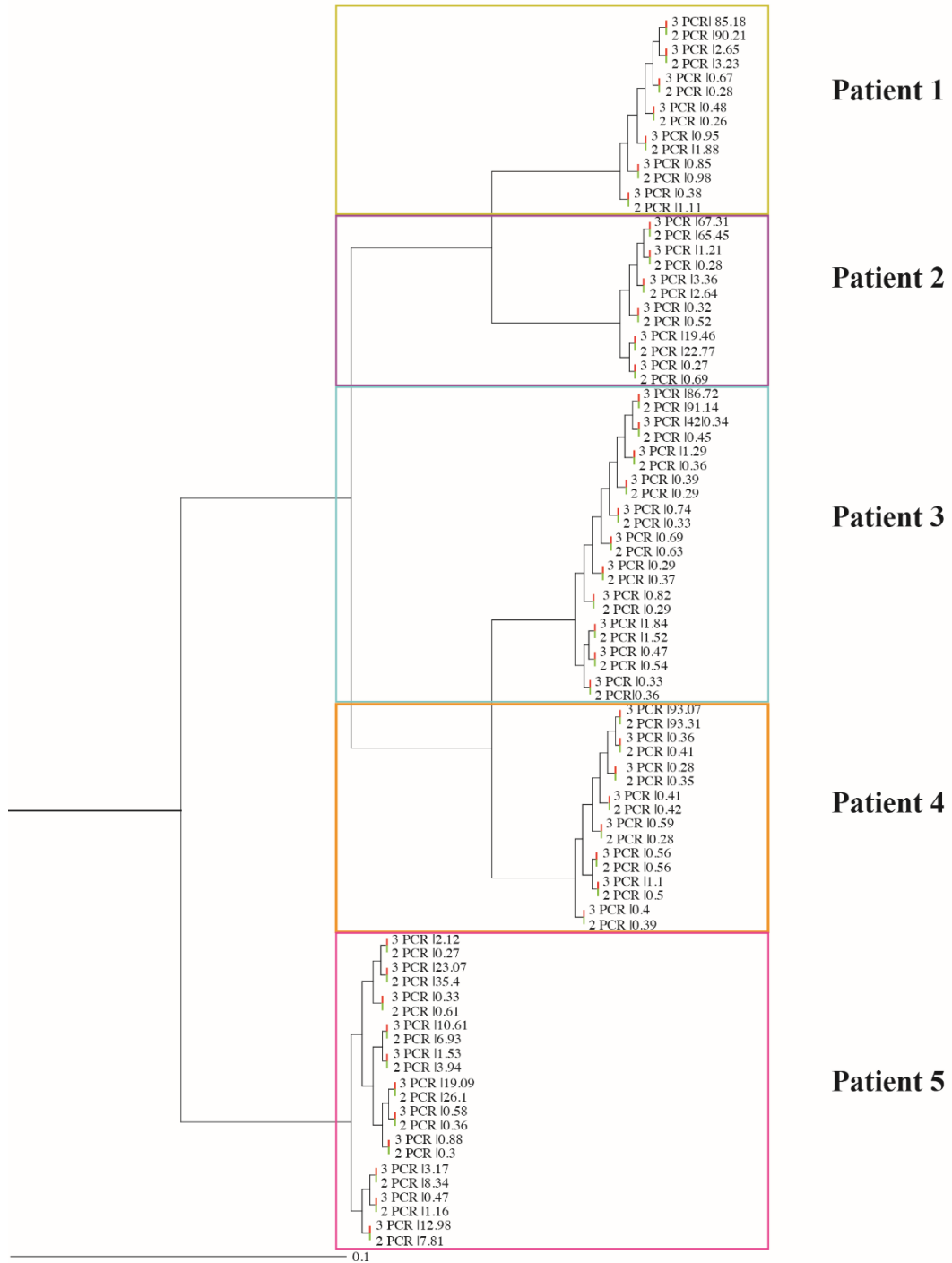
Supplementary Table 3.

Biosample accession numbers for each sample analysed.

Patient	Clinical stage	Biosample accession number
1	IC	SAMN10077196
2	IC	SAMN10077197
3	IC	SAMN10077198
4	IC	SAMN10077199
5	IC	SAMN10077200
6	IC	SAMN10077201
7	IC	SAMN10077202
8	IC	SAMN10077203
9	IC	SAMN10077204
10	IC	SAMN10077205
11	IC	SAMN10077206
12	IC	SAMN10077207
13	IC	SAMN10077208
14	IC	SAMN10077209
15	IC	SAMN10077210
16	IC	SAMN10077211
17	HCC	SAMN10077212
18	HCC	SAMN10077213
19	HCC	SAMN10077214
20	HCC	SAMN10077215
21	HCC	SAMN10077216
22	HCC	SAMN10077217
23	HCC	SAMN10077218
24	HCC	SAMN10077219
25	HCC	SAMN10077220
26	HCC	SAMN10077221
27	HCC	SAMN10077222
28	HCC	SAMN10077223
29	HCC	SAMN10077224
30	HCC	SAMN10077225
31	HCC	SAMN10077226
32	HCC	SAMN10077227
33	LC	SAMN10077228
34	LC	SAMN10077229
35	LC	SAMN10077230

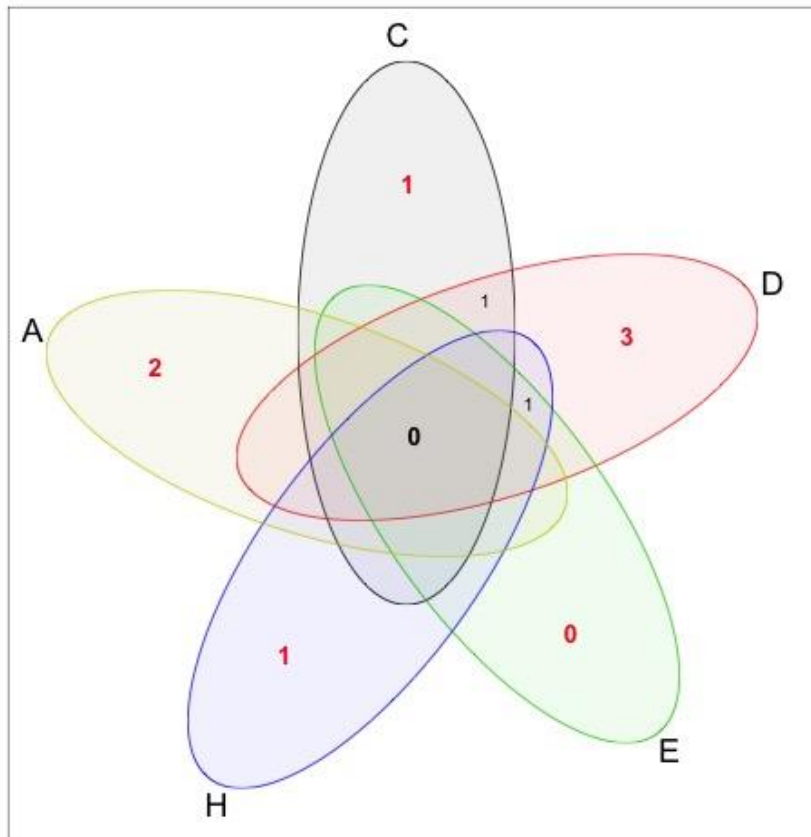
Patient	Clinical stage	Biosample accession number
36	LC	SAMN10077231
37	LC	SAMN10077232
38	LC	SAMN10077233
39	CHB	SAMN10077234
40	CHB	SAMN10077235
41	CHB	SAMN10077236
42	CHB	SAMN10077237
43	CHB	SAMN10077238
44	CHB	SAMN10077239
45	CHB	SAMN10077240
46	CHB	SAMN10077241
47	CHB	SAMN10077242
48	CHB	SAMN10077243
49	CHB	SAMN10077244
50	CHB	SAMN10077245
51	CHB	SAMN10077246
52	CHB	SAMN10077247
53	CHB	SAMN10077248

Supplementary Figure 1.



Comparison of quasispecies population between 3 and 2-round PCR. In order to ensure that 3-round PCR protocol did not affect quasispecies population, we compared the obtained QS haplotypes (in red) with the haplotypes from 2-round PCR (in green) in 5 of the patients included in the study. The tree shows how the haplotypes obtained with 3-round PCR (in red) clustered together with the sequences of 2-round PCR (in green) without observing relevant difference in term of relative frequency. Each patient is highlighted with a different color box.

Supplementary Figure 2.



Genotype-specific mutations. The Venn Diagram shows the number of mutations observed alone or in common in the different viral genotypes. Each genotype is represented by an ellipse at different color (respectively red for genotype D, green for genotype E, blue for genotype H, yellow for genotype A, and grey for genotype C). The number of the genotype-specific mutations is pointed out in red. The number of shared mutations is also reported (in black into the intersection space between the genotype ellipse). Genotype F was no represented because no mutations were detected. In the group of the genotype D-specific mutations was also included the pattern A12S/ P33S/P46S/T36G-D.

CAPÍTULO 5: DISCUSIÓN

Discusión

La infección por el virus de la hepatitis B (VHB) es una de las infecciones más frecuentes en la población mundial, que infecta de forma crónica a más de 257 millones de personas en el mundo, con una tasa de muerte del 15-25% debido a sus complicaciones, fundamentalmente cirrosis y carcinomoma hepatocelular (HCC)³⁷⁴.

Los programas de prevención, la vacuna anti-VHB y los tratamientos disponibles; análogos de nucleos(t)idos (AN) e interferón α (IFN α), basados en la inhibición de la replicación viral, han conseguido disminuir el número de pacientes infectados y frenar la progresión de la enfermedad. Sin embargo, aunque el tratamiento con AN (el más utilizado hoy en día contra la hepatitis crónica B) permite controlar la replicación viral, un pequeño pero relevante porcentaje de pacientes, eficazmente tratados, acaba desarrollando HCC³¹⁵. Esto podría deberse al hecho de que los inhibidores de la RT inhiben la producción de nuevas partículas virales, pero no afectan a la producción intrahepática de proteínas o antígenos virales, algunas de las cuales poseen actividad oncogénica, como es el caso de la proteína X (HBx)⁸². La razón es que los AN no afectan directamente al reservorio natural de la infección, el ADN circular covalentemente cerrado (ADNccc), ni al ADN integrado en el genoma del huésped (ver en el Capítulo 1, los apartados *1.6.3 Formación del ADNccc* y *1.6.4 Transcripción Viral, Formación de Partículas Core*). Esto confirma la necesidad de encontrar una nueva estrategia de tratamiento que permita controlar, no solamente la replicación del virus, sino también la expresión de los antígenos virales.

Este trabajo de tesis doctoral se focaliza en estudiar el gen X del VHB (*HBX*), el cual codifica la HBx, que juega un papel clave en la iniciación de la expresión del VHB⁸² e interactúa con varias vías de señalización celular y regula la expresión de múltiples genes del huésped, lo que afecta a muchas actividades celulares^{275,356,357} (ver en el Capítulo 1 el apartado *1.4.1.4 ORF X y la proteína X*). Dado el papel esencial de esta proteína, se ha considerado que el estudio de la conservación y diversidad genética en la secuencia nucleotídica del *HBX*, así como en la secuencia de aminoácidos de la HBx, podría conducirnos a un mejor entendimiento del comportamiento del VHB en los diferentes estadios clínicos que se pueden producir durante la infección crónica. Además, la identificación, durante la evolución natural de la infección, de zonas altamente conservadas en cualquier estadio de la infección, e incluso bajo el efecto de tratamiento antiviral, evidenciaría zonas o dominios esenciales para la función de esta proteína, que

podrían ser potenciales dianas para el diseño de una nueva terapia antiviral basada en la terapia génica.

Esta tesis doctoral se compone de dos estudios que intentan abordar los puntos antes mencionados. En el primer estudio el objetivo fue identificar las regiones hiper-conservadas en la secuencia nucleotídica del *HBX* y en la secuencia de aminoácidos de la *HBx*, que podrían ser dianas de una nueva estrategia de tratamiento basada en el silenciamiento génico. El segundo estudio amplía el primero, incluyendo un mayor número de pacientes, y sobre todo focalizando la atención sobre el grupo de infección crónica HBeAg (-) anteriormente conocidos como “portadores inactivos”, con el objetivo de identificar factores virológicos en sus quasiespecies (QS, conjunto de genomas virales con un origen común, relacionados entre sí aunque no idénticos) que puedan ser característicos de ellos ayudando así a diferenciarlos de los otros grupos de pacientes con hepatitis crónica. La región objeto de estudio incluye el extremo 5’ del *HBX* y su región no-codificante *upstream* (nucleótidos 1255 a 1611), donde residen varios sitios de iniciación de la transcripción viral¹⁵¹. Esta región se ha escogido porque al principio de esta tesis estaba poco estudiada, mientras que ya se había descrito una presencia significativa de deleciones y / o inserciones (indels) en el extremo 3’ de *HBX*²⁸⁶⁻²⁸⁸ (ver en Capítulo 1 el apartado 2.4.4 *Variabilidad en el ORF X*), lo que sugiere que será más difícil encontrar zonas hiper-conservadas en la región 3’. Es importante destacar que debido a la localización estratégica co-terminal del *HBX*, hace que su secuencia esté presente en todos los transcritos virales, por lo que la región estudiada podría ser una valiosa diana en la terapia génica, permitiendo así la interferencia en la expresión de todos los transcritos virales. Además, la presencia de una QS que muestra diferencias en la complejidad y conservación en esta región, en los diferentes estadios clínicos, podría ser determinante a la hora de definir su capacidad de la replicación viral, conduciendo, por lo tanto, a una diferente evolución clínica.

1. Búsqueda de las regiones hiper-conservadas en la región 5’ del

HBX

Como he argumentado anteriormente, la incapacidad de garantizar una cura definitiva de la infección por el VHB con los tratamientos antivirales actualmente disponibles, empuja a la comunidad científica a buscar nuevas estrategias terapéuticas^{82,175}. Estas nuevas estrategias terapéuticas tienen dos enfoques principales: el primero consiste en potenciar la respuesta inmunitaria del huésped y el segundo está dirigido hacia el desarrollo de terapias génicas dirigidas contra dianas específicas del ciclo replicativo del virus. Las

guías más recientes del tratamiento clínico para la hepatitis B crónica apuntan a que la terapia más prometedora para conseguir una “cura funcional” de la infección, es la combinación de estas dos estrategias³⁰⁵. En la actualidad la cura funcional se considera el punto final del tratamiento, ya que no solo implica la supresión de la replicación viral, sino también la pérdida del HBsAg y silenciamiento transcripcional del ADNccc y el ADN del VHB integrado (ver en el Capítulo 1 el apartado 4.1 *¿Que se entiende por curación de la infección del VHB?*).

Las nuevas tecnologías de la terapia génica pueden permitir la destrucción del ADNccc (ZFNs, TALENs, CRISPR-Cas9...) o su silenciamiento transcripcional (reguladores epigenéticos, siRNA...), por lo tanto, brindan nuevas oportunidades interesantes para el futuro desarrollo de posibles terapias curativas. Entre todos estos tipos de terapia génica dirigida, los siRNAs, por su elevada especificidad, podrían constituir una estrategia altamente prometedora. Estas moléculas son pequeños oligonucleótidos de ARN que silencian genes de manera específica a nivel postranscripcional, a través de una interacción de su secuencia con el ARNm diana, dando como resultado la inhibición de la expresión de la proteína viral³³⁸ (ver en el Capítulo 1 el apartado 4.4.1.4 *ARN pequeños de interferencia [small interference RNA o siRNA]*). El hecho de que la inhibición de la expresión de proteínas virales sea a nivel postranscripcional, sin interaccionar con el genoma viral, da una posible ventaja al siRNA respecto a las terapias que permiten la destrucción del ADNccc, ya que estas son potenciales candidatas para dañar los genes humanos que estén cerca del sitio donde se haya integrado el genoma viral. Un siRNA producido por Arbutus Biopharma Corporation está actualmente en fase II de experimentación clínica en el tratamiento de pacientes con hepatitis crónica. Los resultados preliminares indicaron que la terapia génica fue bien tolerada y condujo a una reducción significativa en los niveles de HBsAg^{340,359}.

Las terapias génicas se basan en el reconocimiento específico de una secuencia diana por un pequeño oligonucleótido o un complejo ARN-proteína. Así pues, la variación de su secuencia diana puede afectar a su efectividad, lo que justifica la búsqueda de regiones hiper-conservadas en el genoma del VHB, cuya secuencia no pueda variar sin afectar a la replicación viral. Con el fin de garantizar la detección de posibles dianas útiles en cualquier estadio clínico y en presencia de cualquiera de los genotipos virales, en el primer estudio de esta tesis se ha seleccionado un grupo heterogéneo de 27 pacientes en diferentes fases de la de la infección crónica por VHB (previamente bien caracterizadas) y con diferentes genotipos virales. En una muestra de plasma de cada uno de ellos se realizó el análisis de la conservación de la región 5' del *HBX* mediante el contenido de información en un alineamiento múltiple de las secuencias obtenidas por secuenciación

masiva (ver en el Capítulo 1 el apartado 2.1.3 *Estudio de la conservación de la quasispecies*), lo que permitió identificar dos regiones híper-conservadas.

La primera región se localizó entre los nucleótidos 1255-1286 (en la región no codificante de los transcritos del *HBX*). Es necesario destacar que los transcritos del gen *HBX* pueden iniciarse en múltiples sitios¹⁵¹, lo que significa que esta región podría no estar completamente presente o incluso no estar presente en todas las formas de los transcritos de este gen, pero sí que se encontrará en todos los otros transcritos virales, gracias a su localización en la parte co-terminal de todos ellos. La segunda región híper-conservada localizada se sitúa entre los nucleótidos 1519 y 1603 dentro de la región codificante. Una región híper-conservada en este intervalo de nucleótidos había sido descrita en un estudio realizado por Karinova M et al.³³³, donde demostraron que el diseño de CRISPR / Cas9 dirigido a esta región en el *HBX* de un genoma de un genotipo A podría modificar tanto el ADNccc episomal como el ADN del VHB integrado, interfiriendo con la replicación del VHB y con la infección de “*novo*” de las líneas celulares de hepatoma. En un estudio posterior, Thongthae et al.³⁷⁵ observaron, mediante un estudio de predicción informática, algunas posibles dianas para el uso de siRNA en una sola secuencia viral del *HBX* (incluida la región no codificante), describiendo la eficacia de este sistema en un estudio *in vitro*. El valor añadido del presente estudio sobre los previamente publicados es que la conservación de las regiones examinadas se estudió directamente por secuenciación masiva de muestras de pacientes, teniendo en cuenta los diferentes genotipos del VHB y los diferentes estadios clínicos de la infección. Esto permitió obtener una muestra de miles de secuencias por cada muestra, con lo cual se garantiza disponer de una representación significativa de las diferentes poblaciones virales (haplotipos) que formaban cada QS.

Al traducir las secuencias de los haplotipos obtenidos a aminoácidos, se detectó una región conservada entre los aa 63 y 76 (codificada por nts 1563-1602, dentro de la segunda región nucleotídica híper-conservada), confirmando los resultados de conservación a nivel nucleotídico. Es de destacar que este fragmento incluye algunos aa de un dominio tipo Kunitz (del aa 58 al 70)⁹⁰. Este tipo de dominios pueden inhibir la función de las enzimas encargadas de la degradación celular, como son las proteasas³⁷⁶. Eso nos sugiere que esta porción de proteína podría conservarse para preservar la integridad de la proteína HBx, protegiéndola de una degradación no deseada. Esto indica su esencialidad para el ciclo replicativo del VHB y justifica el elevado grado de conservación de la secuencia a nivel de nt y aa.

Vale la pena señalar que entre el grupo inicial de 46 pacientes con infección crónica por el VHB bien caracterizados disponibles para el desarrollo de nuestro primer estudio, tuvimos que seleccionar a aquellos pacientes que presentaron niveles de viremia > 2000 UI/mL, lo cual vino marcado por el límite de detección inferior de nuestra técnica de PCR. Por este motivo solo fueron seleccionados 27 pacientes en los que se representaron la mayoría de las etapas clínicas de la infección por el VHB, y con la mayoría de los genotipos del VHB (A a F y H). Aunque el número de pacientes incluidos en este estudio fue limitado, los resultados fueron confirmados en un grupo más grande en el segundo estudio de esta tesis doctoral, en el que se incluyeron 53 pacientes con infección crónica por el VHB HBeAg (-), la forma más frecuente de hepatitis crónica B en España²⁵¹. Como en el primer estudio estos pacientes fueron previamente bien clasificados en infección o hepatitis crónicas y dentro de estos últimos en función de su progresión. Además, la mayoría de los genotipos del VHB también se hallaban representados. Este aumento del número de pacientes en el segundo estudio fue posible al agregar una PCR previa a las descritas en el primer estudio, con unos *primers* externos a la región objeto del estudio, que permitió aumentar el límite de sensibilidad de la PCR hasta 100 UI / mL, permitiendo así la secuenciación de muestras de pacientes con baja actividad replicativa del VHB. Esta PCR extra no influyó de manera significativa en el análisis por secuenciación masiva de la QS en la región estudiada (como se demuestra en la figura 1 del Capítulo 4 el apartado 6. *Material suplementario*:). Cabe destacar que, en este segundo estudio, además de volverse a identificar las regiones híper-conservadas ya identificadas en el primer estudio, se identificó una nueva región entre los nts 1411 y 1435. Las implicaciones biológicas para el VHB de la conservación de esta región se deberían investigar en futuros estudios. A nivel de aa se observó una región híper-conservada entre los aa 63 y 77, de nuevo confirmando los resultados obtenidos en el primer estudio. Además, es preciso señalar que la tecnología utilizada en este primer estudio (GS-Junior plataforma, 454 / Roche) ha sido descatalogada por el proveedor. No obstante, el protocolo utilizado para este trabajo puede adaptarse a las plataformas actualmente disponibles, como las Illumina MiSeq (San Diego, Estados Unidos), como se ha demostrado en el segundo estudio del presente trabajo. Es de destacar que las regiones híper-conservadas aquí identificadas ya se están utilizando como diana en un estudio *in vitro* de siRNA, Los resultados preliminares han sido presentados en el último congreso de la Asociación Americana por el estudio de las enfermedades hepáticas (AALSD) que tuvo lugar en San Francisco en Noviembre 2018 (poster numero 403) (ver *Anexo I*)³⁶⁶.

2. Estudio de la QS de la región 5' de la proteína HBx: el caso de la Infección crónica por el VHB HBeAg (-).

Desde un punto de vista clínico, la infección por el VHB puede dividirse en diferentes estadios clínicos, en función de sus características virológicas (ADN-VHB), bioquímicas (ALT, AST...) y serológicas (HBeAg/anti-HBeAg, HBsAg) y a los cuales se les asocia un protocolo preciso de intervención y tratamiento³⁰⁵. El grupo clínico de los pacientes con infección crónica HBeAg (-) (IC) se caracteriza por una viremia muy baja, ALT normal (< 40 UI/mL) o con elevación leve y un buen pronóstico a largo plazo (ver en el Capítulo 1 el apartado 3.2.2. *Infección crónica*). La identificación de características virológicas específicas en estos pacientes podría proporcionar una valiosa información sobre los mecanismos responsables del control de la replicación viral, determinando así factores que podrían tener una utilidad clínica en el seguimiento (factor pronóstico) y en el tratamiento de los pacientes. Por lo tanto, el grupo de los los pacientes con infección crónica HBeAg (-) es seguramente el de mayor interés. Además, aunque generalmente estos pacientes tengan un pronóstico favorable, el riesgo de desarrollar HCC y/o reactivación de la hepatitis después de la terapia inmunosupresora persiste, y se recomienda realizar un seguimiento periódico continuo^{377,378}.

En este trabajo de tesis doctoral se ha analizado la QS viral de estos pacientes en la región 5' del *HBX* entre los nt 1255 y 1611 con el fin de identificar factores virológicos que puedan explicar la diferente evolución de la enfermedad hepática. Para ello se ha estudiado el grupo de 53 pacientes con infección crónica por el VHB HBeAg (-) del segundo estudio, que incluyó un grupo de 16 pacientes con infección crónica (IC) y se comparó con otro grupo de 37 pacientes con hepatitis crónica en diferentes etapas clínicas: 15 con hepatitis crónica (CHB), 6 con cirrosis (LC) y 16 con HCC. Gracias a la puesta a punto, de una metodología de PCR, fue posible aumentar la sensibilidad de la amplificación de la región de estudio sin que resultara afectada la población viral obtenida con la secuenciación masiva, como se ha comentado en el apartado anterior. Esto permitió incluir un número relevante de pacientes con IC. Es necesario destacar que, debido a la baja viremia que caracteriza a este estadio clínico, hasta la fecha era muy difícil amplificar y secuenciar el genoma del VHB en estos pacientes, por lo que se sabe muy poco sobre las bases virológicas de sus características clínicas específicas.

El estudio de la complejidad de la QS cuantificó la diversidad y la frecuencia de la mezcla dinámica de los haplotipos / poblaciones virales que la forman (ver en el Capítulo 1 el apartado 2.1 *Complejidad de quasiespecies*). El primer resultado a destacar de este análisis ha sido la observación de una mayor complejidad de tipo funcional de la QS de

los IC. Según los resultados obtenidos, los IC y los LC son los pacientes que presentan una mayor complejidad de la QS en términos de los índices funcionales: frecuencia de mutación (Mf) y diversidad nucleotídica (π), en comparación con los CHB y HCC. Hay que considerar que estos índices no consideran la frecuencia relativa de los haplotipos, ya que la Mf mide la diversidad genética con respecto al haplotipo más prevalente, mientras que la π mide la diversidad genética de la población viral como el número promedio de mutaciones por sitio entre cada par de haplotipos en la población viral (ver Capítulo 1 en el apartado 2.1.2 *Índices para el estudio de la diversidad complejidad de la quasispecie*), por lo que los haplotipos con un gran número de mutaciones podrían encontrarse a baja frecuencia.

Los IC mostraron una complejidad, del orden de hasta 5 veces superior, a la de los pacientes con hepatitis crónica con o sin lesión tumoral. Existen pocos estudios que hayan analizado la QS del VHB en este grupo de pacientes. Tan solo hemos encontrado un estudio en el que detectó un aumento en el número de haplotipos en una región que incluía el extremo 3' del gen *HBX* en una madre con infección crónica HBeAg positivo (previamente conocido como inmunotolerante) que seroconvirtió a anti-HBe, pasando a un estado de IC³⁷⁹. De acuerdo con nuestros resultados, la acumulación de cambios de nt en los IC podría influir en la replicación viral observada en este grupo. Hay que destacar que el grupo con más complejidad resultó ser el de los LC, pero, lamentablemente, no pudimos analizar un número significativo de este grupo de pacientes, por lo que no ha sido posible sacar unas conclusiones claras sobre este grupo. Curiosamente los grupos CHB y HCC que generalmente se asocian con la presencia de variantes altamente replicativas y/o carcinogénicas, fueron menos complejos, lo que sugiere que se han seleccionado algunas pocas variantes para garantizar así la replicación y la persistencia del VHB.

El segundo resultado a destacar en este estudio es el análisis de la conservación en los alineamientos múltiples de los haplotipos obtenidos en cada grupo, mediante el contenido de información y su posterior comparación entre ellos. Como se ha comentado en el apartado anterior, dicho análisis confirmó las regiones híper-conservadas descritas en el primer estudio. Además, analizando más en detalle la tendencia de la conservación en la QS viral en los diferentes grupos de pacientes, se evidenció que las QS de los IC eran las más conservadas, sobre todo con respecto al grupo de los CHB. Observando dos regiones donde la diferencia entre la conservación del grupo de los IC y los CHB es más significativa: la primera a nivel de nt, situada en la región no codificante del *HBX* entre nt 1250 y nt 1350, donde se han descrito diferentes sitios de iniciación de transcritos de *HBX* y la segunda nivel aminoacídico la cual recae sobre la región del *HBX* que codifica

el dominio putativo de dimerización de HBx³⁷³ (aa 20-50). Una mayor variabilidad a este nivel podría conferir un estado diferente de configuración tridimensional de la HBx que podría influir en su función de transactivación. La menor conservación de los haplotipos de CHB podría sugerir que la variabilidad en estos niveles podría ayudar a que el VHB se adapte mejor a cambios en el entorno externo, garantizando su replicación. En el caso del grupo de los IC, la mayor conservación de la QS podría indicar una selección de los haplotipos más conservados y probablemente menos replicativos. Esto podría favorecer la persistencia viral, ya que esta replicación a bajo nivel no estimularía la respuesta inmune en la célula ni tampoco produciría un daño hepático inmediato. Sin embargo, esta hipótesis debería ser confirmada en futuros estudios.

Estos resultados (mayor complejidad y al mismo tiempo mayor conservación de la QS de los IC) pueden parecer contradictorios, pero representan una observación adicional de la gran sofisticación de la QS del VHB en los pacientes IC. Nuestros datos sugieren que la quasiespecies de los IC está compuesta por variantes altamente mutadas a baja frecuencia que, aunque no afectan a la conservación de la población principal, aumentan la complejidad de las QS a nivel funcional (Mf y π). La posible interferencia en la replicación del VHB de estos haplotipos menores debe analizarse más a fondo y también se debería analizar la relevancia en la capacidad replicativa del virus en las dos regiones donde la diferencia entre la conservación los pacientes IC y los CHB es más significativa.

3. Estudios de las mutaciones genotipo-específicas.

En la segunda parte del análisis de la QS de los 53 pacientes con infección crónica por el VHB HBeAg (-) del segundo estudio se han analizado las secuencias de los haplotipos de cada paciente respecto a la consensus de su genotipo. Se ha considerado mutación un cambio nucleotídico que conduzca a un cambio a nivel del aa codificado (cambios no silenciosos). El objetivo de este análisis fue identificar mutaciones que puedan discriminar entre los diferentes grupos clínicos.

En una primera fase del análisis se compararon los diferentes grupos clínicos sin tener en cuenta el genotipo de los pacientes. En este análisis no se detectaron mutaciones representadas diferencialmente entre los grupos. Sin embargo, al estudiar la presencia de las mutaciones genotipo-específicas en los distintos grupos clínicos se identificó un patrón de mutaciones (A12S / P33S / P46S / T36G-D) muy frecuente en los haplotipos del genotipo D en el grupo de los IC. Muchos estudios han corroborado la existencia de una relación muy estrecha entre el genotipo viral y el estadio de la enfermedad o la respuesta al tratamiento³⁸⁰. También se ha demostrado que existe una evolución

específica correlacionada con el genotipo del VHB³⁸¹. De hecho ya se han descrito mutaciones específicas de genotipo, en el dominio C-terminal de HBsAg del genotipo D del VHB que han sido asociadas con una viremia <2000 UI / ml³⁸². También se han descrito mutaciones específicas de los genotipos C y D en la región del *HBX* que codifica el extremo C-terminal de HBx que están estrechamente relacionadas con HCC^{286,383}. Estas observaciones corroboran la importancia de genotipar de forma adecuada el VHB, para lo que la NGS podría ser una herramienta muy valiosa, gracias a su capacidad de detectar variantes de la población viral presentes, también a baja frecuencia. Esto permitiría identificar mezclas de genotipos que de lo contrario no se verían con la secuenciación por el método Sanger, ya que tiene un límite de detección de variantes con una frecuencia de 15-20%³⁸⁴. Así pues, el correcto estudio de los diferentes genotipos virales podría ser una valiosa herramienta con el fin de mejorar el seguimiento de la evolución de la enfermedad. En este sentido, cabe destacar que una dificultad en el genotipado del VHB es que no existen patrones de referencia para cada genotipo viral. Para poder solventar este inconveniente y genotipar con fiabilidad los haplotipos obtenidos en los dos estudios de este trabajo de tesis, se adoptó por una estrategia basada en agrupar estos haplotipos por su identidad de secuencia, es decir, según la distancia genética entre ellos. Para ello se estableció un umbral de distancias genéticas máximas entre secuencias del mismo genotipo y mínimas entre secuencias de diferentes genotipos. Finalmente, el genotipo de cada grupo de secuencias se determinó mediante un análisis filogenético con la región entre los nts 1255 y 1611 de 102 secuencias representativas de los genotipos virales A a H extraídas del genoma completo de VHB.

Curiosamente, al analizar este patrón de mutaciones hemos observado que su principal característica viene determinada por la sustitución de un aa hidrófobo (alanina o prolina) por un aa polar (serina). Esto es bastante relevante, considerando que dentro del dominio de dimerización rico en Ser / Pro hay algunos aa polares altamente conservados (como Ser25 y Ser41) que son objetivos de modificaciones post-traduccionales⁸⁹. La adición de un nuevo aa polar, puede considerarse como un posible objetivo de fosforilación y O- β -glicosilación que podría alterar la tridimensionalidad de la HBx y, por consiguiente, limitar su actividad transactivadora. De hecho, el estudio fenotípico *in vitro* de este patrón mutacional, realizado por nuestro propio grupo a raíz de este trabajo (ver *Anexo2*), demostró una reducción en la capacidad de expresión viral de más de un logaritmo en presencia del patrón completo A12S/P33S/P46S/T36D. Queda explorar el rol de estas mutaciones en el plegamiento tridimensional de la proteína X y sus funciones en las modificaciones post-transcripcionales anteriormente citadas.

4. Resumen

En el proyecto de tesis aquí presentado se han desarrollado dos estudios. En el primer estudio se han identificado 2 regiones híper-conservadas en un grupo heterogéneo de pacientes en diferentes fases de la de la infección crónica por el VHB y con diferentes genotipos virales. Esto sugiere que estas regiones podrían ser utilizadas como potenciales dianas terapéuticas para una nueva estrategia basada en el silenciamiento génico que sea efectiva en todos los estadios clínicos y en presencia de todos o la mayoría de los genotipos virales. Aunque todavía el estudio de los siRNA específicos para estas regiones es a un nivel preliminar, los datos aquí mostrados constituyen un buen punto de partida para el diseño de dicha estrategia.

En el segundo estudio los resultados de conservación han sido confirmados en un número más amplio de pacientes con infección crónica por VHB HBeAg (-), clasificados en distintos grupos según la fase de infección o hepatitis crónicas y dentro de estos últimos en función de su progresión. Esto reafirma la validez de estos resultados para la creación de una terapia que sea pangotípica y panclínica. Además, el segundo estudio ha sido focalizado mayoritariamente en el grupo de los IC, evidenciando, en estos pacientes, una QS viral mucho más sofisticada; caracterizada por la presencia de un alto número de haplotipos altamente mutados, que probablemente se encuentren a muy bajas frecuencias en la QS, por lo que no afectan a la conservación. Esto podría sugerir que la mayor sofisticación de la QS viral en los infectados crónicos sea el primer determinante de la baja replicación del VHB en estos pacientes, con la selección de haplotipos más conservados como signo de una menor capacidad de adaptarse al ambiente exterior, llevando por lo tanto a un virus que replica menos.

Asimismo, en este último estudio se ha evidenciado un grupo de mutaciones genotipo-específicas. En concreto se ha descrito un patrón mutacional en los haplotipos de genotipo D de los IC asociado a una probable reducción de la expresión viral. Este patrón mutacional genotipo-específico acentúa aún más la necesidad de genotipar adecuadamente el virus en el seguimiento de los pacientes infectados por el VHB, ya que podría ser un indicador de baja replicación viral.

5. Perspectivas de futuro:

Muchos programas de investigación están en curso para desarrollar nuevos tratamientos que tienen como objetivo principal intentar conseguir una verdadera curación del VHB, sin riesgo de recaída al retirarlos, ni de progresión de la enfermedad hepática. El

silenciamiento génico podría ser una valiosa estrategia para bloquear la expresión de antígenos virales y limitar la evolución de la enfermedad. En este sentido, nuestro grupo está intentando buscar otras regiones híper-conservadas que puedan ser candidatas a nuevas dianas de terapia génica. Asimismo, nuestro grupo también colabora con el grupo del Centro de Investigación Médica Aplicada (CIMA) de terapia génica de enfermedades hepáticas (IP: Gloria González-Aseguinolozza) en el diseño y estudio de siRNAs que usen como diana las regiones híper-conservadas identificadas a lo largo de este proyecto. Además, estas regiones y otras nuevas serán confirmadas directamente en la molécula diana de este tratamiento, es decir en los ARNm virales, mediante el estudio de la QS del VHB a nivel de los transcritos virales en nuevos proyectos de nuestro grupo.

La presencia de mutaciones genotipo-específicas y la coexistencia en la misma población viral de diferentes genotipos resaltan la importancia de clasificar adecuadamente el genotipo del VHB. Con este objetivo, nuestro grupo está ahora llevando a cabo un proyecto que permita el correcto genotipaje del VHB por secuenciación masiva, permitiendo al mismo tiempo identificar mezclas de genotipos diferentes e incluso inferir variantes que sean producto de la recombinación entre dos o más genotipos diferentes. La clasificación genotípica de las muestras procesadas en este nuevo proyecto se basa en la misma estrategia utilizada en este trabajo de tesis (agrupar estos haplotipos por su identidad de secuencia y el análisis filogenético con secuencias representativas de los diferentes genotipos).

La clasificación de los pacientes con infección crónica por VHB HBeAg (-) en infección o hepatitis crónica en ocasiones resulta difícil. Algunos pacientes que pueden tener niveles de ADN del VHB de entre 2.000 UI / ml y 20.000 UI / ml acompañados de una ALT dentro de los rangos de normalidad o ligeramente elevada y sin evidencias de daño hepático³⁷⁰ quedan dentro de la “zonas gris”. La detección de un patrón de mutaciones asociadas a una expresión viral reducida podría ser útil para ayudar a la clasificación de estos pacientes. Este patrón mutacional se podría confirmar en un grupo más grande de pacientes con IC y analizarla en un estudio prospectivo de pacientes en la zona gris relacionándolo con su evolución clínica (o IC o CHB).

El hecho de que la presencia de este patrón pueda afectar a la expresión del virus, abre una nueva puerta en el estudio de la estructura tridimensional de la proteína HBx, y de sus posibles modificaciones post-trasduccionales, y de su importancia en la transactivación de la expresión del virus de la hepatitis B.

CAPÍTULO 6: CONCLUSIONES

1. Conclusiones:

1. Los dos estudios de esta tesis han confirmado la presencia de 2 regiones híper-conservadas en la región 5' del gen *HBX*, la primera, entre los nucleótidos 1255-1286, en la región no codificante de los transcritos del *HBX* y la segunda, entre los nucleótidos 1519 y 1603, en su región codificante, que se conservó tanto a nivel de nt como a nivel de aa. La utilidad de estas regiones como potenciales dianas para la terapia génica antiviral basada en el silenciamiento génico con siRNA está siendo testada. Esta terapia podría ser efectiva en todos los estadios clínicos y en presencia de todos o la mayoría de los genotipos virales.
2. A nivel aminoacídico, se detectó en ambos estudios una región altamente conservada entre los aa 63 y 76, que incluye una porción de un dominio tipo *Kunitz-like*, que podría ser esencial para el ciclo replicativo del VHB. Esta región en la HBx corresponde con una secuencia de nt híper-conservada, incluida dentro de la segunda región híper-conservada (nt 1519-1603) previamente descrita.
3. En el segundo estudio se detectó otra región híper-conservada, en la región codificante del *HBX*, entre los nucleótidos 1411 y 1435, cuya utilidad como potencial diana para la terapia génica y las posibles implicaciones biológicas de su conservación, deberían ser estudiadas en futuros estudios.
4. Los pacientes con infección crónica por el VHB HBeAg (-) en diferentes estadios clínicos siguen patrones de conservación de la región 5' del *HBX* similares. Sin embargo, se observan diferentes tendencias en la conservación de los segmentos entre los nucleótidos 1250 y 1350 y la región aminoacídica que codifica el dominio putativo de dimerización de la HBx (aa 20-50), entre pacientes con infección y hepatitis crónica. La relevancia en la capacidad replicativa del virus de estas dos regiones debería ser estudiada en futuros estudios.
5. Se ha evidenciado una complejidad y conservación mayor de la quasiespecies en los pacientes con infección crónica HBeAg (-) (IC), que en los pacientes con hepatitis crónica HBeAg (-) con carcinoma hepatocelular (HCC) o sin éste (CHB). Esto podría ser debido a la presencia de un grupo de variantes altamente mutadas que se encuentran en

baja frecuencia en la quasiespecies de los IC, por lo que afectan poco a la conservación. Por otro lado, la menor complejidad y conservación de los CHB y HCC podría estar asociada a la selección de algunas pocas variantes altamente replicativas y/o carcinogénicas.

6. La presencia de un patrón de mutaciones en las secuencias de los pacientes con IC de genotipo D (A12S / P33S / P46S / T36G-D) podría relacionarse con la baja replicación viral observada en estos pacientes y sugiriendo una evolución de la población viral de forma genotipo-específica. Esta observación refuerza la necesidad del genotipado adecuado del virus durante el seguimiento del paciente con infección crónica por el VHB. Las implicaciones de este patrón de mutaciones sobre la función transactivadora de la HBx, o su efecto sobre su estructura tridimensional, o las posibles modificaciones post-trasduccionales que pueda ocasionar, deberían ser analizadas en futuros estudios.

ANEXO

1. Anexo 1

Abstract y poster del trabajo presentado en el último congreso de la Asociación Americana por el estudio de las enfermedades hepáticas (AALSD) que tuvo lugar en San Francisco en noviembre 2018 (poster numero 403), del estudio preliminar de la eficacia del siRNA *in vitro*

Antiviral Activity of a New siRNA Targeting a Hyper-Conserved Region in Hepatitis B X Gene Transcripts: Preliminary Data.

Sheila Maestro^{1,2}, **Maria Francesca Cortese**^{3,4}, Carolina González³, David Tabemero³, Mar Riveiro-Barciela⁵, Prof. Maria Buti⁵, Gloria González-Aseguinolaza^{1,2} and Francisco Rodríguez-Frías⁴, (1)Gene Therapy and Regulation of Gene Expression Program, Center for Applied Medical Research (CIMA), (2)Instituto De Investigación Sanitaria De Navarra (IdiSNA), (3)Biochemistry and Microbiology/Liver Pathology Unit, Vall D'hebron University Hospital, (4)Liver Unit, Vall D'hebron Research Institute, (5)Department of Internal Medicine/Liver Unit, Vall D'hebron University Hospital

Abstract Text:

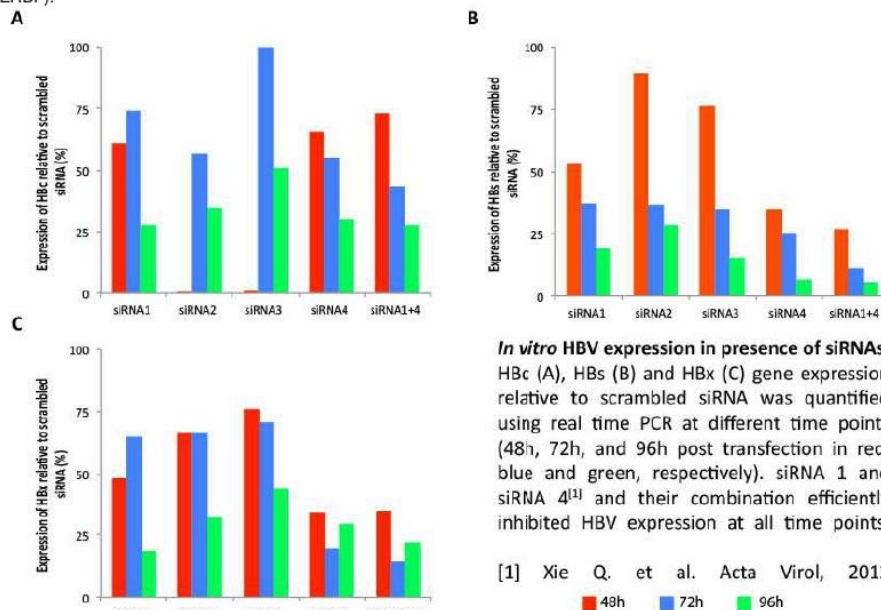
Background: Functional cure of chronic hepatitis B (CHB) by standard nucleos(t)ide analog-based treatment is rarely achieved, with the risk of hepatocellular carcinoma (HCC) remaining present. The hepatitis B X gene (*HBX*) plays a key role in this process. Using next-generation sequencing (NGS) a previous study by our group (González *et al.* World J Gastroenterol 2018;24:2095) identified two hyper-conserved regions in 5' end of *HBX* [nucleotide (nt) 1255-1286; nt 1519-1603] regardless of the patients' clinical stage or HBV genotype. The aim of this study was to determine the antiviral activity of a small interference RNA (siRNA) battery targeting these regions.

Methods: siRNAs were identified using the iScore Designer. In order to evaluate their antiviral activity, HepG2-NTCP cells were infected with purified HBV, and 48 hours post-infection they were transfected with siRNAs (alone or in combination) using lipofectamine. A known siRNA targeting X transcript (siRNA4) and a non-specific scrambled siRNA were used as controls. The expression of hepatitis B X (HBx), core (HBc) and surface (HBs) antigen mRNAs was determined by real time PCR at different time points.

Results: We identified 3 potential siRNAs (1-3) targeting the described hyper-conserved regions. Mock-infected cells were productively infected and expressed HBs, HBc and HBx transcripts. siRNAs efficiently reduced gene expression from 48h to 96h post-transfection (Figure). Of note, siRNA1 and siRNA4 were more potent inhibitors of viral expression (percentage of HBx expression relative to scrambled siRNA: 18.6 and 29.5% for siRNA1 and siRNA4, respectively vs. 32.1 and 44% for siRNA2 and siRNA3 at 96h, Figure panel C). The combination of siRNA1+4 further increased the inhibition of gene expression with a percentage of HBV antigen expression relative to scrambled siRNA of 27.5, 5.6 and 22% for HBc, HBs and HBx, respectively at 96h (Figure).

Conclusion: By targeting a hyper-conserved region of *HBX*, siRNA1 efficiently reduces HBV gene expression *in vitro*. The use of this siRNA alone or in combination with other known siRNAs could help to control HBV infection and prevent disease progression and HCC development. Further *in vitro* and *in vivo* studies are required to confirm these results.

Funding: Instituto de Salud Carlos III (grant PI15/00856) co-financed by the European Regional Development Fund (ERDF).



#403. Antiviral Activity of a New siRNA Targeting a Hyper-Conserved Region in Hepatitis B X Gene Transcripts: Preliminary Data.

Sheila Maestro^{1,2}, María Francesca Cortese^{3,4}, Carolina González³, David Tabernero^{3,5}, Dr. Mar Riveiro Barciela^{5,6}, Dr. María Butús⁷, Gloria González-Aseginolaza^{1,2} and Francisco Rodríguez-Frías^{3,5}

1. Gene Therapy and Regulation of Gene Expression Program, Center for Applied Medical Research (CIMA), 2. Instituto De Investigación Sanitaria De Navarra (IdISNA), 3. Biochemistry and Microbiology/Liver Pathology Unit, Vall D'hebron University Hospital, 4. Liver Unit, Vall D'hebron Research Institute, 5. Cibersalud, Instituto De Salud Carlos III, 6. Department of Internal Medicine/Liver Unit, Vall D'hebron University Hospital, 7. Hospital Universitari Vall d'Hebron, Barcelona, Spain

The author have no conflict of interest to declare

INTRODUCTION

Standard nucleos(t)ide analog-based treatment efficiently suppressed HBV replication, however functional cure is rarely achieved. The hepatitis B X gene (HBX) plays a key role in HBV replication and disease progression. Using Next-Generation Sequencing (NGS) we previously identified two hyperconserved regions in the 5'-end of HBX [nucleotide (nt) 1255-1286; nt 1519-1603]⁽¹⁾ regardless of the patients' clinical stages or HBV genotype.

AIM

The aim of this study was to determine the in vitro antiviral activity of a small interference RNA (siRNA) battery targeting these regions.

METHODS

siRNAs promotes gene silencing by binding the RNA target in a sequence-specific way thus activating the RNA-Induced Silencing Complex (RISC) and promoting the RNA target degradation. In this study siRNAs were identified using the iScore Designer starting from the HBX 5'-end region previously published⁽¹⁾.

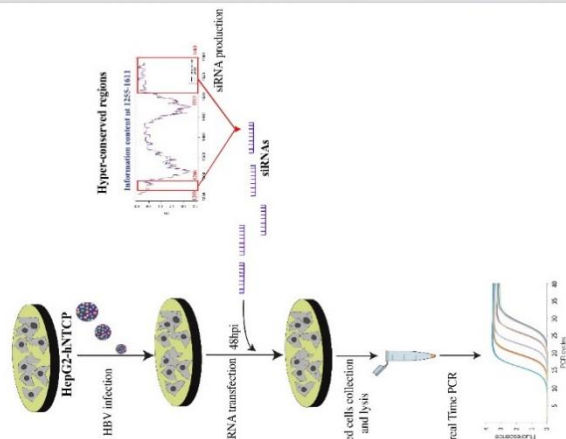
HepG2-hNTCP cells were infected with purified HBV particles (producing from HepG2.2.15 cells, which stably expressed HBV)

48h post infection (pi) infected cells were transfected with siRNAs (alone or in combination) using lipofectamine.

A scramble siRNA was used as negative control. A siRNA specific for HBX (nt 1660-1679, siRNA4) and previously published⁽²⁾ was used as positive control.

Cells were collected at different time points (48, 72 and 96hpi) and lysed with TRIZOL in order to extract viral RNA

HBV RNAs were quantified through RT-real Time PCR using primers hybridizing the different HBV RNA forms (HBcAg, HBsAg and HBxAg mRNAs)



⁽¹⁾Gonzalez C et al. WJG 2018
⁽²⁾Xie Q, et al. Acta Virol. 2012

RESULTS

siRNAs designing

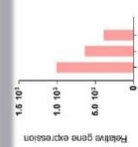
By considering Best iScore in iScore designer, 10 potential siRNAs were identified, however only 3 were specific for the hyper-conserved regions (identified by different colors in figure: siRNA1 in red box, siRNA2 in green box and siRNA 3 in yellow box)

siRNA positions in 5'-end. The hyper-conserved regions are highlighted in bold font

In vitro HBV infection

HBV efficiently infected HepG2-hNTCP cells and productively expressed viral HBcAg, HBsAg and HBxAg mRNAs at 48hpi.

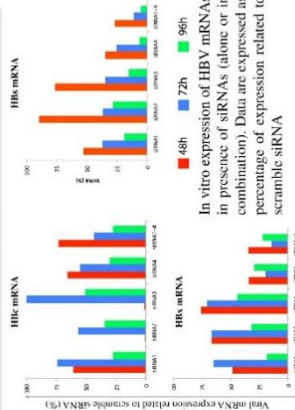
The bar graph highlights the relative expression of HBcAg, HBsAg and HBxAg mRNAs, quantified by RT-Real Time PCR, at 48hpi



siRNA-related HBV mRNAs inhibition

siRNA1 and siRNA4 were more potent inhibitors of viral expression (percentage of HBx mRNA expression related to the scramble siRNA of 18.6 and 29.5% for siRNA 1 and 4 at 96hpi, vs 32.1 and 44% for respectively siRNA2 and 3.

The combination of siRNA1+siRNA4 further increased the inhibition with a percentage of HBV antigens expression relative to scrambled siRNA of 27.5, 5.6 and 22% for HBc, HBs and HBx, respectively at 96h.



CONCLUSIONS

By targeting a hyper-conserved region of HBX, siRNA1 efficiently reduces HBV gene expression in vitro. siRNA1 alone or in combination with other known siRNAs could help the control of HBV infection. Further in vitro and in vivo studies are needed to confirm these results.

ACKNOWLEDGEMENTS

Study funded by Instituto de Salud Carlos III (grant PI15/00856), co-financed by the European Regional Development Fund (ERDF).

CONTACT Dr. María Francesca Cortese maria.cortese@vhiv.org

2. Anexo 2

Estudios *in vitro* sobre el efecto de las mutaciones sobre replicación del VHB

Las mutaciones que presentaron una frecuencia significativamente diferente entre los IC y los demás grupos de pacientes se probaron *in vitro*, para comprobar su efecto sobre la eficiencia replicativa (*fitness*) del VHB. En primer lugar, estas mutaciones se introdujeron en el plásmido pCRII.HBV.ayw por mutagénesis dirigida. Dicho plásmido contiene un monómero de VHB. Una vez obtenidos los mutantes, el genoma lineal del VHB mutado y no (*wild-type*), se obtuvo por corte enzimático mediante el enzima de restricción EcoRI. Este monómero lineal se utilizó para la transfección de células HepG2-NTCP (Ver detalles en el Capítulo 4 en la apartado 5. *Manuscrito completo: HBeAg negative chronic infected patients: sophisticated viral quasispecies, with a genotype-related pattern of mutation in Hepatitis B X gene*). El flujo de trabajo llevado a cabo se muestra en la siguiente figura 56.

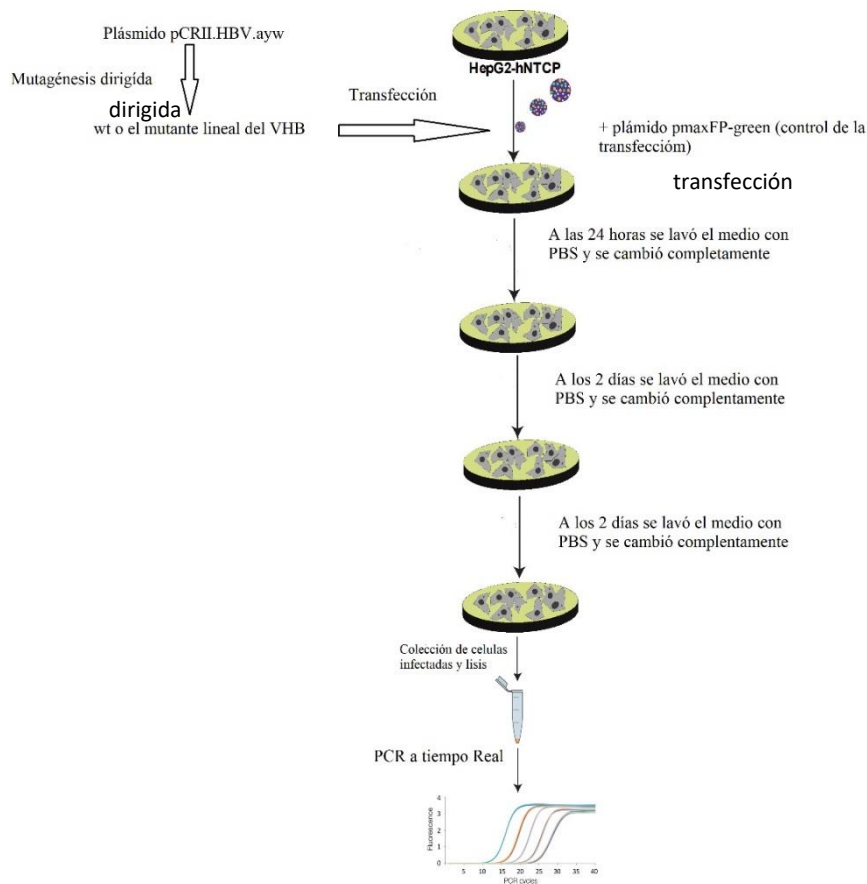


Figura 56. Flujo de trabajo del estudio *in vitro*

Los resultados obtenidos en este estudio se muestran en el Capítulo 4 en la apartado 5.
Manuscrito completo: HBeAg negative chronic infected patients: sophisticated viral quasispecies, with a genotype-related pattern of mutation in Hepatitis X gene.

BIBLIOGRAFÍA

1. Bibliografía

1. Ott JJJ, Stevens GAA, Groeger J, Wiersma STT. Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine*. 2012;30(12):2212-2219. doi:10.1016/j.vaccine.2011.12.116
2. Who hepatitis. <http://www.who.int/mediacentre/factsheets/fs204/es/>. Published 2017.
3. A L. Eine icterus epidemic. *Berl Klin Wochenschr*. 1985;22:20-23.
4. Block TM, Alter HJ, London WT, Bray M. A historical perspective on the discovery and elucidation of the hepatitis B virus. *Antiviral Res*. 2016;131:109-123. doi:10.1016/j.antiviral.2016.04.012
5. BLUMBERG BS, ALTER HJ, VISNICH S. A "NEW" ANTIGEN IN LEUKEMIA SERA. *JAMA*. 1965;191:541-546. <http://www.ncbi.nlm.nih.gov/pubmed/14239025>.
6. Almeida JD, Rubenstein D, Stott EJ. New antigen-antibody system in Australia-antigen-positive hepatitis. *Lancet (London, England)*. 1971;2(7736):1225-1227. <http://www.ncbi.nlm.nih.gov/pubmed/4143591>.
7. Dane DS, Cameron CH, Briggs M. Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet (London, England)*. 1970;1(7649):695-698. <http://www.ncbi.nlm.nih.gov/pubmed/4190997>.
8. 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(5733):646-650. doi:10.1038/281646a0
9. John E. Bennett & Raphael Dolin & Martin J. Blase. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*,. (Churchill Livingstone, ed.); 2000.
10. International Committee on Taxonomy of Viruses (ICTV). <https://talk.ictvonline.org/>.

11. Hahn CM, Iwanowicz LR, Cornman RS, Conway CM, Winton JR BV. Characterization of a novel hepadnavirus in the white sucker (*Catostomus commersonii*) from the Great Lakes region of the United States. *J Virol.* 2015;89:11801-11811.
12. Gust ID, Burrell CJ, Coulepis AG, Robinson WS, Zuckerman AJ. Taxonomic classification of human hepatitis B virus. *Intervirology.* 1986;25(1):14-29. doi:10.1159/000149651
13. Caballero A, Taberner D, Buti M, Rodriguez-Frias F. Hepatitis B virus: The challenge of an ancient virus with multiple faces and a remarkable replication strategy. *Antiviral Res.* 2018;158(July):34-44. doi:10.1016/j.antiviral.2018.07.019
14. Heermann KH, Goldmann U, Schwartz W, Seyffarth T, Baumgarten H, Gerlich WH. Large surface proteins of hepatitis B virus containing the pre-s sequence. *J Virol.* 1984;52(2):396-402. <http://www.ncbi.nlm.nih.gov/pubmed/6492255>.
15. Wynne SA, Crowther RA, Leslie AG. The crystal structure of the human hepatitis B virus capsid. *Mol Cell.* 1999;3(6):771-780. <http://www.ncbi.nlm.nih.gov/pubmed/10394365>.
16. Hu J, Liu K. Complete and incomplete hepatitis B virus particles: Formation, function, and application. *Viruses.* 2017;9(3):1-17. doi:10.3390/v9030056
17. Beck J, Nassal M. Hepatitis B virus replication. *World J Gastroenterol.* 2007;13(1):48-64. doi:10.3748/wjg.v13.i1.48
18. Nassal M. Hepatitis B viruses: reverse transcription a different way. *Virus Res.* 2008;134(1-2):235-249. doi:10.1016/j.virusres.2007.12.024
19. Kann M, Gerlich WH. Effect of core protein phosphorylation by protein kinase C on encapsidation of RNA within core particles of hepatitis B virus. *J Virol.* 1994;68(12):7993-8000. <http://www.ncbi.nlm.nih.gov/pubmed/7966589>.
20. Schädler S, Hildt E. HBV life cycle: entry and morphogenesis. *Viruses.* 2009;1(2):185-209. doi:10.3390/v1020185
21. Luangsay S, Zoulim F. Structure and Molecular Virology. In: *Viral Hepatitis*. Oxford, UK: John Wiley & Sons, Ltd; 2013:63-80. doi:10.1002/9781118637272.ch5

22. Gilbert RJC, Beales L, Blond D, et al. Hepatitis B small surface antigen particles are octahedral. *Proc Natl Acad Sci U S A*. 2005;102(41):14783-14788. doi:10.1073/pnas.0505062102
23. Hu J, Liu K. Complete and incomplete hepatitis B virus particles: Formation, function, and application. *Viruses*. 2017;9(3):1-17. doi:10.3390/v9030056
24. Huovila AP, Eder AM, Fuller SD. Hepatitis B surface antigen assembles in a post-ER, pre-Golgi compartment. *J Cell Biol*. 1992;118(6):1305-1320. <http://www.ncbi.nlm.nih.gov/pubmed/1522109>.
25. 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*. 2015;90(7):3330-3341. doi:10.1128/JVI.03109-15
26. Sattler F, Robinson WS, Okamoto H, et al. Hepatitis B viral DNA molecules have cohesive ends. *J Virol*. 1988;32(1):226-233. doi:10.1099/0022-1317-69-10-2575
27. Okamoto H, Tsuda F, Sakugawa H, et al. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol*. 1988;69 (Pt 10):2575-2583. doi:10.1099/0022-1317-69-10-2575
28. Kidd AH, Kidd-Ljunggren K. A revised secondary structure model for the 3'-end of hepatitis B virus pregenomic RNA. *Nucleic Acids Res*. 1996;24(17):3295-3301. <http://www.ncbi.nlm.nih.gov/pubmed/8811080>.
29. Ruiz-Opazo N, Chakraborty PR, Shafritz DA. Evidence for supercoiled hepatitis B virus DNA in chimpanzee liver and serum Dane particles: possible implications in persistent HBV infection. *Cell*. 1982;29(1):129-136. <http://www.ncbi.nlm.nih.gov/pubmed/6286137>.
30. Bock CT, Schwinn S, Locarnini S, et al. Structural organization of the hepatitis B virus minichromosome. *J Mol Biol*. 2001;307(1):183-196. doi:10.1006/jmbi.2000.4481
31. Rodríguez-Frias F, Jardi R. Virología molecular del virus de la hepatitis B. *Enferm Infecc Microbiol Clin*. 2008;26:2-10. doi:10.1016/S0213-005X(08)76514-5

32. Wooddell CI, Yuen M, Chan HL-Y, et al. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. *Sci Transl Med.* 2017;9(409):32-36. doi: 10.1126/scitranslmed.aan0241
33. Nassal M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut.* 2015;64(12):1972-1984. doi:10.1136/gutjnl-2015-309809
34. Charnay P, Pourcel C, Louise A, Fritsch A, Tiollais P. Cloning in *Escherichia coli* and physical structure of hepatitis B virion DNA. *Proc Natl Acad Sci U S A.* 1979;76(5):2222-2226. <http://www.ncbi.nlm.nih.gov/pubmed/377294>.
35. 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(5733):646-650. <http://www.ncbi.nlm.nih.gov/pubmed/399327>.
36. Pasek M, Goto T, Gilbert W, et al. Hepatitis B virus genes and their expression in *E. coli*. *Nature.* 1979;282(5739):575-579. <http://www.ncbi.nlm.nih.gov/pubmed/399329>.
37. Datta S, Chatterjee S, Veer V, Chakravarty R. Molecular biology of the hepatitis B virus for clinicians. *J Clin Exp Hepatol.* 2012;2(4):353-365. doi: 10.1016/j.jceh.2012.10.003
38. Garralda AC. Complejidad de la cuasiespecie del virus de la hepatitis B en la región X/preCore: asociación con la evolución de la infección con y sin tratamiento antiviral.
39. Chang C, Enders G, Sprengel R, Peters N, Varmus HE, Ganem D. Expression of the precore region of an avian hepatitis B virus is not required for viral replication. *J Virol.* 1987;61(10):3322-3325. <http://www.ncbi.nlm.nih.gov/pubmed/3041052>.
40. Pumpens P, Grens E. HBV core particles as a carrier for B cell/T cell epitopes. *Intervirology.* 2001;44(2-3):98-114. doi:10.1159/000050037
41. Lang T, Lo C, Skinner N, Locarnini S, Visvanathan K, Mansell A. The hepatitis B e antigen (HBeAg) targets and suppresses activation of the toll-like receptor signaling pathway. *J Hepatol.* 2011;55(4):762-769. doi:10.1016/j.jhep.2010.12.042

42. Milich DR, Schödel F, Hughes JL, Jones JE, Peterson DL. The hepatitis B virus core and e antigens elicit different Th cell subsets: antigen structure can affect Th cell phenotype. *J Virol.* 1997;71(3):2192-2201.
<http://www.ncbi.nlm.nih.gov/pubmed/9032353>.
43. Visvanathan K, Skinner NA, Thompson AJ V, et al. Regulation of Toll-like receptor-2 expression in chronic hepatitis B by the precore protein. *Hepatology.* 2007;45(1):102-110. doi:10.1002/hep.21482
44. Milich DR, Jones JE, Hughes JL, Price J, Raney AK, McLachlan A. Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance in utero? *Proc Natl Acad Sci U S A.* 1990;87(17):6599-6603.
<http://www.ncbi.nlm.nih.gov/pubmed/2395863>.
45. Ito K, Kim K-H, Lok AS-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(8):3507-3517. doi:10.1128/JVI.02348-08
46. Yang H-I, Lu S-N, Liaw Y-F, et al. Hepatitis B e antigen and the risk of hepatocellular carcinoma. *N Engl J Med.* 2002;347(3):168-174. doi:10.1056/NEJMoa013215
47. Melegari M, Wolf SK, Schneider RJ. Hepatitis B virus DNA replication is coordinated by core protein serine phosphorylation and HBx expression. *J Virol.* 2005;79(15):9810-9820. doi:10.1128/JVI.79.15.9810-9820.2005
48. Kann M, Sodeik B, Vlachou A, Gerlich WH, Helenius A. Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex. *J Cell Biol.* 1999;145(1):45-55. <http://www.ncbi.nlm.nih.gov/pubmed/10189367>.
49. Tang H, Oishi N, Kaneko S, Murakami S. Molecular functions and biological roles of hepatitis B virus x protein. *Cancer Sci.* 2006;97(10):977-983. doi:10.1111/j.1349-7006.2006.00299.x
50. Venkatakrisnan B, Zlotnick A. The Structural Biology of Hepatitis B Virus: Form and Function. *Annu Rev Virol.* 2016;3(1):429-451. doi:10.1146/annurev-virology-110615-042238

51. Vörös J, Urbanek A, Rautureau GJP, et al. Large-scale production and structural and biophysical characterizations of the human hepatitis B virus polymerase. *J Virol.* 2014;88(5):2584-2599. doi:10.1128/JVI.02575-13
52. Jones SA, Hu J. Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. *Emerg Microbes Infect.* 2013;2(9):e56-e56. doi:10.1038/emi.2013.56
53. Feng H, Hu K. Structural characteristics and molecular mechanism of hepatitis B virus reverse transcriptase. *Virol Sin.* 2009;24(6):509-517. doi:10.1007/s12250-009-3076-6
54. Junker-Niepmann M, Bartenschlager R, Schaller H. A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.* 1990;9(10):3389-3396. <http://www.ncbi.nlm.nih.gov/pubmed/2209549>.
55. Bartenschlager R, Junker-Niepmann M, Schaller H. The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J Virol.* 1990;64(11):5324-5332. <http://www.ncbi.nlm.nih.gov/pubmed/2214019>.
56. Kim S, Lee J, Ryu W-S. Four Conserved Cysteine Residues of the Hepatitis B Virus Polymerase Are Critical for RNA Pregenome Encapsidation. *J Virol.* 2009;83(16):8032-8040. doi:10.1128/JVI.00332-09
57. Daga PR, Duan J, Doerksen RJ. Computational model of hepatitis B virus DNA polymerase: molecular dynamics and docking to understand resistant mutations. *Protein Sci.* 2010;19(4):796-807. doi:10.1002/pro.359
58. Walsh AW, Langley DR, Colonno RJ, Tenney DJ. Mechanistic characterization and molecular modeling of hepatitis B virus polymerase resistance to entecavir. *PLoS One.* 2010;5(2):e9195. doi:10.1371/journal.pone.0009195
59. Das K, Xiong X, Yang H, et al. Molecular modeling and biochemical characterization reveal the mechanism of hepatitis B virus polymerase resistance to lamivudine (3TC) and emtricitabine (FTC). *J Virol.* 2001;75(10):4771-4779. doi:10.1128/JVI.75.10.4771-4779.2001

60. Li J, Du Y, Liu X, et al. Binding sensitivity of adefovir to the polymerase from different genotypes of HBV: molecular modeling, docking and dynamics simulation studies. *Acta Pharmacol Sin.* 2013;34(2):319-328. doi:10.1038/aps.2012.146
61. Stuyver LJ, Locarnini SA, Lok A, et al. Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. *Hepatology.* 2001;33(3):751-757. doi:10.1053/jhep.2001.22166
62. Chen Y, Marion PL. Amino acids essential for RNase H activity of hepadnaviruses are also required for efficient elongation of minus-strand viral DNA. *J Virol.* 1996;70(9):6151-6156.
<http://www.ncbi.nlm.nih.gov/pubmed/8709240>.
63. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell.* 2005;121(7):1005-1016. doi:10.1016/j.cell.2005.04.024
64. Tavis JE, Cheng X, Hu Y, et al. The hepatitis B virus ribonuclease H is sensitive to inhibitors of the human immunodeficiency virus ribonuclease H and integrase enzymes. *PLoS Pathog.* 2013;9(1):e1003125. doi:10.1371/journal.ppat.1003125
65. Chen Y, Robinson WS, Marion PL. Selected mutations of the duck hepatitis B virus P gene RNase H domain affect both RNA packaging and priming of minus-strand DNA synthesis. *J Virol.* 1994;68(8):5232-5238.
<http://www.ncbi.nlm.nih.gov/pubmed/8035519>.
66. Ramírez C, Gregori J, Buti M, et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res.* 2013;98(2):273-283. doi:10.1016/j.antiviral.2013.03.007
67. 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(2):613-620. <http://www.ncbi.nlm.nih.gov/pubmed/2153228>.
68. Locarnini S, Bowden S. Hepatitis B surface antigen quantification: not what it seems on the surface. *Hepatology.* 2012;56(2):411-414. doi:10.1002/hep.25732

69. Eble BE, MacRae DR, Lingappa VR, Ganem D. Multiple topogenic sequences determine the transmembrane orientation of the hepatitis B surface antigen. *Mol Cell Biol.* 1987;7(10):3591-3601. doi:10.1128/MCB.7.10.3591.Updated
70. Short JM, Chen S, Roseman AM, Butler PJG, Crowther RA. Structure of hepatitis B surface antigen from subviral tubes determined by electron cryomicroscopy. *J Mol Biol.* 2009;390(1):135-141. doi:10.1016/j.jmb.2009.04.059
71. Seitz S, Urban S, Antoni C, Böttcher B. Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. *EMBO J.* 2007;26(18):4160-4167. doi:10.1038/sj.emboj.7601841
72. Eble BE, Lingappa VR, Ganem D. The N-terminal (pre-S2) domain of a hepatitis B virus surface glycoprotein is translocated across membranes by downstream signal sequences. *J Virol.* 1990;64(3):1414-1419. <http://www.ncbi.nlm.nih.gov/pubmed/2304150>.
73. Schmitt S, Glebe D, Alving K, et al. Analysis of the pre-S2 N- and O-linked glycans of the M surface protein from human hepatitis B virus. *J Biol Chem.* 1999;274(17):11945-11957. <http://www.ncbi.nlm.nih.gov/pubmed/10207016>.
74. Bruss V, Hagelstein J, Gerhardt E, Galle PR. Myristylation of the large surface protein is required for hepatitis B virus in vitro infectivity. *Virology.* 1996;218(2):396-399. doi:10.1006/viro.1996.0209
75. De Falco S, Ruvo M, Verdoliva A, et al. N-terminal myristylation of HBV preS1 domain enhances receptor recognition. *J Pept Res.* 2001;57(5):390-400. <http://www.ncbi.nlm.nih.gov/pubmed/11350599>.
76. Prange R, Clemen A, Streeck RE. Myristylation is involved in intracellular retention of hepatitis B virus envelope proteins. *J Virol.* 1991;65(7):3919-3923. <http://www.ncbi.nlm.nih.gov/pubmed/2041101>.
77. Gazina E, Gallina A, Milanese G. Common localization of retention determinants in hepatitis B virus L protein from different strains. *J Gen Virol.* 1996;77 (Pt 12):3069-3075. doi:10.1099/0022-1317-77-12-3069
78. Glebe D, Bremer CM. The molecular virology of hepatitis B virus. *Semin Liver Dis.* 2013;33(2):103-112. doi:10.1055/s-0033-1345717

79. Bruss V, Lu X, Thomssen R, Gerlich WH. Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. *EMBO J*. 1994;13(10):2273-2279. <http://www.ncbi.nlm.nih.gov/pubmed/8194518>.
80. Yan H, Zhong G, Xu G, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife*. 2012;2012(1):1-28. doi:10.7554/eLife.00049
81. Feitelson MA, Miller RH. X gene-related sequences in the core gene of duck and heron hepatitis B viruses. *Proc Natl Acad Sci U S A*. 1988;85(16):6162-6166. <http://www.ncbi.nlm.nih.gov/pubmed/3261866>.
82. Belloni L, Pollicino T, De Nicola F, et al. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc Natl Acad Sci*. 2009;106(47):19975-19979. doi:10.1073/pnas.0908365106
83. Geng M. Molecular mechanism of hepatitis B virus X protein function in hepatocarcinogenesis. *World J Gastroenterol*. 2015;21(38):10732. doi:10.3748/wjg.v21.i38.10732
84. Minor MM, Slagle BL. Hepatitis B virus HBx protein interactions with the ubiquitin proteasome system. *Viruses*. 2014;6(11):4683-4702. doi:10.3390/v6114683
85. Bouchard MJ, Schneider RJ. The Enigmatic X Gene of Hepatitis B Virus. *J Virol*. 2004;78(23):12725-12734. doi:10.1128/JVI.78.23.12725-12734.2004
86. Kim H, Lee S-A, Kim B-J. X region mutations of hepatitis B virus related to clinical severity. *World J Gastroenterol*. 2016;22(24):5467-5478. doi:10.3748/wjg.v22.i24.5467
87. Martín-Vílchez S, Moreno-Otero R, Sanz-Cameno P. Mecanismos de acción y efectos de la proteína X (HBx) en la patogenia de la hepatitis crónica B. *Med Clin (Barc)*. 2013;140(11):508-513. doi:10.1016/j.medcli.2012.09.012
88. van Hemert FJ, van de Klundert MAA, Lukashov V V, Kootstra NA, Berkhout B, Zaaijer HL. Protein X of hepatitis B virus: origin and structure similarity with the central domain of DNA glycosylase. *PLoS One*. 2011;6(8):e23392. doi:10.1371/journal.pone.0023392

89. Hernández S, Venegas M, Brahm J, Villanueva RA. The viral transactivator HBx protein exhibits a high potential for regulation via phosphorylation through an evolutionarily conserved mechanism. *Infect Agent Cancer*. 2012;7(1):27. doi:10.1186/1750-9378-7-27
90. Datta S, Banerjee A, Chandra PK, et al. Analysis of hepatitis B virus X gene phylogeny, genetic variability and its impact on pathogenesis: Implications in Eastern Indian HBV carriers. *Virology*. 2008;382(2):190-198. doi:10.1016/j.virol.2008.09.007
91. Geng M, Xin X, Bi L-Q, Zhou L-T, Liu X-H. Molecular mechanism of hepatitis B virus X protein function in hepatocarcinogenesis. *World J Gastroenterol*. 2015;21(38):10732. doi:10.3748/wjg.v21.i38.10732
92. Slagle BL, Bouchard MJ. Hepatitis B Virus X and Regulation of Viral Gene Expression. *Cold Spring Harb Perspect Med*. 2016;6(3):a021402. doi:10.1101/cshperspect.a021402
93. Arii M, Takada S, Koike K. Identification of three essential regions of hepatitis B virus X protein for trans-activation function. *Oncogene*. 1992;7(3):397-403.
94. Minor M, Slagle B. Hepatitis B Virus HBx Protein Interactions with the Ubiquitin Proteasome System. *Viruses*. 2014;6(11):4683-4702. doi:10.3390/v6114683
95. Sirma H, Weil R, Rosmorduc O, et al. Cytosol is the prime compartment of hepatitis B virus X protein where it colocalizes with the proteasome. *Oncogene*. 1998;16(16):2051-2063. doi:10.1038/sj.onc.1201737
96. Höhne M, Schaefer S, Seifer M, Feitelson MA, Paul D, Gerlich WH. Malignant transformation of immortalized transgenic hepatocytes after transfection with hepatitis B virus DNA. *EMBO J*. 1990;9(4):1137-1145. <http://www.ncbi.nlm.nih.gov/pubmed/2323335>.
97. Henkler F, Hoare J, Waseem N, et al. Intracellular localization of the hepatitis B virus HBx protein. *J Gen Virol*. 2001;82(Pt 4):871-882. doi:10.1099/0022-1317-82-4-871
98. Cha M-Y, Ryu D-K, Jung H-S, Chang H-E, Ryu W-S. Stimulation of hepatitis B virus genome replication by HBx is linked to both nuclear and cytoplasmic HBx expression. *J Gen Virol*. 2009;90(4):978-986. doi:10.1099/vir.0.009928-0

99. Kew MC. Hepatitis B virus x protein in the pathogenesis of hepatitis B virus-induced hepatocellular carcinoma. *J Gastroenterol Hepatol*. 2011;26 Suppl 1:144-152. doi:10.1111/j.1440-1746.2010.06546.x
100. Bouchard MJ, Schneider RJ. The Enigmatic X Gene of Hepatitis B Virus. *J Virol*. 2004;78(23):12725-12734. doi:10.1128/JVI.78.23.12725-12734.2004
101. Decorsière A, Mueller H, van Breugel PC, et al. Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature*. 2016;531(7594):386-389. doi:10.1038/nature17170
102. Murphy CM, Xu Y, Li F, et al. Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep*. 2016;16(11):2846-2854. doi:10.1016/j.celrep.2016.08.026
103. Niu C, Livingston CM, Li L, et al. *The Smc5 / 6 Complex Restricts HBV When Localized to ND10 without Inducing an Innate Immune Response and Is Counteracted by the HBV X Protein Shortly after Infection.*; 2017. doi:10.1371/journal.pone.0169648
104. Belloni L, Pollicino T, De Nicola F, et al. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc Natl Acad Sci*. 2009;106(47):19975-19979. doi:10.1073/pnas.0908365106
105. Pollicino T, Belloni L, Raffa G, et al. Hepatitis B virus replication is regulated by the acetylation status of hepatitis B virus cccDNA-bound H3 and H4 histones. *Gastroenterology*. 2006;130(3):823-837. doi:10.1053/j.gastro.2006.01.001
106. Quasdorff M, Protzer U. Control of hepatitis B virus at the level of transcription. *J Viral Hepat*. 2010;17(8):527-536. doi:10.1111/j.1365-2893.2010.01315.x
107. Wang F, Marshall CB, Ikura M. Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target recognition. *Cell Mol Life Sci*. 2013;70(21):3989-4008. doi:10.1007/s00018-012-1254-4
108. Mukherji A, Janbandhu VC, Kumar V. HBx-dependent cell cycle deregulation involves interaction with cyclin E/A-cdk2 complex and destabilization of p27Kip1. *Biochem J*. 2007;401(1):247-256. doi:10.1042/BJ20061091

109. Keng VW, Tschida BR, Bell JB, Largaespada DA. Modeling hepatitis B virus X-induced hepatocellular carcinoma in mice with the Sleeping Beauty transposon system. *Hepatology*. 2011;53(3):781-790. doi:10.1002/hep.24091
110. Huo TI, Wang XW, Forgues M, et al. Hepatitis B virus X mutants derived from human hepatocellular carcinoma retain the ability to abrogate p53-induced apoptosis. *Oncogene*. 2001;20(28):3620-3628. doi:10.1038/sj.onc.1204495
111. Tarn C, Lee S, Hu Y, Ashendel C, Andrisani OM. Hepatitis B virus X protein differentially activates RAS-RAF-MAPK and JNK pathways in X-transforming versus non-transforming AML12 hepatocytes. *J Biol Chem*. 2001;276(37):34671-34680. doi:10.1074/jbc.M104105200
112. Feitelson MA, Bonamassa B, Arzumanyan A. The roles of hepatitis B virus-encoded X protein in virus replication and the pathogenesis of chronic liver disease. *Expert Opin Ther Targets*. 2014;18(3):293-306. doi:10.1517/14728222.2014.867947
113. Wang M-D, Wu H, Huang S, et al. HBx regulates fatty acid oxidation to promote hepatocellular carcinoma survival during metabolic stress. *Oncotarget*. 2016;7(6):6711-6726. doi:10.18632/oncotarget.6817
114. Dan Yue, Zhang Y, Cheng L, et al. Hepatitis B virus X protein (HBx)-induced abnormalities of nucleic acid metabolism revealed by (1)H-NMR-based metabonomics. *Sci Rep*. 2016;6:24430. doi:10.1038/srep24430
115. Kumar M, Jung SY, Hodgson AJ, Madden CR, Qin J, Slagle BL. Hepatitis B virus regulatory HBx protein binds to adaptor protein IPS-1 and inhibits the activation of beta interferon. *J Virol*. 2011;85(2):987-995. doi:10.1128/JVI.01825-10
116. Zhang Z, Torii N, Furusaka A, Malayaman N, Hu Z, Liang TJ. Structural and Functional Characterization of Interaction between Hepatitis B Virus X Protein and the Proteasome Complex. *J Biol Chem*. 2000;275(20):15157-15165. doi:10.1074/jbc.M910378199
117. Becker SA, Lee TH, Butel JS, Slagle BL. Hepatitis B virus X protein interferes with cellular DNA repair. *J Virol*. 1998;72(1):266-272. <http://www.ncbi.nlm.nih.gov/pubmed/9420223>.

118. Majano P, Lara-Pezzi E, López-Cabrera M, Apolinario A, Moreno-Otero R, García-Monzón C. Hepatitis B virus X protein transactivates inducible nitric oxide synthase gene promoter through the proximal nuclear factor kappaB-binding site: evidence that cytoplasmic location of X protein is essential for gene transactivation. *Hepatology*. 2001;34(6):1218-1224. doi:10.1053/jhep.2001.29626
119. Lara-Pezzi E, Majano PL, Gómez-Gonzalo M, et al. The hepatitis B virus X protein up-regulates tumor necrosis factor alpha gene expression in hepatocytes. *Hepatology*. 1998;28(4):1013-1021. doi:10.1002/hep.510280416
120. Guidotti LG, Chisari F V. Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu Rev Immunol*. 2001;19:65-91. doi:10.1146/annurev.immunol.19.1.65
121. Lu H-Z, Zhou J-H. Hepatitis B virus X protein up-regulates tumor necrosis factor- α expression in cultured mesangial cells via ERKs and NF- κ B pathways. *Asian Pac J Trop Biomed*. 2013;3(3):217-222. doi:10.1016/S2221-1691(13)60053-2
122. Wei C, Ni C, Song T, et al. The hepatitis B virus X protein disrupts innate immunity by downregulating mitochondrial antiviral signaling protein. *J Immunol*. 2010;185(2):1158-1168. doi:10.4049/jimmunol.0903874
123. Friedman SL. Mechanisms of hepatic fibrogenesis. *Gastroenterology*. 2008;134(6):1655-1669. doi:10.1053/j.gastro.2008.03.003
124. Martín-Vílchez S, Sanz-Cameno P, Rodríguez-Muñoz Y, et al. The hepatitis B virus X protein induces paracrine activation of human hepatic stellate cells. *Hepatology*. 2008;47(6):1872-1883. doi:10.1002/hep.22265
125. Holotnakova T, Tylkova L, Takacova M, et al. Role of the HBx oncoprotein in carbonic anhydrase 9 induction. *J Med Virol*. 2010;82(1):32-40. doi:10.1002/jmv.21671
126. Yoo Y-GY-GY-G, Oh SH, Park ES, et al. Hepatitis B virus X protein enhances transcriptional activity of hypoxia-inducible factor-1alpha through activation of mitogen-activated protein kinase pathway. *J Biol Chem*. 2003;278(40):39076-39084. doi:10.1074/jbc.M305101200

127. Yoo Y-G, Na T-Y, Seo H-W, et al. Hepatitis B virus X protein induces the expression of MTA1 and HDAC1, which enhances hypoxia signaling in hepatocellular carcinoma cells. *Oncogene*. 2008;27(24):3405-3413. doi:10.1038/sj.onc.1211000
128. Kim YC, Song KS, Yoon G, Nam MJ, Ryu WS. Activated ras oncogene collaborates with HBx gene of hepatitis B virus to transform cells by suppressing HBx-mediated apoptosis. *Oncogene*. 2001;20(1):16-23. doi:10.1038/sj.onc.1203840
129. Park NH, Chung YH. [Molecular mechanisms of hepatitis B virus-associated hepatocellular carcinoma]. *Korean J Hepatol*. 2007;13(3):320-340. <http://www.ncbi.nlm.nih.gov/pubmed/17898549>.
130. Molina-Jiménez F, Benedicto I, Murata M, et al. Expression of pituitary tumor-transforming gene 1 (PTTG1)/securin in hepatitis B virus (HBV)-associated liver diseases: evidence for an HBV X protein-mediated inhibition of PTTG1 ubiquitination and degradation. *Hepatology*. 2010;51(3):777-787. doi:10.1002/hep.23468
131. Renard Â, Levy R, Molina T, et al. The hepatitis B virus X gene potentiates c-myc -induced liver oncogenesis in transgenic mice. 1997;2.
132. Zhang Y, Liu H, Yi R, et al. Hepatitis B virus whole-X and X protein play distinct roles in HBV-related hepatocellular carcinoma progression. *J Exp Clin Cancer Res*. 2016;35(1):87. doi:10.1186/s13046-016-0366-3
133. Feitelson MA, Reis HMGPVGP V, Tufan NL, et al. Putative roles of hepatitis B x antigen in the pathogenesis of chronic liver disease. *Cancer Lett*. 2009;286(1):69-79. doi:10.1016/j.canlet.2008.12.010
134. Tian Y, Yang W, Song J, Wu Y, Ni B. Hepatitis B virus X protein-induced aberrant epigenetic modifications contributing to human hepatocellular carcinoma pathogenesis. *Mol Cell Biol*. 2013;33(15):2810-2816. doi:10.1128/MCB.00205-13
135. Kim DH, Kang HS, Kim K-H. Roles of hepatocyte nuclear factors in hepatitis B virus infection. *World J Gastroenterol*. 2016;22(31):7017-7029. doi:10.3748/wjg.v22.i31.7017

136. Lee GH, Wasser S, Lim SG. Hepatitis B pregenomic RNA splicing--the products, the regulatory mechanisms and its biological significance. *Virus Res.* 2008;136(1-2):1-7. doi:10.1016/j.virusres.2008.05.007
137. Zheng Y, Li J, Ou J. Regulation of hepatitis B virus core promoter by transcription factors HNF1 and HNF4 and the viral X protein. *J Virol.* 2004;78(13):6908-6914. doi:10.1128/JVI.78.13.6908-6914.2004
138. Quarleri J. Core promoter: A critical region where the hepatitis B virus makes decisions. *World J Gastroenterol.* 2014;20(2):425-435. doi:10.3748/wjg.v20.i2.425
139. Shaul Y, Rutter WJ, Laub O. A human hepatitis B viral enhancer element. *EMBO J.* 1985;4(2):427-430. <http://www.ncbi.nlm.nih.gov/pubmed/3926485>.
140. Moolla N, Kew M, Arbuthnot P. Regulatory elements of hepatitis B virus transcription. *J Viral Hepat.* 2002;9(5):323-331. <http://www.ncbi.nlm.nih.gov/pubmed/12225325>.
141. Yu X, Mertz JE. Promoters for synthesis of the pre-C and pregenomic mRNAs of human hepatitis B virus are genetically distinct and differentially regulated. *J Virol.* 1996;70(12):8719-8726. <http://www.ncbi.nlm.nih.gov/pubmed/8970999>.
142. Liang TJ. Hepatitis B: the virus and disease. *Hepatology.* 2009;49(5 Suppl):S13-21. doi:10.1002/hep.22881
143. Yuh CH, Ting LP. The genome of hepatitis B virus contains a second enhancer: cooperation of two elements within this enhancer is required for its function. *J Virol.* 1990;64(9):4281-4287. <http://www.ncbi.nlm.nih.gov/pubmed/2166817>.
144. Raney AK, Easton AJ, McLachlan A. Characterization of the minimal elements of the hepatitis B virus large surface antigen promoter. *J Gen Virol.* 1994;75 (Pt 10):2671-2679. doi:10.1099/0022-1317-75-10-2671
145. Siddiqui A, Jameel S, Mapoles J. Transcriptional control elements of hepatitis B surface antigen gene. *Proc Natl Acad Sci U S A.* 1986;83(3):566-570. <http://www.ncbi.nlm.nih.gov/pubmed/3456153>.
146. Zhou DX, Yen TS. The hepatitis B virus S promoter comprises A CCAAT motif and two initiation regions. *J Biol Chem.* 1991;266(34):23416-23421. <http://www.ncbi.nlm.nih.gov/pubmed/1840596>.

147. Lu CC, Yen TS. Activation of the hepatitis B virus S promoter by transcription factor NF-Y via a CCAAT element. *Virology*. 1996;225(2):387-394. doi:10.1006/viro.1996.0613
148. Lu CC, Chen M, Ou JH, Yen TS. Key role of a CCAAT element in regulating hepatitis B virus surface protein expression. *Virology*. 1995;206(2):1155-1158. doi:10.1006/viro.1995.1042
149. Raney AK, Milich DR, McLachlan A. Complex regulation of transcription from the hepatitis B virus major surface antigen promoter in human hepatoma cell lines. *J Virol*. 1991;65(9):4805-4811. <http://www.ncbi.nlm.nih.gov/pubmed/1651407>.
150. Zhang P, Raney AK, McLachlan A. Characterization of the hepatitis B virus X- and nucleocapsid gene transcriptional regulatory elements. *Virology*. 1992;191(1):31-41. <http://www.ncbi.nlm.nih.gov/pubmed/1413508>.
151. Treinin M, Laub O. Identification of a promoter element located upstream from the hepatitis B virus X gene. *Mol Cell Biol*. 1987;7(1):545-548. <http://www.ncbi.nlm.nih.gov/pubmed/3494193>.
152. Guo WT, Bell KD, Ou JH. Characterization of the hepatitis B virus EnhI enhancer and X promoter complex. *J Virol*. 1991;65(12):6686-6692. <http://www.ncbi.nlm.nih.gov/pubmed/1658368>.
153. Zheng YW, Riegler J, Wu J, Yen TS. Novel short transcripts of hepatitis B virus X gene derived from intragenic promoter. *J Biol Chem*. 1994;269(36):22593-22598. <http://www.ncbi.nlm.nih.gov/pubmed/8077209>.
154. Fukai K, Takada S, Yokosuka O, Saisho H, Omata M, Koike K. Characterization of a specific region in the hepatitis B virus enhancer I for the efficient expression of X gene in the hepatic cell. *Virology*. 1997;236(2):279-287. doi:10.1006/viro.1997.8750
155. Altinel K, Hashimoto K, Wei Y, et al. Single-Nucleotide Resolution Mapping of Hepatitis B Virus Promoters in Infected Human Livers and Hepatocellular Carcinoma. *J Virol*. 2016;90(23):10811-10822. doi:10.1128/JVI.01625-16
156. Hoofnagle JH. Reactivation of hepatitis B. *Hepatology*. 2009;49(5 Suppl):S156-65. doi:10.1002/hep.22945

157. Heise T, Sommer G, Reumann K, Meyer I, Will H, Schaal H. The hepatitis B virus PRE contains a splicing regulatory element. *Nucleic Acids Res.* 2006;34(1):353-363. doi:10.1093/nar/gkj440
158. Yu X, Mertz JE. Differential regulation of the pre-C and pregenomic promoters of human hepatitis B virus by members of the nuclear receptor superfamily. *J Virol.* 1997;71(12):9366-9374. <http://www.ncbi.nlm.nih.gov/pubmed/9371596>.
159. Doitsh G, Shaul Y. Enhancer I predominance in hepatitis B virus gene expression. *Mol Cell Biol.* 2004;24(4):1799-1808. <http://www.ncbi.nlm.nih.gov/pubmed/14749394>.
160. Lamontagne RJ, Bagga S, Bouchard MJ. Hepatitis B virus molecular biology and pathogenesis. *Hepatoma Res.* 2016;2:163-186. doi:10.20517/2394-5079.2016.05
161. Leistner CM, Gruen-Bernhard S, Glebe D. Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cell Microbiol.* 2008;10(1):122-133. doi:10.1111/j.1462-5822.2007.01023.x
162. Xie Y, Zhai J, Deng Q, Tiollais P, Wang Y, Zhao M. Entry of hepatitis B virus: mechanism and new therapeutic target. *Pathol Biol (Paris).* 2010;58(4):301-307. doi:10.1016/j.patbio.2010.04.001
163. Li J, Tong S. From DCPD to NTCP: the long journey towards identifying a functional hepatitis B virus receptor. *Clin Mol Hepatol.* 2015;21(3):193-199. doi:10.3350/cmh.2015.21.3.193
164. 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(3):985-994. doi:10.1002/hep.26125
165. Stieger B, Hagenbuch B, Landmann L, Höchli M, Schroeder A, Meier PJ. In situ localization of the hepatocytic Na⁺/Taurocholate cotransporting polypeptide in rat liver. *Gastroenterology.* 1994;107(6):1781-1787. <http://www.ncbi.nlm.nih.gov/pubmed/7958692>.
166. Seeger C, Mason WS. Molecular biology of hepatitis B virus infection. *Virology.* 2015;479-480:672-686. doi:10.1016/j.virol.2015.02.031

167. Rabe B, Glebe D, Kann M. Lipid-Mediated Introduction of Hepatitis B Virus Capsids into Nonsusceptible Cells Allows Highly Efficient Replication and Facilitates the Study of Early Infection Events. *J Virol.* 2006;80(11):5465-5473. doi:10.1128/JVI.02303-05
168. Panté N, Kann M. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol Biol Cell.* 2002;13(2):425-434. doi:10.1091/mbc.01-06-0308
169. Rabe B, Vlachou A, Panté N, Helenius A, Kann M. Nuclear import of hepatitis B virus capsids and release of the viral genome. *Proc Natl Acad Sci U S A.* 2003;100(17):9849-9854. doi:10.1073/pnas.1730940100
170. Gao W, Hu J. Formation of hepatitis B virus covalently closed circular DNA: removal of genome-linked protein. *J Virol.* 2007;81(12):6164-6174. doi:10.1128/JVI.02721-06
171. Cui X, McAllister R, Boregowda R, et al. Does Tyrosyl DNA Phosphodiesterase-2 Play a Role in Hepatitis B Virus Genome Repair? *PLoS One.* 2015;10(6):e0128401. doi:10.1371/journal.pone.0128401
172. Schreiner S, Nassal M. A role for the host DNA damage response in hepatitis B virus cccDNA formation—and beyond? *Viruses.* 2017;9(5). doi:10.3390/v9050125
173. Levrero M, Pollicino T, Petersen J, Belloni L, Raimondo G, Dandri M. Control of cccDNA function in hepatitis B virus infection. *J Hepatol.* 2009;51(3):581-592. doi:10.1016/j.jhep.2009.05.022
174. Waters CA, Strande NT, Wyatt DW, Pryor JM, Ramsden DA. Nonhomologous end joining: a good solution for bad ends. *DNA Repair (Amst).* 2014;17:39-51. doi:10.1016/j.dnarep.2014.02.008
175. Yang W, Summers J. Illegitimate replication of linear hepadnavirus DNA through nonhomologous recombination. *J Virol.* 1995;69(7):4029-4036. <http://www.ncbi.nlm.nih.gov/pubmed/7769660>.
176. Tang H, McLachlan A. A pregenomic RNA sequence adjacent to DR1 and complementary to epsilon influences hepatitis B virus replication efficiency. *Virology.* 2002;303(1):199-210. <http://www.ncbi.nlm.nih.gov/pubmed/12482672>.

177. Hu J, Seeger C. Hepadnavirus Genome Replication and Persistence. 2018;1-16. doi:10.1101/cshperspect.a021386
178. Abraham TM, Loeb DD. The topology of hepatitis B virus pregenomic RNA promotes its replication. *J Virol.* 2007;81(21):11577-11584. doi:10.1128/JVI.01414-07
179. Bardens A, Döring T, Stieler J, Prange R. Alix regulates egress of hepatitis B virus naked capsid particles in an ESCRT-independent manner. *Cell Microbiol.* 2011;13(4):602-619. doi:10.1111/j.1462-5822.2010.01557.x
180. Summers J, Smith PM, Huang MJ, Yu MS. Morphogenetic and regulatory effects of mutations in the envelope proteins of an avian hepadnavirus. *J Virol.* 1991;65(3):1310-1317. <http://www.ncbi.nlm.nih.gov/pubmed/1995945>.
181. Lin Y-Y, Liu C, Chien W-H, et al. New Insights into the Evolutionary Rate of Hepatitis B Virus at Different Biological Scales. *J Virol.* 2015;89(7):3512-3522. doi:10.1128/JVI.03131-14
182. Kramvis A. Genotypes and genetic variability of hepatitis B virus. *Intervirology.* 2014;57(3-4):141-150. doi:10.1159/000360947
183. Domingo E, Gomez J. Quasispecies and its impact on viral hepatitis. *Virus Res.* 2007;127(2):131-150. doi:10.1016/j.virusres.2007.02.001
184. Lim SG, Cheng Y, Guindon S, et al. Viral Quasi-Species Evolution During Hepatitis Be Antigen Seroconversion. *Gastroenterology.* 2007;133(3):951-958. doi:10.1053/j.gastro.2007.06.011
185. Cheng Y, Guindon S, Rodrigo A, et al. Cumulative viral evolutionary changes in chronic hepatitis B virus infection precedes hepatitis B e antigen seroconversion. *Gut.* 2013;62(9):1347-1355. doi:10.1136/gutjnl-2012-302408
186. Chen L, Zhang Q, Yu D, Wan M, Zhang X. Early changes of hepatitis B virus quasispecies during lamivudine treatment and the correlation with antiviral efficacy. *J Hepatol.* 2009;50(5):895-905. doi:10.1016/j.jhep.2008.12.018
187. Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. *Microbiol Mol Biol Rev.* 2012;76(2):159-216. doi:10.1128/MMBR.05023-11

188. Liu F, Chen L, Yu D-M, et al. Evolutionary patterns of hepatitis B virus quasispecies under different selective pressures: correlation with antiviral efficacy. *Gut*. 2011;60(9):1269-1277. doi:10.1136/gut.2010.226225
189. Gregori J, Perales C, Rodriguez-Frias F, Esteban JI, Quer J, Domingo E. Viral quasispecies complexity measures. *Virology*. 2016;493:227-237. doi: 10.1016/j.virol.2016.03.017
190. 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(11):7861-7884. doi:10.3390/ijms12117861
191. Jardi R, Rodriguez-Frias F, Taberner D, et al. Use of the Novel INNO-LiPA Line Probe Assay for Detection of Hepatitis B Virus Variants That Confer Resistance to Entecavir Therapy. *J Clin Microbiol*. 2009;47(2):485-488. doi:10.1128/JCM.01678-08
192. Lowe CF, Merrick L, Harrigan PR, Mazzulli T, Sherlock CH, Ritchie G. Implementation of Next-Generation Sequencing for Hepatitis B Virus Resistance Testing and Genotyping in a Clinical Microbiology Laboratory. *J Clin Microbiol*. 2016;54(1):127-133. doi:10.1128/JCM.02229-15
193. Hannachi N, Fredj N Ben, Bahri O, et al. Molecular analysis of HBV genotypes and subgenotypes in the Central-East region of Tunisia. *Virol J*. 2010;7:302. doi:10.1186/1743-422X-7-302
194. Jardi R, Buti M, Rodriguez-Frias F, et al. Rapid detection of lamivudine-resistant hepatitis B virus polymerase gene variants. *J Virol Methods*. 1999;83(1-2):181-187. <http://www.ncbi.nlm.nih.gov/pubmed/10598095>.
195. Chayama K, Suzuki Y, Kobayashi M, et al. Emergence and takeover of YMDD motif mutant hepatitis B virus during long-term lamivudine therapy and re-takeover by wild type after cessation of therapy. *Hepatology*. 1998;27(6):1711-1716. doi:10.1002/hep.510270634
196. Allen MI, Gauthier J, DesLauriers M, et al. Two sensitive PCR-based methods for detection of hepatitis B virus variants associated with reduced susceptibility to lamivudine. *J Clin Microbiol*. 1999;37(10):3338-3347. <http://www.ncbi.nlm.nih.gov/pubmed/10488202>.

197. Cane PA, Cook P, Ratcliffe D, Mutimer D, Pillay D. Use of real-time PCR and fluorimetry to detect lamivudine resistance-associated mutations in hepatitis B virus. *Antimicrob Agents Chemother.* 1999;43(7):1600-1608.
<http://www.ncbi.nlm.nih.gov/pubmed/10390210>.
198. Tran N, Berne R, Chann R, et al. European Multicenter Evaluation of High-Density DNA Probe Arrays for Detection of Hepatitis B Virus Resistance Mutations and Identification of Genotypes. *J Clin Microbiol.* 2006;44(8):2792-2800. doi:10.1128/JCM.00295-06
199. Wang R-S. Detection of YMDD mutants using universal template real-time PCR. *World J Gastroenterol.* 2006;12(8):1308. doi:10.3748/wjg.v12.i8.1308
200. Malmström S, Hannoun C, Lindh M. Mutation analysis of lamivudine resistant hepatitis B virus strains by TaqMan PCR. *J Virol Methods.* 2007;143(2):147-152. doi:10.1016/j.jviromet.2007.03.001
201. Mignardi M, Nilsson M. Fourth-generation sequencing in the cell and the clinic. *Genome Med.* 2014;6(4):31. doi:10.1186/gm548
202. Smith LM, Sanders JZ, Kaiser RJ, et al. Fluorescence detection in automated DNA sequence analysis. *Nature.* 321(6071):674-679. doi:10.1038/321674a0
203. Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res.* 1985;13(7):2399-2412. <http://www.ncbi.nlm.nih.gov/pubmed/4000959>.
204. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975;94(3):441-448. <http://www.ncbi.nlm.nih.gov/pubmed/1100841>.
205. Zoulim F, Locarnini S. Management of treatment failure in chronic hepatitis B. *Journal of hepatology.* doi:10.1016/S0168-8278(12)60012-9
206. Rodriguez-Frias F, Buti M, Tabernero D, Homs M. Quasispecies structure, cornerstone of hepatitis B virus infection: Mass sequencing approach. *World J Gastroenterol.* 2013;19(41):6995. doi:10.3748/wjg.v19.i41.6995

207. Ramírez C, Gregori J, Buti M, et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res.* 2013;98(2):273-283. doi:10.1016/j.antiviral.2013.03.007
208. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333-351. doi:10.1038/nrg.2016.49
209. Merriman B, Ion Torrent R&D Team, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis.* 2012;33(23):3397-3417. doi:10.1002/elps.201200424
210. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005;437(7057):376-380. doi:10.1038/nature03959
211. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31-46. doi:10.1038/nrg2626
212. An introduction to Next Generation Sequencing technology (NGS). <http://www.illumina.com/technology/next-generation-sequencing.html>.
213. Illumina. Guía breve de flujo de trabajo de principio a fin con Custom Protocol Selector. 2015:100. https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/translations/miseq-system-guide-15027617-01-esp.pdf.
214. Lahens NF, Ricciotti E, Smirnova O, et al. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics.* 2017;18(1):602. doi:10.1186/s12864-017-4011-0
215. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009;323(5910):133-138. doi:10.1126/science.1162986
216. Carter J-M, Hussain S. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome open Res.* 2017;2:23. doi:10.12688/wellcomeopenres.11246.3

217. Freytag S, Tian L, Lönnstedt I, Ng M, Bahlo M. Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. *F1000Research*. 2018;7:1297. doi:10.12688/f1000research.15809.2
218. Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. Deep sequencing: becoming a critical tool in clinical virology. *J Clin Virol*. 2014;61(1):9-19. doi:10.1016/j.jcv.2014.06.013
219. Homs M, Caballero A, Gregori J, et al. Clinical application of estimating hepatitis b virus quasispecies complexity by massive sequencing: Correlation between natural evolution and on-treatment evolution. *PLoS One*. 2014;9(11). doi:10.1371/journal.pone.0112306
220. Nishijima N, Marusawa H, Ueda Y, et al. Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. *PLoS One*. 2012;7(4):e35052. doi:10.1371/journal.pone.0035052
221. Rodriguez-Frías F, Tabernero D, Quer J, et al. Ultra-deep pyrosequencing detects conserved genomic sites and quantifies linkage of drug-resistant amino acid changes in the hepatitis B virus genome. *PLoS One*. 2012;7(5):e37874. doi:10.1371/journal.pone.0037874
222. Homs M, Buti M, Tabernero D, et al. Quasispecies dynamics in main core epitopes of hepatitis B virus by ultra-deep-pyrosequencing. *World J Gastroenterol*. 2012;18(42):6096-6105. doi:10.3748/wjg.v18.i42.6096
223. Margeridon-Thermet S, Svarovskaia ES, Babrzadeh F, et al. Low-level persistence of drug resistance mutations in hepatitis B virus-infected subjects with a past history of Lamivudine treatment. *Antimicrob Agents Chemother*. 2013;57(1):343-349. doi:10.1128/AAC.01601-12
224. Margeridon-Thermet S, Shulman NS, Ahmed A, et al. Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J Infect Dis*. 2009;199(9):1275-1285. doi:10.1086/597808
225. Homs M, Buti M, Quer J, et al. 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(19):8457-8471. doi:10.1093/nar/gkr451

226. Solmone M, Vincenti D, Prosperi MCF, Bruselles A, Ippolito G, Capobianchi MR. Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. *J Virol*. 2009;83(4):1718-1726. doi:10.1128/JVI.02011-08
227. Campbell PJ, Pleasance ED, Stephens PJ, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A*. 2008;105(35):13081-13086. doi:10.1073/pnas.0801523105
228. Loman NJ, Misra R V, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. 2012;30(5):434-439. doi:10.1038/nbt.2198
229. Schories M, Peters T, Rasenack J. Isolation, characterization and biological significance of hepatitis B virus mutants from serum of a patient with immunologically negative HBV infection. *J Hepatol*. 2000;33(5):799-811. <http://www.ncbi.nlm.nih.gov/pubmed/11097490>.
230. Görzer I, Guelly C, Trajanoski S, Puchhammer-Stöckl E. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. *J Virol Methods*. 2010;169(1):248-252. doi:10.1016/j.jviromet.2010.07.040
231. Tatematsu K, Tanaka Y, Kurbanov F, et al. A Genetic Variant of Hepatitis B Virus Divergent from Known Human and Ape Genotypes Isolated from a Japanese Patient and Provisionally Assigned to New Genotype J. *J Virol*. 2009;83(20):10538-10547. doi:10.1128/JVI.00462-09
232. Araujo NM, Waizbort R, Kay A. Hepatitis B virus infection from an evolutionary point of view: how viral, host, and environmental factors shape genotypes and subgenotypes. *Infect Genet Evol*. 2011;11(6):1199-1207. doi:10.1016/j.meegid.2011.04.017
233. Kramvis A, Kew M, François G. Hepatitis B virus genotypes. *Vaccine*. 2005;23(19):2409-2423. doi:10.1016/j.vaccine.2004.10.045
234. Bartholomeusz A, Schaefer S. Hepatitis B virus genotypes: comparison of genotyping methods. *Rev Med Virol*. 2004;14(1):3-16. doi:10.1002/rmv.400

235. Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, Kew MC. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol.* 2008;80(1):27-46. doi:10.1002/jmv.21049
236. Hardie DR WC. Analysis of the preS1 gene of hepatitis B virus (HBV) to define epidemiologically linked and un-linked infections in South Africa. *Arch Virol.* 1997;142(9):1829-1841.
237. Luca AS, Ursu RG, Teuşan T, Luca CM, Pleşca CE, Iancu LS. THE NEED FOR HBV GENOTYPING: A COST-EFFICIENT APPROACH. *Rev Med Chir Soc Med Nat Iasi.* 119(4):982-987. <http://www.ncbi.nlm.nih.gov/pubmed/26793838>.
238. Rodriguez-Frias F, Jardi R, Buti M, et al. Hepatitis B virus genotypes and G1896A precore mutation in 486 Spanish patients with acute and chronic HBV infection. *J Viral Hepat.* 2006;13(5):343-350. doi:10.1111/j.1365-2893.2005.00691.x
239. Sunbul M. Hepatitis B virus genotypes: Global distribution and clinical importance. *World J Gastroenterol.* 2014;20(18):5427. doi:10.3748/wjg.v20.i18.5427
240. Tong S, Revill P. Overview of hepatitis B viral replication and genetic variability. *J Hepatol.* 2016;64(1):S4-S16. doi:10.1016/j.jhep.2016.01.027
241. Mizokami M, Nakano T, Orito E, et al. Hepatitis B virus genotype assignment using restriction fragment length polymorphism patterns. *FEBS Lett.* 1999;450(1-2):66-71. <http://www.ncbi.nlm.nih.gov/pubmed/10350059>.
242. Sugauchi F, Orito E, Ichida T, et al. Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *J Virol.* 2002;76(12):5985-5992. <http://www.ncbi.nlm.nih.gov/pubmed/12021331>.
243. Cui C, Shi J, Hui L, et al. The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J Gen Virol.* 2002;83(Pt 11):2773-2777. doi:10.1099/0022-1317-83-11-2773

244. Abdou Chekaraou M, Brichtler S, Mansour W, et al. A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *J Gen Virol*. 2010;91(Pt 6):1609-1620. doi:10.1099/vir.0.018127-0
245. Schaefer S. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol*. 2007;13(1):14-21.
<http://www.ncbi.nlm.nih.gov/pubmed/17206751>.
246. Shi W, Zhang Z, Ling C, et al. Hepatitis B virus subgenotyping: history, effects of recombination, misclassifications, and corrections. *Infect Genet Evol*. 2013;16:355-361. doi:10.1016/j.meegid.2013.03.021
247. Paraskevis D, Magiorkinis G, Magiorkinis E, et al. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology*. 2013;57(3):908-916. doi:10.1002/hep.26079
248. Lauber C, Seitz S, Mattei S, et al. Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. *Cell Host Microbe*. 2017;22(3):387-399.e6. doi:10.1016/j.chom.2017.07.019
249. Yousif M, Mudawi H, Bakhiet S, Glebe D, Kramvis A. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. *BMC Infect Dis*. 2013;13:328. doi:10.1186/1471-2334-13-328
250. McMahon BJ. The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. *Hepatol Int*. 2009;3(2):334-342. doi:10.1007/s12072-008-9112-z
251. Papatheodoridis G, Buti M, Cornberg M, Janssen H, Mutimer D, Pol S et al. Guía de práctica clínica de las EASL: tratamiento de la infección crónica por el virus de la hepatitis B. *J hepato*. 2012;57:167-185. doi:10.1016/j.jmpmed.2011.06.012
252. Erhardt A, Blondin D, Hauck K, et al. Response to interferon alfa is hepatitis B virus genotype dependent: genotype A is more sensitive to interferon than genotype D. *Gut*. 2005;54(7):1009-1013. doi:10.1136/gut.2004.060327
253. Buster EHCJ, Flink HJ, Cakaloglu Y, et al. Sustained HBeAg and HBsAg loss after long-term follow-up of HBeAg-positive patients treated with peginterferon alpha-2b. *Gastroenterology*. 2008;135(2):459-467. doi:10.1053/j.gastro.2008.05.031

254. Li F, Zhang D, Li Y, et al. Whole genome characterization of hepatitis B virus quasispecies with massively parallel pyrosequencing. *Clin Microbiol Infect.* 2015;21(3):280-287. doi:10.1016/j.cmi.2014.10.007
255. Zhang Z-H. Genetic variation of hepatitis B virus and its significance for pathogenesis. *World J Gastroenterol.* 2016;22(1):126. doi:10.3748/wjg.v22.i1.126
256. Hsu C-W, Yeh C-T. Emergence of hepatitis B virus S gene mutants in patients experiencing hepatitis B surface antigen seroconversion after peginterferon therapy. *Hepatology.* 2011;54(1):101-108. doi:10.1002/hep.24363
257. Caligiuri P. Overview of hepatitis B virus mutations and their implications in the management of infection. *World J Gastroenterol.* 2016;22(1):145. doi:10.3748/wjg.v22.i1.145
258. Salpini R, Colagrossi L, Bellocchi MC, et al. Hepatitis B surface antigen genetic elements critical for immune escape correlate with hepatitis B virus reactivation upon immunosuppression. *Hepatology.* 2015;61(3):823-833. doi:10.1002/hep.27604
259. Lee S-A, Kim K-J, Kim D-W, Kim B-J. Male-specific W4P/R mutation in the pre-S1 region of hepatitis B virus, increasing the risk of progression of liver diseases in chronic patients. *J Clin Microbiol.* 2013;51(12):3928-3936. doi:10.1128/JCM.01505-13
260. Kim JH, Park YK, Park E-S, Kim K-H. Molecular diagnosis and treatment of drug-resistant hepatitis B virus. *World J Gastroenterol.* 2014;20(19):5708-5720. doi:10.3748/wjg.v20.i19.5708
261. Childs-Kean LM, Egelund EF, Jourjy J. Tenofovir Alafenamide for the Treatment of Chronic Hepatitis B Monoinfection. *Pharmacotherapy.* 2018;38(10):1051-1057. doi:10.1002/phar.2174
262. Zoulim F, Locarnini S. Optimal management of chronic hepatitis B patients with treatment failure and antiviral drug resistance. *Liver Int.* 2013;33:116-124. doi:10.1111/liv.12069
263. Roseman AM, Borschukova O, Berriman JA, Wynne SA, Pumpens P, Crowther RA. Structures of hepatitis B virus cores presenting a model epitope and their complexes with antibodies. *J Mol Biol.* 2012;423(1):63-78. doi:10.1016/j.jmb.2012.06.032

264. Bertoletti A, Gehring AJ. The immune response during hepatitis B virus infection. *J Gen Virol*. 2006;87(Pt 6):1439-1449. doi:10.1099/vir.0.81920-0
265. Bozkaya H, Ayola B, Lok AS. High rate of mutations in the hepatitis B core gene during the immune clearance phase of chronic hepatitis B virus infection. *Hepatology*. 1996;24(1):32-37. doi:10.1002/hep.510240107
266. Locarnini S, McMillan J, Bartholomeusz A. The hepatitis B virus and common mutants. *Semin Liver Dis*. 2003;23(1):5-20. doi:10.1055/s-2003-37587
267. Okamoto H, Yotsumoto S, Akahane Y, et al. Hepatitis B viruses with precore region defects prevail in persistently infected hosts along with seroconversion to the antibody against e antigen. *J Virol*. 1990;64(3):1298-1303.
<http://www.ncbi.nlm.nih.gov/pubmed/2304145>.
268. Song B-C, Cui XJ, Kim HU, Cho Y-K. Sequential accumulation of the basal core promoter and the precore mutations in the progression of hepatitis B virus-related chronic liver disease. *Intervirology*. 2006;49(5):266-273. doi:10.1159/000093456
269. Besharat S, Poustchi H, Mohamadkhani A, et al. Association of Mutations in the Basal Core Promoter and Pre-core Regions of the Hepatitis B Viral Genome and Longitudinal Changes in HBV Level in HBeAg Negative Individuals: Results From a Cohort Study in Northern Iran. *Hepat Mon*. 2015;15(2):e23875. doi:10.5812/hepatmon.23875
270. Kaymakoglu S, Baran B, Onel D, Badur S, Atamer T, Akyuz F. Acute hepatitis B due to immune-escape mutations in a naturally immune patient. *Acta Gastroenterol Belg*. 2014;77(2):262-265.
<http://www.ncbi.nlm.nih.gov/pubmed/25090827>.
271. Ma Q, Wang Y. Comprehensive analysis of the prevalence of hepatitis B virus escape mutations in the major hydrophilic region of surface antigen. *J Med Virol*. 2012;84(2):198-206. doi:10.1002/jmv.23183
272. Wang Y, Lau SH, Sham JS-T, Wu M-C, Wang T, Guan X-Y. Characterization of HBV integrants in 14 hepatocellular carcinomas: association of truncated X gene and hepatocellular carcinogenesis. *Oncogene*. 2004;23(1):142-148. doi:10.1038/sj.onc.1206889

273. Ma N-F, Lau SH, Hu L, et al. COOH-terminal truncated HBV X protein plays key role in hepatocarcinogenesis. *Clin Cancer Res.* 2008;14(16):5061-5068. doi:10.1158/1078-0432.CCR-07-5082
274. Feitelson MA, Reis HMGP V, Liu J, Lian Z, Pan J. Hepatitis B virus X antigen (HBxAg) and cell cycle control in chronic infection and hepatocarcinogenesis. *Front Biosci.* 2005;10:1558-1572. <http://www.ncbi.nlm.nih.gov/pubmed/15769646>.
275. Tu H, Bonura C, Giannini C, et al. Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Res.* 2001;61(21):7803-7810. <http://www.ncbi.nlm.nih.gov/pubmed/11691796>.
276. Liu X-H, Lin J, Zhang S-H, et al. COOH-terminal deletion of HBx gene is a frequent event in HBV-associated hepatocellular carcinoma. *World J Gastroenterol.* 2008;14(9):1346-1352. <http://www.ncbi.nlm.nih.gov/pubmed/18322946>.
277. Iavarone M, Trabut J-B, Delpuech O, et al. Characterisation of hepatitis B virus X protein mutants in tumour and non-tumour liver cells using laser capture microdissection. *J Hepatol.* 2003;39(2):253-261. <http://www.ncbi.nlm.nih.gov/pubmed/12873823>.
278. Jang JW, Chun J-Y, Park YM, et al. Mutational complex genotype of the hepatitis B virus X /precore regions as a novel predictive marker for hepatocellular carcinoma. *Cancer Sci.* 2012;103(2):296-304. doi:10.1111/j.1349-7006.2011.02170.x
279. Chen Q-Y, Harrison TJ, Sabin CA, et al. The Effect of HBV Genotype C on the Development of HCC Differs Between Wild-Type Viruses and Those With BCP Double Mutations (T(1762)A(1764)). *Hepat Mon.* 2014;14(2):e16214. doi:10.5812/hepatmon.16214
280. Lee J-H, Han K-H, Lee JM, Park JH, Kim H-S. Impact of hepatitis B virus (HBV) x gene mutations on hepatocellular carcinoma development in chronic HBV infection. *Clin Vaccine Immunol.* 2011;18(6):914-921. doi:10.1128/CVI.00474-10
281. Kwun HJ, Jang KL. 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(7):2202-2213. doi:10.1093/nar/gkh553

282. Buckwold VE, Xu Z, Chen M, Yen TS, Ou JH. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. *J Virol.* 1996;70(9):5845-5851.
<http://www.ncbi.nlm.nih.gov/pubmed/8709203>.
283. Hakami A, Ali A, Hakami A. Effects of hepatitis B virus mutations on its replication and liver disease severity. *Open Virol J.* 2013;7:12-18.
doi:10.2174/1874357901307010012
284. Geng X, Harry BL, Zhou Q, et al. Hepatitis B virus X protein targets the Bcl-2 protein CED-9 to induce intracellular Ca²⁺ increase and cell death in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A.* 2012;109(45):18465-18470.
doi:10.1073/pnas.1204652109
285. Kim H-J, Park J-H, Jee Y, et al. Hepatitis B virus X mutations occurring naturally associated with clinical severity of liver disease among Korean patients with chronic genotype C infection. *J Med Virol.* 2008;80(8):1337-1343.
doi:10.1002/jmv.21219
286. Ali A, Abdel-Hafiz H, Suhail M, et al. Hepatitis B virus, HBx mutants and their role in hepatocellular carcinoma. *World J Gastroenterol.* 2014;20(30):10238-10248. doi:10.3748/wjg.v20.i30.10238
287. Lee S-A, Mun H-S, Kim H, et al. Naturally occurring hepatitis B virus X deletions and insertions among Korean chronic patients. *J Med Virol.* 2011;83(1):65-70.
doi:10.1002/jmv.21938
288. Peng Y, Liu B, Hou J, et al. Naturally occurring deletions/insertions in HBV core promoter tend to decrease in hepatitis B e antigen-positive chronic hepatitis B patients during antiviral therapy. *Antivir Ther.* 2015;20(6):623-632.
doi:10.3851/IMP2955
289. Caballero A, Gregori J, Buti M, et al. Insertions and/or deletions in the main regulatory region of hepatitis B virus suggest multicoding of the X protein. *J Hepatol.* 2015;62(Supplement 2):S523. doi:10.1016/S0168-8278(15)30762-5
290. Zamora-Barquero Gastroenterólogo H, México H, Verdugo RM, et al. Hepatitis b. *asscat.* doi:10.1016/j.mpmed.2011.06.012

291. Hepatitis B. AEP. <http://vacunasaep.org/documentos/manual/cap-29#3>.
292. ISCIII. Ministerio de ciencia, innovacion y universidades. <http://www.isciii.es/ISCIII/es/contenidos/fd-servicios-cientifico-tecnicos/fd-vigilancias-alertas/fd-boletines/fd-boletin-epidemiologico-semanal-red/semanas-antiores-2018.shtml>.
293. Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viral Hepat.* 2004;11(2):97-107. <http://www.ncbi.nlm.nih.gov/pubmed/14996343>.
294. Alter MJ. Epidemiology and prevention of hepatitis B. *Semin Liver Dis.* 2003;23(1):39-46. doi:10.1055/s-2003-37583
295. Nakamoto Y, Kaneko S. Mechanisms of viral hepatitis induced liver injury. *Curr Mol Med.* 2003;3(6):537-544. <http://www.ncbi.nlm.nih.gov/pubmed/14527085>.
296. McMahon BJ. The natural history of chronic hepatitis B virus infection. *Hepatology.* 2009;49(5 Suppl):S45-55. doi:10.1002/hep.22898
297. Pungpapong S, Kim WR, Poterucha JJ. Natural history of hepatitis B virus infection: an update for clinicians. *Mayo Clin Proc.* 2007;82(8):967-975. doi:10.4065/82.8.967
298. Levrero M, Zucman-Rossi J. Mechanisms of HBV-induced hepatocellular carcinoma. *J Hepatol.* 2016;64(1 Suppl):S84-S101. doi:10.1016/j.jhep.2016.02.021
299. Trépo C, Chan HLY, Lok A. Hepatitis B virus infection. *Lancet.* 2014;384(9959):2053-2063. doi:10.1016/S0140-6736(14)60220-8
300. Dienstag JL. Hepatitis B virus infection. *N Engl J Med.* 2008;359(14):1486-1500. doi:10.1056/NEJMra0801644
301. Lee WM, Squires RH, Nyberg SL, Doo E, Hoofnagle JH. Acute liver failure: Summary of a workshop. *Hepatology.* 2008;47(4):1401-1415. doi:10.1002/hep.22177
302. Alegre F, Moreno D, Quiroga J. Infección aguda por el VHB. *An Sist Sanit Navar.* 2010;27:17-26. doi:10.4321/s1137-66272004000400003

303. Buti M, García-Samaniego J, Prieto M, et al. Documento de consenso de la AEEH sobre el tratamiento de la infección por el virus de la hepatitis B (2012). *Gastroenterol Hepatol*. 2012;35(7):512-528. doi:10.1016/j.gastrohep.2012.04.006
304. Samal J, Kandpal M, Vivekanandan P. Molecular mechanisms underlying occult hepatitis B virus infection. *Clin Microbiol Rev*. 2012;25(1):142-163. doi:10.1128/CMR.00018-11
305. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol*. 2017;67(0):370–398. doi:10.1016/j.jhep.2017.03.021
306. Hadziyannis SJ, Papatheodoridis G V. Hepatitis B e antigen-negative chronic hepatitis B: natural history and treatment. *Semin Liver Dis*. 2006;26(2):130-141. doi:10.1055/s-2006-939751
307. Knöll A, Pietrzyk M, Loss M, Goetz WA, Jilg W. Solid-organ transplantation in HBsAg-negative patients with antibodies to HBV core antigen: low risk of HBV reactivation. *Transplantation*. 2005;79(11):1631-1633. <http://www.ncbi.nlm.nih.gov/pubmed/15940056>.
308. Infección por el virus de la hepatitis B. sefh. http://formasefh.sefh.es/tecnifarmh/curso-farmacoterapia-infecciosa-paciente-externo/tema03_pagina04.php.
309. Wieland SF, Chisari F V. Stealth and cunning: hepatitis B and hepatitis C viruses. *J Virol*. 2005;79(15):9369-9380. doi:10.1128/JVI.79.15.9369-9380.2005
310. Isogawa M, Chung J, Murata Y, Kakimi K, Chisari F V. CD40 activation rescues antiviral CD8⁺ T cells from PD-1-mediated exhaustion. *PLoS Pathog*. 2013;9(7):e1003490. doi:10.1371/journal.ppat.1003490
311. Boni C, Laccabue D, Lampertico P, et al. Restored function of HBV-specific T cells after long-term effective therapy with nucleos(t)ide analogues. *Gastroenterology*. 2012;143(4):963-73.e9. doi:10.1053/j.gastro.2012.07.014
312. Kappus MR, Sterling RK. Extrahepatic manifestations of acute hepatitis B virus infection. *Gastroenterol Hepatol (N Y)*. 2013;9(2):123-126. <http://www.ncbi.nlm.nih.gov/pubmed/23983659>.

313. Zeisel MB, Lucifora J, Mason WS, et al. Towards an HBV cure: state-of-the-art and unresolved questions--report of the ANRS workshop on HBV cure. *Gut*. 2015;64(8):1314-1326. doi:10.1136/gutjnl-2014-308943
314. Lok AS, Zoulim F, Dusheiko G, Ghany MG. Hepatitis B cure: From discovery to regulatory approval. *Hepatology*. 2017;66(4):1296-1313. doi:10.1002/hep.29323
315. Testoni B, Levrero M, Zoulim F. Challenges to a Cure for HBV Infection. *Semin Liver Dis*. 2017;37(3):231-242. doi:10.1055/s-0037-1606212
316. Chen C-J, Yang H-I. Natural history of chronic hepatitis B REVEALed. *J Gastroenterol Hepatol*. 2011;26(4):628-638. doi:10.1111/j.1440-1746.2011.06695.x
317. Santantonio TA, Fasano M. Chronic hepatitis B: Advances in treatment. *World J Hepatol*. 2014;6(5):284-292. doi:10.4254/wjh.v6.i5.284
318. van Zonneveld M, Flink HJ, Verhey E, et al. The safety of pegylated interferon alpha-2b in the treatment of chronic hepatitis B: predictive factors for dose reduction and treatment discontinuation. *Aliment Pharmacol Ther*. 2005;21(9):1163-1171. doi:10.1111/j.1365-2036.2005.02453.x
319. Janssen HLA, van Zonneveld M, Senturk H, et al. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. *Lancet (London, England)*. 365(9454):123-129. doi:10.1016/S0140-6736(05)17701-0
320. Lau GKK, Piratvisuth T, Luo KX, et al. Peginterferon Alfa-2a, lamivudine, and the combination for HBeAg-positive chronic hepatitis B. *N Engl J Med*. 2005;352(26):2682-2695. doi:10.1056/NEJMoa043470
321. Chang T-T, Lai C-L, Kew Yoon S, et al. Entecavir treatment for up to 5 years in patients with hepatitis B e antigen-positive chronic hepatitis B. *Hepatology*. 2010;51(2):422-430. doi:10.1002/hep.23327
322. Striki A, Manolakopoulos S, Deutsch M, et al. Hepatitis B s antigen kinetics during treatment with nucleos(t)ides analogues in patients with hepatitis B e antigen-negative chronic hepatitis B. *Liver Int*. 2017;37(11):1642-1650. doi:10.1111/liv.13432

323. Lee KS, Kweon Y-O, Um S-H, et al. Efficacy and safety of entecavir versus lamivudine over 5 years of treatment: A randomized controlled trial in Korean patients with hepatitis B e antigen-negative chronic hepatitis B. *Clin Mol Hepatol*. 2017;23(4):331-339. doi:10.3350/cmh.2016.0040
324. Chang J, Guo F, Zhao X, Guo J-T. Therapeutic strategies for a functional cure of chronic hepatitis B virus infection. *Acta Pharm Sin B*. 2014;4(4):248-257. doi:10.1016/j.apsb.2014.05.002
325. Lin C-L, Kao J-H. Review article: novel therapies for hepatitis B virus cure - advances and perspectives. *Aliment Pharmacol Ther*. 2016;44(3):213-222. doi:10.1111/apt.13694
326. Volz T, Allweiss L, Ben MBarek M, et al. The entry inhibitor Myrcludex-B efficiently blocks intrahepatic virus spreading in humanized mice previously infected with hepatitis B virus. *J Hepatol*. 2013;58(5):861-867. doi:10.1016/j.jhep.2012.12.008
327. Lucifora J, Xia Y, Reisinger F, et al. Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA. *Science*. 2014;343(6176):1221-1228. doi:10.1126/science.1243462
328. Cradick TJ, Keck K, Bradshaw S, Jamieson AC, McCaffrey AP. Zinc-finger nucleases as a novel therapeutic strategy for targeting hepatitis B virus DNAs. *Mol Ther*. 2010;18(5):947-954. doi:10.1038/mt.2010.20
329. Weber ND, Stone D, Sedlak RH, et al. AAV-mediated delivery of zinc finger nucleases targeting hepatitis B virus inhibits active replication. *PLoS One*. 2014;9(5):e97579. doi:10.1371/journal.pone.0097579
330. Bloom K, Ely A, Arbuthnot P. Recent advances in use of gene therapy to treat hepatitis B virus infection. *Adv Exp Med Biol*. 2015;848:31-49. doi:10.1007/978-1-4939-2432-5_2
331. Chen J, Zhang W, Lin J, et al. An efficient antiviral strategy for targeting hepatitis B virus genome using transcription activator-like effector nucleases. *Mol Ther*. 2014;22(2):303-311. doi:10.1038/mt.2013.212
332. Maepa MB, Roelofse I, Ely A, Arbuthnot P. Progress and Prospects of Anti-HBV Gene Therapy Development. *Int J Mol Sci*. 2015;16(8):17589-17610. doi:10.3390/ijms160817589

333. Karimova M, Beschorner N, Dammermann W, et al. CRISPR/Cas9 nickase-mediated disruption of hepatitis B virus open reading frame S and X. *Sci Rep.* 2015;5:1-16. doi:10.1038/srep13734
334. Ramanan V, Shlomai A, Cox DBT, et al. CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. *Sci Rep.* 2015;5(1):10833. doi:10.1038/srep10833
335. Cai D, Mills C, Yu W, et al. Identification of disubstituted sulfonamide compounds as specific inhibitors of hepatitis B virus covalently closed circular DNA formation. *Antimicrob Agents Chemother.* 2012;56(8):4277-4288. doi:10.1128/AAC.00473-12
336. Zhang W, Chen J, Wu M, et al. PRMT5 restricts hepatitis B virus replication through epigenetic repression of covalently closed circular DNA transcription and interference with pregenomic RNA encapsidation. *Hepatology.* 2017;66(2):398-415. doi:10.1002/hep.29133
337. Livingston CM, Ramakrishnan D, Strubin M, Fletcher SP, Beran RK. Identifying and Characterizing Interplay between Hepatitis B Virus X Protein and Smc5/6. *Viruses.* 2017;9(4). doi:10.3390/v9040069
338. Gish RG, Yuen M-FF, Chan HLY, et al. Synthetic RNAi triggers and their use in chronic hepatitis B therapies with curative intent. *Antiviral Res.* 2015;121:97-108. doi:10.1016/j.antiviral.2015.06.019
339. Liang TJ, Block TM, McMahon BJ, et al. Present and future therapies of hepatitis B: From discovery to cure. *Hepatology.* 2015;62(6):1893-1908. doi:10.1002/hep.28025
340. Agarwal K, Gane E, Cheng W, et al. HBcrAg , HBV-RNA Declines in A Phase 2a Study Evaluating the Multi-Dose Activity of ARB-1467 in HBeAg-Positive and Negative Virally Suppressed Patients With Hepatitis B. :1-16.
341. Tavis JE, Lomonosova E. NVR 3-778 Plus Pegylated Interferon- α Treatment for Chronic Hepatitis B Viral Infections: Could $1 + 1 = 3$? *Gastroenterology.* 2018;154(3):481-482. doi:10.1053/j.gastro.2018.01.011
342. Testoni B, Durantel D, Zoulim F. Novel targets for hepatitis B virus therapy. *Liver Int.* 2017;37 Suppl 1:33-39. doi:10.1111/liv.13307

343. Vaillant A. Nucleic acid polymers: Broad spectrum antiviral activity, antiviral mechanisms and optimization for the treatment of hepatitis B and hepatitis D infection. *Antiviral Res.* 2016;133:32-40. doi:10.1016/j.antiviral.2016.07.004
344. 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(6):e0156667. doi:10.1371/journal.pone.0156667
345. Cai CW, Lomonosova E, Moran EA, et al. Hepatitis B virus replication is blocked by a 2-hydroxyisoquinoline-1,3(2H,4H)-dione (HID) inhibitor of the viral ribonuclease H activity. *Antiviral Res.* 2014;108:48-55. doi:10.1016/j.antiviral.2014.05.007
346. Pei Y, Wang C, Yan SF, Liu G. Past, Current, and Future Developments of Therapeutic Agents for Treatment of Chronic Hepatitis B Virus Infection. *J Med Chem.* 2017;60(15):6461-6479. doi:10.1021/acs.jmedchem.6b01442
347. Janssen HLA, Brunetto MR, Kim YJ, et al. Safety, efficacy and pharmacodynamics of vesatolimod (GS-9620) in virally suppressed patients with chronic hepatitis B. *J Hepatol.* 2018;68(3):431-440. doi:10.1016/j.jhep.2017.10.027
348. Gane EJ, Lim Y-S, Gordon SC, et al. The oral toll-like receptor-7 agonist GS-9620 in patients with chronic hepatitis B virus infection. *J Hepatol.* 2015; 63(2):320-328. doi:10.1016/j.jhep.2015.02.037
349. Janssen HLA, Brunetto MR, Kim YJ, et al. Safety, efficacy and pharmacodynamics of vesatolimod (GS-9620) in virally suppressed patients with chronic hepatitis B. *J Hepatol.* 2018;68(3):431-440. doi:10.1016/j.jhep.2017.10.027
350. Korolowicz KE, Iyer RP, Czerwinski S, et al. Antiviral Efficacy and Host Innate Immunity Associated with SB 9200 Treatment in the Woodchuck Model of Chronic Hepatitis B. *PLoS One.* 2016;11(8):e0161313. doi:10.1371/journal.pone.0161313

351. Fulda S. Promises and Challenges of Smac Mimetics as Cancer Therapeutics. *Clin Cancer Res.* 2015;21(22):5030-5036. doi:10.1158/1078-0432.CCR-15-0365
352. Ebert G, Allison C, Preston S, et al. Eliminating hepatitis B by antagonizing cellular inhibitors of apoptosis. *Proc Natl Acad Sci U S A.* 2015;112(18):5803-5808. doi:10.1073/pnas.1502400112
353. Zhang E, Kosinska A, Lu M, Yan H, Roggendorf M. Current status of immunomodulatory therapy in chronic hepatitis B, fifty years after discovery of the virus: Search for the “magic bullet” to kill cccDNA. *Antiviral Res.* 2015;123:193-203. doi:10.1016/j.antiviral.2015.10.009
354. Brahmania M, Feld J, Arif A, Janssen HLA. New therapeutic agents for chronic hepatitis B. *Lancet Infect Dis.* 2016;16(2):e10-21. doi:10.1016/S1473-3099(15)00436-3
355. Gallo Echeverri S, Caraballo Cordovez C, Orozco Marín M, et al. Tratamiento actual y nuevas terapias contra la infección crónica por el virus de la Hepatitis B. *Rev Colomb Gastroenterol.* 2017;32(2):131. doi:10.22516/25007440.138
356. Wang Q, Zhang W, Ye L, Zhang X. A mutant of HBx (HBxDelta127) promotes hepatoma cell growth via sterol regulatory element binding protein 1c involving 5-lipoxygenase. *Acta Pharmacol Sin.* 2010;31(3):367-374. doi:10.1038/aps.2010.5
357. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 1989;8(12):3867-3874. doi:10.1093/emboj/16.6.1248
358. Levrero M, Testoni B, Zoulim F. ScienceDirect HBV cure : why , how , when ? *Curr Opin Virol.* 2016;18:135-143. doi:10.1016/j.coviro.2016.06.003
359. Eley T, Russ R, Streinu-cercel A, et al. Pharmacokinetics and exploratory exposure-response of siRNAs administered monthly as ARB-001467 (ARB-1467) in a Phase 2a study in HBeAg positive and negative virally suppressed subjects with chronic hepatitis B. 001467.
360. Pujol FH, Navas M-CC, Hainaut P, Chemin I. Worldwide genetic diversity of HBV genotypes and risk of hepatocellular carcinoma. *Cancer Lett.* 2009;286(1):80-88. doi:10.1016/j.canlet.2009.07.013

361. Team RC, R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. doi:10.1038/sj.hdy.6800737
362. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80. doi:10.1186/gb-2004-5-10-r80
363. H. Pages, P. Aboyoun, R. Gentleman SD. Biostrings: String objects representing biological sequences, and matching algorithms. *R Packag Key citeulike11644278*.
364. Gavrillov K, Saltzman WM. Therapeutic siRNA: principles, challenges, and strategies. *Yale J Biol Med.* 2012;85(2):187-200. <http://www.ncbi.nlm.nih.gov/pubmed/22737048>.
365. Schneider TD. Information content of individual genetic sequences. *J Theor Biol.* 1997;189(4):427-441. doi:10.1006/jtbi.1997.0540
366. Sheila Maestro, Maria Francesca Cortese, Carolina González, David Tabernero, Mar Riveiro-Barciela M, Buti GG-AFR-F. Antiviral Activity of a New siRNA Targeting a Hyper-Conserved Region in Hepatitis B X Gene Transcripts: Preliminary Data. 2018.
367. Soria ME, Gregori J, Chen Q, et al. Pipeline for specific subtype amplification and drug resistance detection in hepatitis C virus. *BMC Infect Dis.* 2018;18(1):446. doi:10.1186/s12879-018-3356-6
368. González C, Tabernero D, Cortese MF, et al. Detection of hyper-conserved regions in hepatitis B virus X gene potentially useful for gene therapy. *World J Gastroenterol.* 2018;24(19):2095-2107. doi:10.3748/wjg.v24.i19.2095
369. Ringelhan M, O'Connor T, Protzer U, Heikenwalder M. The direct and indirect roles of HBV in liver cancer: prospective markers for HCC screening and potential therapeutic targets. *J Pathol.* 2015;235(2):355-367. doi:10.1002/path.4434
370. Papatheodoridis G V. Hepatitis B virus treatment: Which patients can have treatment deferred? *Clin Liver Dis.* 2013. doi:10.1002/clid.160

371. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27(21):2957-2963. doi:10.1093/bioinformatics/btr507
372. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 1998;8(3):186-194. <http://www.ncbi.nlm.nih.gov/pubmed/9521922>.
373. Datta S, Chatterjee S, Veer V. Recent advances in molecular diagnostics of hepatitis B virus. 2014;20(40):14615-14625. doi:10.3748/wjg.v20.i40.14615
374. Lavanchy D, Kane M. Global Epidemiology of Hepatitis B Virus Infection. In: ; 2016:187-203. doi:10.1007/978-3-319-22330-8_9
375. Thongthae N, Payungporn S, Poovorawan Y, T-Thienprasert NP. A rational study for identification of highly effective siRNAs against hepatitis B virus. *Exp Mol Pathol*. 2014;97(1):120-127. doi:10.1016/j.yexmp.2014.06.006
376. Lee J, Chung JH, Kim HMH, Kim DW, Kim HMH. Designed nucleases for targeted genome editing. *Plant Biotechnol J*. 2016;14(2):448-462. doi:10.1111/pbi.12465
377. Slagle BL, Bouchard MJ. Role of HBx in hepatitis B virus persistence and its therapeutic implications. *Curr Opin Virol*. 2018;30:32-38. doi:10.1016/j.coviro.2018.01.007
378. Taida T, Arai M, Kanda T, et al. The prognosis of hepatitis B inactive carriers in Japan: a multicenter prospective study. *J Gastroenterol*. 2017;52(1):113-122. doi:10.1007/s00535-016-1229-6
379. Jones LR, Sede M, Manrique JM, Quarleri J. Virus evolution during chronic hepatitis B virus infection as revealed by ultradeep sequencing data. 2019;(2016):435-444. doi:10.1099/jgv.0.000344
380. Rajoriya N, Combet C, Zoulim F, Janssen HLA. Review How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B . Time for an individualised approach ? *J Hepatol*. 2017;67(6):1281-1297. doi:10.1016/j.jhep.2017.07.011

381. Li S, Wang Z, Li Y, Ding G. Adaptive evolution of proteins in hepatitis B virus during divergence of genotypes. *Sci Rep.* 2017;7(1):1-11. doi:10.1038/s41598-017-02012-8
382. Mirabelli C, Surdo M, Van Hemert F, et al. Specific mutations in the C-terminus domain of HBV surface antigen significantly correlate with low level of serum HBV-DNA in patients with chronic HBV infection. *J Infect.* 2015;70(3):288-298. doi:10.1016/j.jinf.2014.10.015
383. Salpini R, Surdo M, Cortese MF, et al. The novel HBx mutation F30V correlates with hepatocellular carcinoma in vivo, reduces hepatitis B virus replicative efficiency and enhances anti-apoptotic activity of HBx N terminus in vitro. *Clin Microbiol Infect.* November 2018. doi:10.1016/j.cmi.2018.11.017
384. Tsiatis AC, Norris-kirby A, Rich RG, et al. Comparison of Sanger Sequencing , Pyrosequencing , and Melting Curve Analysis for the Detection of KRAS Mutations Diagnostic and Clinical Implications. *J Mol Diagnostics.* 2010;12(4):425-432. doi:10.2353/jmoldx.2010.090188