UAB

Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

ANALYSIS OF THE NATURAL EVOLUTION OF THE HEPATITIS B VIRUS QUASISPECIES IN THE PRECORE/CORE REGIONS AND TREATMENT EFFECTS USING CONVENTIONAL TECHNIQUES AND ULTRA-DEEP PYROSEQUENCING

Memòria presentada per **Maria Homs i Riba** per optar al grau de Doctor per la Universitat Autònoma de Barcelona

Directors de la tesi

Dr. Francisco Rodriguez-Frias i Dra. Maria Buti Ferret

Tutora de la tesi

Dra Anna Meseguer Navarro

Barcelona, 2012

Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

ANALYSIS OF THE NATURAL EVOLUTION OF THE HEPATITIS B VIRUS QUASISPECIES IN THE PRECORE/CORE REGIONS AND TREATMENT EFFECTS USING CONVENTIONAL TECHNIQUES AND ULTRA-DEEP PYROSEQUENCING

Maria Homs Riba

INDEX

AGRAÏMENTS

La part dels agraïments d'una tesis suposo que és la part més personal, per això m'agradaria poder incloure totes aquelles persones que m'han ajudat en la feina realitzada en aquests 4 anys i mig. Doncs bé, primer de tot dir que segurament em deixaré algú, però per totes aquelles persones a qui no surt el seu nom però que han format part d'aquest llarg camí… gràcies!

Primer de tot, vull donar les gràcies a l'investigador principal del grup del CIBERehd, el Dr. Rafael Esteban ja que el seu reconeixament i prestigi ens dóna força i permet consolidar-nos com a grup de recerca.

D'una forma molt especial, vull agraïr-li a la codirectora de la tesi, la Dra. Maria Buti, el seu entusiasme, les infinites ganes de treballar i la seva capacitat de simplificar qualsevol problema (fins al punt que la vas a veure al seu despatx per un problema i surts d'allà sense saber perquè la visitaves…). La seva *hiper*activitat, m'ha estimulat en seguir la recerca aquests anys i a més a més, i per damunt de tot, ha fet un gran vot de confiança en mi i per això, li estaré sempre agraïda.

Agraeixo infinitament al Dr. Francisco Rodriguez-Frias com m'ha ensenyat a "pensar", "postular", "hipotetitzar", "suggerir"… El Paco és el codirector i la persona del laboratori que ha coordinat la meva feina diàriament. La seva capacitat per trobar explicacions als resultats, la seva motivació i dedicació a la feina han donat com a resultat aquesta tesi. Sense les seves idees, aquesta tesi tindria la meitat de gruix. Les seves ganes de fer recerca són inacabables i amb el dia a dia ho transmetent, però del que li estic més agraïda és que sempre m'ha entès i m'ha fet sentir una persona important en el laboratori. Molt sincerament, li vull donar les gràcies.

També vull donar les gràcies als altres membres del CIBERehd, el Dr. Rosend Jardí, la Dra. Mélanie Schaper i el David Tabernero, amb qui he treballat dia a dia i formen part de la feina d'aquests anys. Al laboratori, he conviscut diariament amb amb la Tona, la Montse i el Gerard. Els tres formen un grup bastant peculiar i fan que a partir de les 3 de la tarda el laboratori es quedi buit, i hi faltin bromes, riures, música i… alegria! A més, en l'últim any hem tingut noves incorporacions d'algunes residents, com la Sílvia, la Clara i la Marta. Espero que la Sílvia es quedi amb nosaltres durant molt de temps perquè a part de ser una gran treballadora, és un

encant! I la Clara, amb qui he treballat bastant els últims mesos (i espero seguir un temps més) li dóno les gràcies per la seva eficiència i eficància treballant.

De tot cor puc dir, que tant el Paco, el Rosend i la Mélanie, com el David, la Tona, la Montse, el Gerard, la Sílvia i la Clara han fet que les hores de feina fossin més alegres i agradables. Per tots els esmorzars al bar, els vermuts que ens hem organitzat al laboratori, els dinars amb tupper i les converses tant sinceres que hem tingut …gràcies! Amb el permís de tots ells, un agraïment molt especial a la Tona, una (petita) gran persona, gran consellera, conseqüent, treballadora, viatjera, amb una cultura envejable i culé fins a la mort… des d'aquí (el lloc on em troba cada matí i em despedeix al migdia) li dóno les gràcies de manera molt especial.

Vull dedicar un agraïment als membres del grup de Malalties Hepàtiques de l'Institut de Recerca, perquè pujar allà dalt és una "pausa agradable". Evidentment, no puc deixar de dedicar unes línies al Dr. Josep Quer, perquè sense ell no hagués arribat fins al final. A part de ser un gran professional, m'ha escoltat sempre incondicionalment i ha estat l'optimisme personificat en tots els moments. Suposo que qui llegeixi aquestes línies i conegui al Josep, m'entendrà perfectament i fins i tot, potser pensarà que em quedo curta, pero Josep, de veritat… moltes, moltíssimes gràcies per tot!

No vull deixar d'agraïr la feina que ha fet el personal de la UCTS de l'Institut de Recerca, la Paqui i la Rosa; el personal de la UEB (Isra, Alejandro i Àlex) i el famós BioPep (formalment conegut com Josep Gregori) que s'ha incorporat recentment pero ha fet una feina inimaginable.

Ja per acabar, vull donar les gràcies a la Celine, per les seves aportacions, suggerències i la correcció de l'anglès en els articles, a més de la correcció que ens ha fet de la tesi.

Moltes gràcies a tots!

A part de l'ambient professional, vull agraïr a totes les persones que m'han acompanyat fora de la feina. El primer i més especial, el (*meu*) David. La seva alegria, simplicitat, comprensió i la seva capacitat d'organitzar coses pel temps lliure (excursions, escalades, correr, bici…) han estat claus per finalitzar la tesi; especialment aquests últims mesos.

Gràcies als meus pares, a l'Anna, l'Edu, el Pep i la Marta, que m'han fet companyia, em transmeten alegria continuament i sempre estan disposats a ajudar-me i escoltar-me. Mil gràcies family! I ja també aprofito per afegir un agraïmet especial al Pipo per les seves visites al laboratori.

No vull deixar-me als meus amics i amigues que han estat el millor esbarjo entre tan estrés; pels sopars junts, les festes, els caps de setmana, calçotades, vacances, per tot el que hem compartit junts … gràcies!

AKNOWLEDGEMENTS

ABBREVIATIONS

AA   Amino Acid

ALT   Alanine Aminotransferase

DNA   Deoxyribonucleic Acid

dNTP   deoxi Nucletide Triphosphate

HBV   Hepatitis B Virus

HCV   Hepatitis C Virus

HIV   Human Immunodeficiency Virus

HCC   Hepatocellular carcinoma

HLA   Human leukocyte antigen

IFN   Interferon

IL   Interleukin

IG   Immunoglobulin

NT   Nucleotide

ORF   Open reading frames

PCR   Polymerase Chain Reaction

RNA   Ribonucleic Acid

RT   Retrotranscriptase

TNF   Tumor necrosis factors

UDPS   Ultra-deep Pyrosequencing

All HBV genome positions described in the present thesis are referred to the single site targeted by EcoRI restriction enzyme.

The letter placed before the number corresponds to the wild type nucleotide or amino acid, whereas the letter placed after the number indicates the mutated form.

Examples:

a) G1896A, G is changed to A in NT position 1896.

b) pcW28X, the W (tryptophan) in codon 28 of the preCore region is substituted by a stop codon.

c) rtM204V, the M (methionine) in codon 204 of the viral reverse transcriptase is changed to V (valine).

Introduction

1   INTRODUCTION

1.1 Hepatitis B Virus Virology

1.1.1 History

Alter and Blumberg discovered hepatitis B virus (HBV) in 1965 (1) with detection of a new blood antigen in an Australian aborigine. At that time, it was called Australian antigen, and currently it is known as HBV surface antigen. In 1974, Dane et al. detected HBV viral particles by electron microscopy, and in 1979, the viral genome was completely sequenced (2). By the early 1980s, a vaccine had been obtained from plasma of HBV-infected patients (3). Currently, the HBV vaccine is produced by recombinant DNA technology.

1.1.2 Classification

HBV is a member of the *Hepadnaviridae* family, genus *Orthohepadnavirus*. As its family name indicates, these viruses preferably infect the liver and have a DNA genome: Hepa (liver)-dna-viridae (DNA virus). HBV has also been detected in other tissues, such as kidney and pancreas, and in mononuclear cells (4,5).

1.1.3 Epidemiology

Hepatitis B virus infection has a worldwide distribution. The disability-adjusted life year (DALY) is a measure of overall disease burden, expressed as the number of years lost due to illness, disability, or early death. Figure 1 shows the most recent DALY data (updated in 2004) attending to HBV distribution. It has been estimated that more than 2 billion people (around one third of the total world population) have been infected by HBV. Of these, approximately 360 million are now chronically infected and have a high risk of developing serious liver disease, liver cancer, or liver-related death. It is estimated that HBV causes 500,000 – 700,000 deaths per year worldwide (6).

Figure 1 DALY for hepatitis B infection per 100,000 habitants (6)

Figure 1 DALY for hepatitis B infection per 100,000 habitants (6)

no data
less than 10
10-20
20-30
30-40
40-50
50-80
80-100
100-120
120-150
150-200
200-500
more than 500
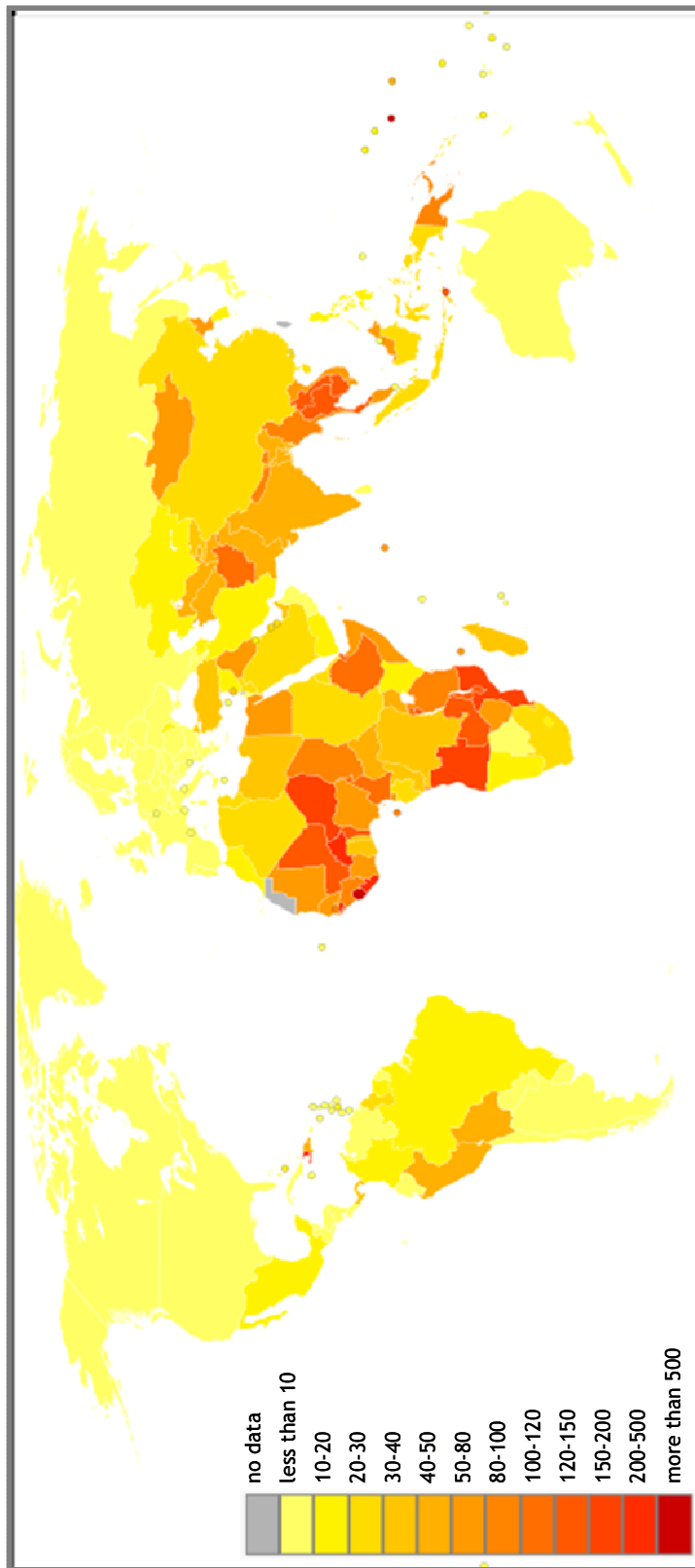
1.1.4 The Viral Particles

HBV virions circulate in blood as double-shelled particles, 40 to 42 nm in diameter, with an outer envelope and an inner nucleocapsid (Figure 2). A double layer phospholipid membrane from the host hepatocyte and hepatitis B surface antigens (HBsAg) constitute the outer envelope. The inner nucleocapsid has a diameter of 30 to 32 nm and consists of hepatitis B Core antigens (HBcAg), encoded by the Core gene of the viral genome. A total of 180 to 240 HBcAg monomers constitute the nucleocapsid, which adopts an icosahedral conformation. The viral genome, viral polymerase and some host proteins that participate in the viral replicative cycle are located within the nucleocapsid.

Figure 2 Conformation of the HBV circulating virions in blood.
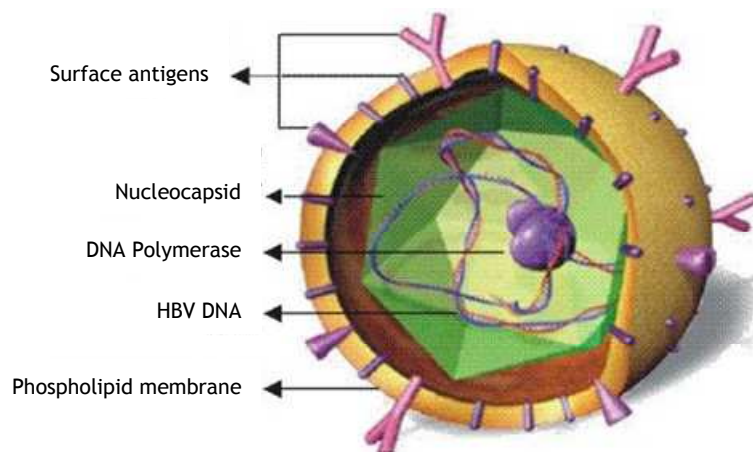
1.1.5 HBV genes and proteins

HBV has a circular DNA genome around 3.2 kb in length that is not fully double-stranded. The negative strand, which is not completely closed (its 5' and 3' ends are not covalently linked), contains all the information of the HBV genome. The positive DNA strand has variable lengths and contains an incomplete copy of the HBV genome. Figure 3 represents HBV genome organization; all the regions indicated in Figure 3 are described in Table 1.

Within the nucleocapsid, the DNA adopts a relaxed circular conformation (rcDNA) that is maintained by the double base pairing of both DNA strands. In addition, the viral polymerase is covalently linked to the 5' end of the rcDNA negative strand, while a short ribonucleotide is

linked to the 5' end of the positive strand. However, inside the nucleus of the hepatocyte, HBV-DNA adopts a covalently, closed, circular DNA conformation.

Figure 3 HBV genome organization. The negative strand (-) is presented in dark grey and the positive (+) in light grey. The polymerase (P) is linked to the 5´ end of the (-) strand and the ribonucleotide (RP) to the 5´ end of the (+) strand. The four promoters are indicated in pink letters: the surface (Sp), pre-surface (preS1p), X (Xp), and Core (Cp) promoter. The two enhancers (En), the direct repeat regions (DR1 and DR2), the polyadenylation site (PolyA) and the analogue region of retroviruses (U5 LR) are also shown.

The four open reading frames are depicted in colors: surface (S, orange), polymerase (P, red), X (X, blue) and core (C, green).
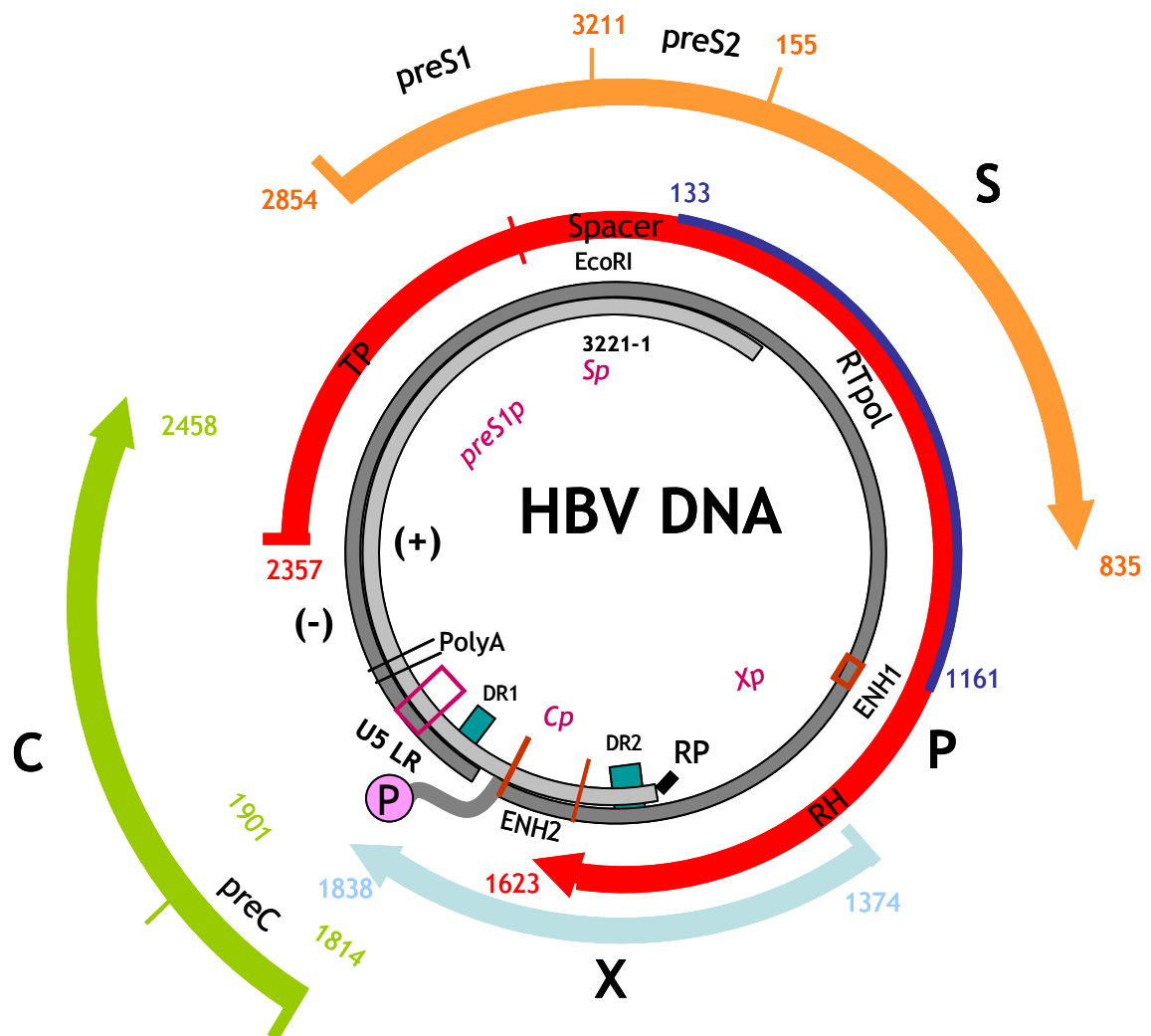
Table 1 Position of the HBV regions, described in Figure 3.

Regions	Positions in HBV genome
Core promoter	1620-1785
X promoter	1255-1390
preS1 promoter	2712-2824
Surface promoter	3152-11
Direct Repeat region 1	1824-1843
Direct Repeat region 2	1590-1600
Enhancer I	1180-1198
Enhancer II	1627-1774
Polyadenylation site	1916-1960
U5 LR	1855-1905

The HBV genome has four, highly overlapping open reading frames (ORFs): the Core, Polymerase, preSurface-Surface and X. Transcription of the four ORFs synthesizes five mRNA (Table 2): preCore-RNA, pregenomic-RNA (pgRNA), preS1-RNA, preS2/S-RNA, and X-RNA. These five transcripts code for seven viral proteins: hepatits B "e" antigen (HBeAg), polymerase, HBcAg, the three types of surface antigens (SHBs, MHBs, and LHBs), and the x antigen (HBxAg) (Table 2).

Table 2 Transcripts and viral antigens synthesized from the HBV genome

Transcripts	Length	Viral Proteins	Molecular Weight
preCore	3.5 kb	HBeAg	16 kDa
pgRNA	3.5 kb	HBcAg	21 kDa
		Polymerase	90 kDa
preS1	2.4 kb	LHBs	39kDa
preS2/S	2.1 kb	MHBs	33 kDa
		SHBs	24kDa
x	0.9 kb	HBxAg	16.5 kDa

pgRNA and preCore RNA are the largest transcripts and contain a redundant sequence at the 3' end. This sequence corresponds to the region from transcription initiation of all the mRNAs (nt position 1789 of the HBV genome) up to the polyadenylation site (nt position 1916-1960). The pgRNA redundant sequence coincides with the encapsidation signal, a secondary structure repeated at both 5' and 3' that is essential for viral replication (7) and mainly corresponds to the preCore region. A detailed description of the encapsidation signal is presented in section 1.1.7.1, Encapsidation Signal.

1.1.5.1 The surface gene

The surface ORF (nt positions 2848-835) codes for the three types of surface antigens comprising the outer envelope: small (SHBs, 24 kDa), medium-sized (MHBs, 33 kDa) and large (LHBs, 39 kDa). LHBs is translated from position 2848 and MHBs, from position 3205. SHBs is translated from position 155 and is the most abundant type. The three HBsAg types share the complete SHBs sequence and the common region from amino acid 124 to 147 contains the main antigenic loop, also called the "a" determinant. This antigenic loop activates the primary response of neutralizing antibodies in HBV infection.

Inside the preS-S region and between positions 2950 and 3125 of the HBV genome, there is a highly variable sequence that has been used to distinguish four major HBV serotypes (adr, adw, ayr, ayw).

HBsAg can self-assemble without the HBV genome, adopting structural spheres or filamentous forms. These non-infective structures account for an important percentage of the population of viral particles (>99.9%) (8), and can be detected in HBV-infected serum by electron microscopy (Figure 4). The infective particles that contain the HBV genome, known as Dane particles, are observed in a lower percentage (1:1000 to 1:10000) in serum of infected patients (8).

Figure 4 Electron microscopy image of serum of a chronic HBV infected patient (9).

Dane particles

Sphere forms

Filamentous forms
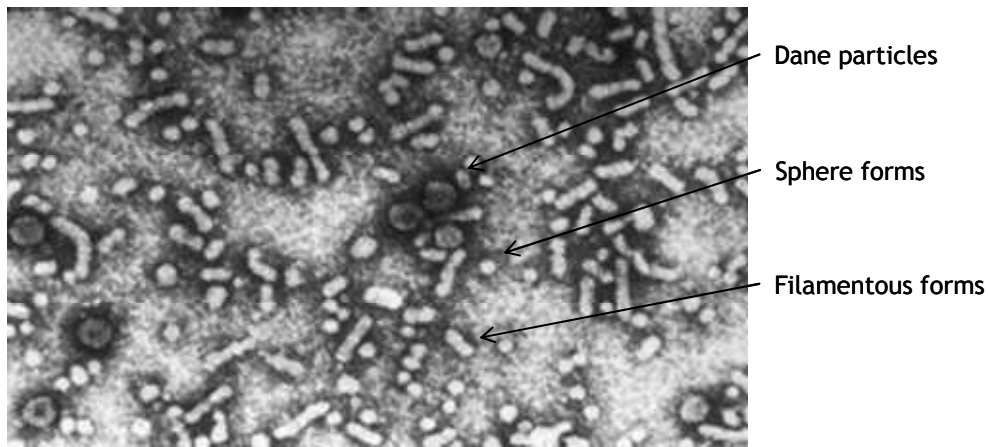
1.1.5.2 The preCore and Core gene

The preCore/Core region (nt positions 1814-2452) contains two initiation codons in positions 1814 and 1901. The region between these codons is known as the preCore, whereas the region from nucleotide position 1901 to 2452 is the Core region. If transcription starts from the first initiation codon (position 1814), the synthesized peptide (known as p25) contains both the preCore and Core region and has a molecular weight of 25 kDa (Figure 5). The preCore region has a signal sequence that leads p25 to a secretion pathway, and cleavage of the first 19 amino acids at the N-terminal region and the last 34 amino acids at the C-terminal region is produced. The resulting protein is HBeAg (17 kDa), which is released from the infected cell together with HBV particles.

The function of HBeAg has not been completely defined, but it is mainly believed to act as an immune tolerogen and to contribute to establishment of HBV infection (10). Interestingly, HBeAg is not detected in all chronically HBV-infected patients with viral activity, and it is a central parameter for clinical classification of chronic hepatitis B patients. Absence of HBeAg seems to be associated with selection of HBV preCore variants that abolish or significantly decrease HBeAg synthesis. Among all the preCore mutations that inhibit HBeAg expression (Figure 5), the most common in the Mediterranean area is a substitution of G to A at position 1896 (G1896A), which induces a change in codon 28 from a tryptophan (TGG) to a stop codon (TAG) (11-14). The minor G1897A change has similar consequences in codon 28 (TGG to TGA)

(15). Moreover, the G1896A change is often detected together with a G to A substitution in position 1899 (A1899G).

Additional changes in the first (A1814C, T1815, G1816A) and second codon (C1817A) of the preCore region result in HBeAg non-expression, due to changes in the ATG start codon or conversion of glutamine (CAA) to a stop codon (TAA) in the second codon. In addition, mutations in the basic core promoter, mainly A1762T and G1764A, are commonly related with HBeAg non-detection. These changes produce a strong decrease in preCore mRNA expression (13,14).

Figure 5 HBeAg and the HBcAg are products of the Core gene. Some of the most common mutations that abolish HBeAg expression described in the preCore region are indicated in the lower part or the figure.
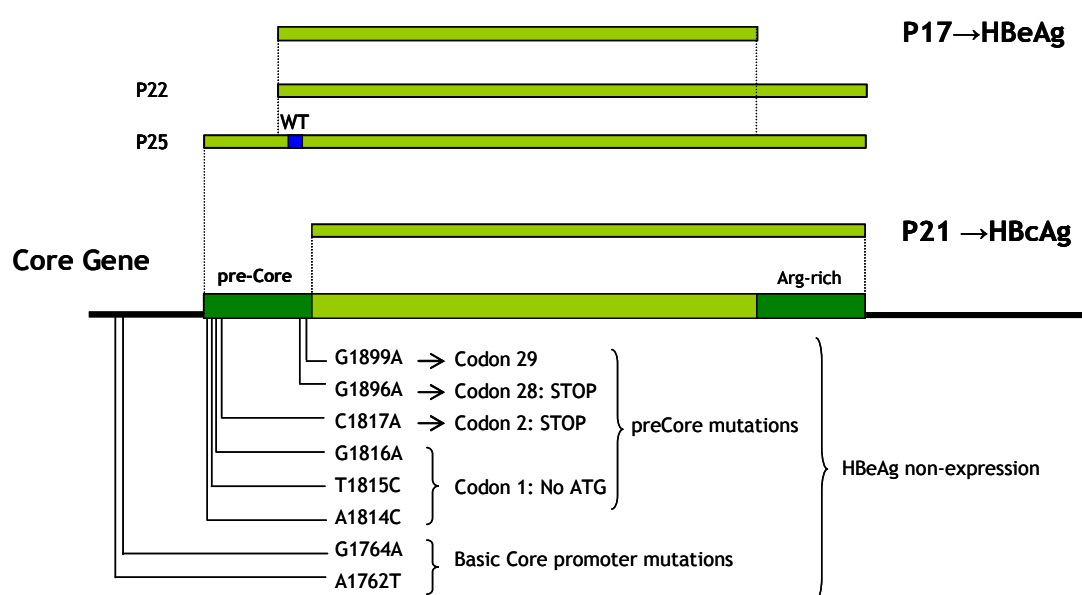
If transcription is initiated at the second initiation codon from the preCore/Core region (position 1901), the viral capsid (HBcAg, 21 kDa) is coded. The Core contains two domains, the assembly domain (AA 1 to 149) and the protamine domain (AA 150 to 183). The latter contains an arginine-rich region responsible for RNA packaging (16).

The core monomers adopt an alpha conformation (Figure 6). Two monomers associate by intermolecular disulfide bridges to form compact dimers (Figure 6). HBcAg contains four

cysteine (Cys) residues at positions 48, 61, 107, and 183. Cys61 is always implicated in the inter-chain disulfide bonds and Cys48 is partly implicated. Both Cys61 and Cys48 interact with identical residues of another monomer. Cys107 is a free thiol buried within the particle structure, whereas Cys183 participates in a disulfide bond with Cys183 of the HBcAg partner. The dimers have a 12- to 15-Å diameter pore that is proposed to be an essential structure to enable access to the nucleotides within the capsid for viral replication (17,18).

Figure 6 Structure of HBV Core monomers (a), association of two monomers to make a dimer (b), and conformation of the capsid (c) (18).
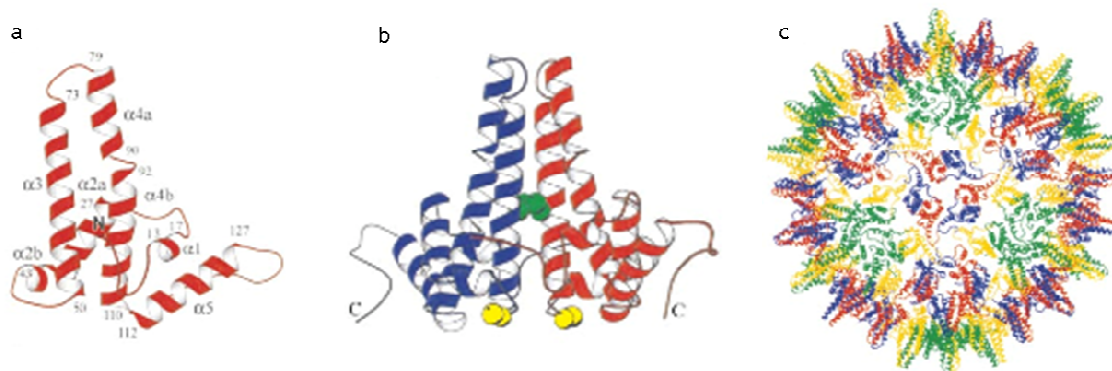
HBcAg and HBeAg share most of their sequences, and are highly cross-reactive at the T-cell level. Nevertheless, B cells recognize HBcAg and HBeAg differently, and the immune response against them appears to be independently regulated (19). The HBcAg immune response seems to be greater than the response against HBeAg or HBsAg (20). Detection of different types of antibodies against HBcAg (anti-HBcAg) in different phases of the infection seems to indicate that HBcAg is the most immunogenic of all the HBV antigens. Anti-HBcAg IgG persists for many years and has slowly decreasing titers, whereas anti-HBcAg IgM appears early in acute infection, and its titers fluctuate with clinical reactivation and strongly correlate with hepatocellular injury (21).

Several intrinsic characteristics of HBcAg might be the cause of its high immunogenicity: unique tridimensional folding, the presence of a region that interacts with immunoglobulins outside the classic antibody binding site, the presence of many CD4+ T-cell epitopes, and the presence of encapsulated nucleic acids (20). These properties, together with findings from immunogenic studies have led to the definition of different regions (Table 3) able to stimulate B-cells, T-helper lymphocytes (Th), and cytotoxic T-lymphocytes (CTLs) (22-25).

Table 3 Amino acid positions of the Core gene that determine epitopic regions

	B cells	Th cells restricted to HLA II	CTL cells restricted to HLA I
Amino acid regions	**74-84** 126-133	1-20 28-47 **50-69** 72-90 81-105 90-99 108-122 111-125 117-131 120-139 126-146 141-165	18-27 88-96 107-115 117-125 141-151

The regions in bold have a particular immunodominant role (20).

Most studies are focused on two main epitopic regions, Th50-69 and B74-84, whose localization in the Core monomer is indicated in Figure 7. Some additional HBV-specific HLA-class II restricted CD4+ cell responses have been characterized, mainly in patients with self-limited acute hepatitis (24,26,27). Th CD4+ cells of patients with self-limited hepatitis target multiple epitopes within the nucleocapsid protein. In this sense, the sequence covering the amino acid 50 to 69 region can stimulate Th cells in 90% of patients tested, irrespective of the HLA-class II profile (24). Importantly, two immunodominant T cell epitopes covering amino acid 1 to 20 and 50 to 69 within HBcAg were found to induce a significant CD4+ Th-cell response, independently of the HLA-class II haplotype (26). In addition to these regions, other minor Th-cell epitopes have been defined (Table 3) (24,28). Epitopic regions in HBcAg that stimulate CD8+ T-cells (CTL) have also been described, but all of them are restricted to HLA genotype; for example, the epitope from amino acid 18 to 27 (Th18-27) is an HLA-A2-restricted site. The major B recognition site of the Core gene is located at amino acid positions 74 to 84 (B74-84) (25), and this region also contains the main immunodominant region (MIR, AA 78-82), found in the spike of the HBc tridimensional structure (Figure 7).

In patients with chronic hepatitis, HBV strains with point mutations or large deletions have been detected in the Core gene, and most are localized in the regions described as epitopic

(25,29). Rigorous longitudinal analyses of these changes are lacking, and it is not clear whether they are a cause of severe liver damage or subsequently modulate immune response and viral clearance (29).

Figure 7 Secondary structure of the Core monomer. The main alpha conformation domains that adopt the Core monomer are highlighted in green. Th50-69 and B74-84 are also indicated.
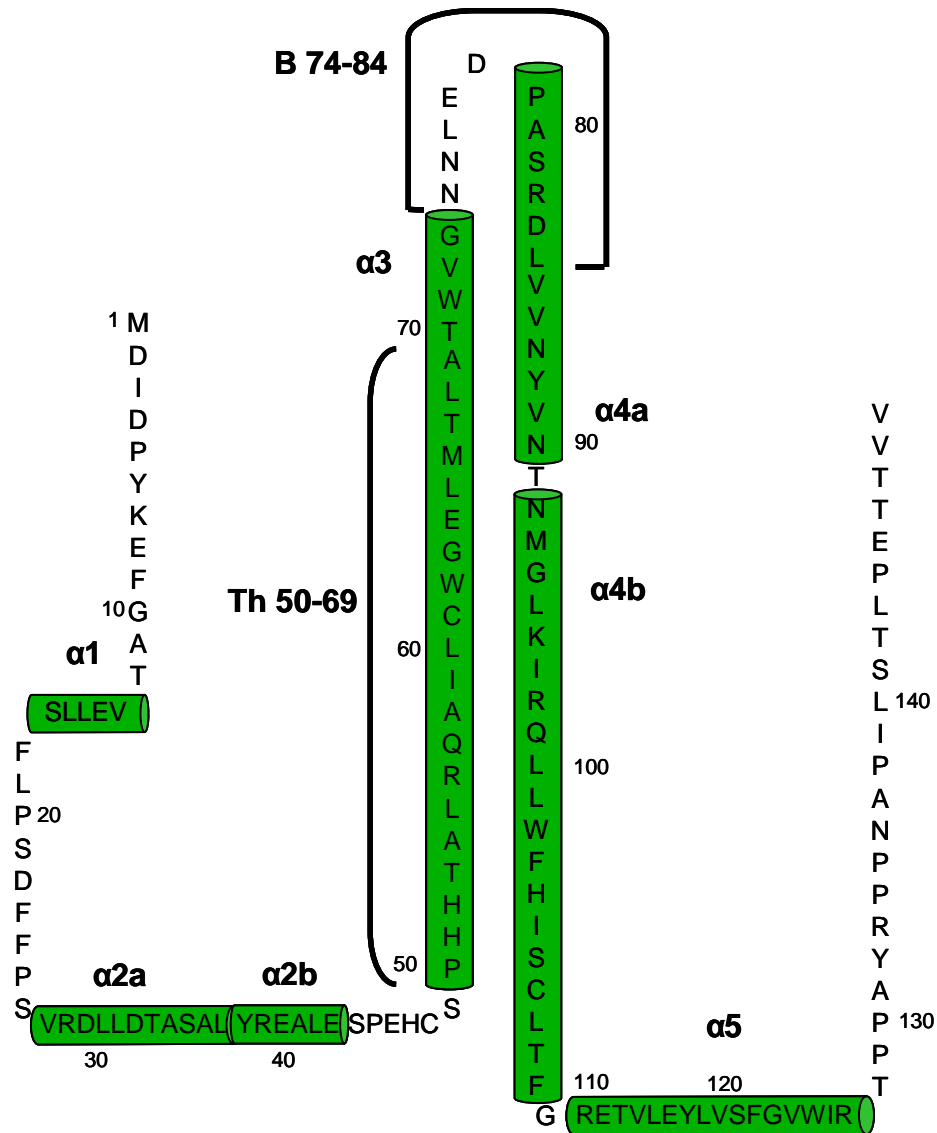
1.1.5.3 The polymerase gene

The P ORF (nt positions 2357-1623), the largest HBV gene, encodes a unique 90-kDa protein, the viral polymerase. This protein is a multifunctional enzyme involved in DNA synthesis and pgRNA retrotranscription. It contains four domains: the terminal protein (TP), a spacer region, the polymerase, and RNaseH. The TP is located at the N-terminal end and acts as a primer for synthesis of the negative DNA strand. A spacer region located next to the TP is described to be dispensable for enzyme function; this region easily tolerates mutations (30). The RNaseH domain is located at the C-terminal of the polymerase gene and degrades pgRNA during the replication cycle. Between RNaseH and the spacer region is the proper viral polymerase, with both DNA polymerase and retrotranscriptase (RT) activities. The RT retrotranscribes pgRNA to a negative DNA strand and polymerase activity synthesizes the incomplete positive DNA from the negative strand.

RT activity can copy pgRNA to DNA, but approximately 1.4 to 7.9 x 10^{-5} substitutions/site/year are generated in the process (29). The catalytic center of RT activity (nt position 736-747) is mapped at AA positions 551 to 557 of the HBV genome, which corresponds to the tyrosine-methionine–aspartate–aspartate (YMDD) sequence. The YMDD motif contains two of the three essential aspartate residues of the polymerase. Attending to the specific RT region nomenclature, position 348 of the polymerase gene corresponds to the first AA of the RT; hence, the catalytic motif is located at AA position 203 to 206 ($Y_{203}M_{204}D_{205}D_{206}$) (31).

The RT region is highly conserved in retroviruses and Hepadnaviruses, and the YMDD catalytic motif is identical to the motif observed in HIV-1 (32). Neither the HBV polymerase nor its RT domain has been crystalized. However, the HBV RT tridimensional structure has been modelled from the available crystal structures of HIV-1 RT (Figure 8) (33,34). Regardless of their amino acid sequences and differences in the structural domain, both polymerases appear to have a common "right-hand" configuration with thumb, palm, and fingers domains (Figure 8) (35,36). The palm domain, which includes the YMDD catalytic motif, appears to be the active site and catalyses the addition of nucleotides in nascent HBV DNA (phosphoryl transfer reaction). The fingers domain facilitates interactions with the incoming natural nucleotides and the template base to which it is paired; and the thumb domain may play a role in positioning the duplex DNA, processivity, and translocation (35). Nucleoside analogues and host nucleotides bind at a site located in the palm subdomain, adjacent to the 3' terminus of the nascent strand.

Figure 8 Tridimensional structure of the HBV retrotranscriptase domain (33).
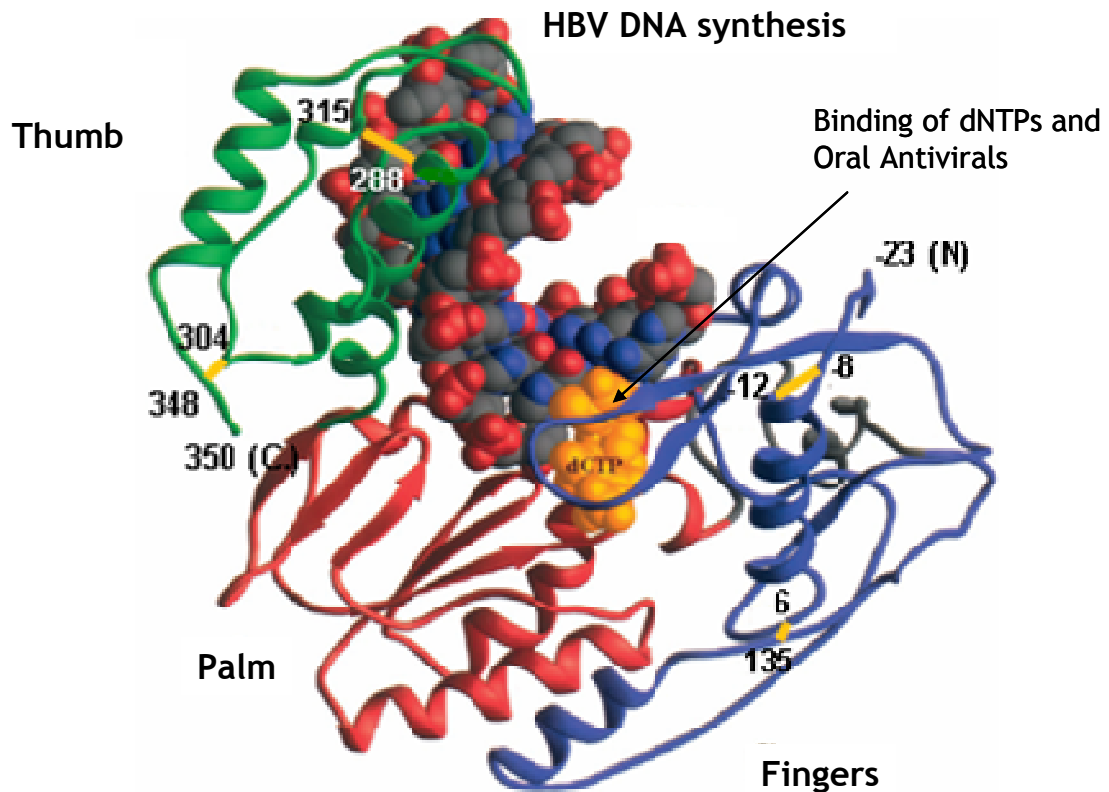
Oral antiviral drugs, also referred to as nucleos(t)ide analogues (NUCs), bind as natural polymerase substrates, and act as competitive inhibitors by blocking elongation of the new HBV-DNA strand. Incorporation of NUCs in the nascent strand results in an absence of available 3'OH-free ends, preventing replication. These drugs induce selection of HBV variants carrying amino acid substitutions in the RT domain, which can cause changes in the tertiary structure of the enzyme and result in a reduction in the affinity of the drug, decreasing its antiviral activity. In consequence, NUCs will efficiently inhibit wild type HBV variants present in the viral quasispecies, while variants carrying resistant mutations will maintain their replicative activity. The percentage of resistant variants in the quasispecies may increase under NUC treatment and ultimately be selected as the major variant, causing treatment failure. Thus, mutated strains have high clinical interest because they can confer resistance to oral antiviral treatments. A detailed description of the mutations selected under different antiviral treatments is described in section 1.2.3, Antiviral Treatments (37,38).

1.1.5.4 The X gene

The X ORF (position 1374-1838) codes for the 16.5-kDa HBxAg, which is reported to modulate certain signal pathways, protein degradation, cell death, and carcinogenesis (39). HBxAg can be divided into 6 domains (A-F), with different degrees of conservation. The C-terminus of the protein (domains C-E) seems to be the transactivating portion, while the N- terminus (domain A) can repress HBxAg transactivation activity (40). The N-terminus is believed to avoid excessive HBxAg transactivation and to play a role in a self-regulatory mechanism of X gene expression. It has been demonstrated that HBxAg stimulates viral replication and is an absolute requirement for *in vivo* and *in vitro* HBV replication (41).

The correlation between HBxAg and HCC development has been extensively investigated. The oncogenic roles of HBxAg include activation of a variety of transcription factors, such as nuclear factor nB, activator protein cAMP, and activating transcription factor 2. The x antigen is also involved in interactions with cellular oncogenes, such as Ras, Src, and c-jun, stimulation of cytoplasmic signal transduction pathways (Ras-Raf), and participation in cell stress–induced pathways (39).

1.1.6 Viral replication cycle

HBV circulates in blood as a double-shelled particle (Figure 2) and replicates within the hepatocyte, as is described in Figure 9. preS1-HBsAg interacts with a hepatocyte receptor and allows entrance of the viral particle inside the liver cell, forming an endosome (42) (Figure 9a). Within the endosome, the outer envelope is released and the nucleocapsid remains in the cytoplasm. The HBcAg dimers contain a nuclear localization signal. Thus, the capsid interacts with the nuclear membrane and releases the partial double-stranded viral genome (rcDNA) inside the nucleus (43). The polymerase linked to the 5' (+) DNA strand and the short ribonucleotide added to the 5'(-) DNA strand are then eliminated by unknown mechanisms. The viral polymerase completes synthesis of a positive strand and the HBV genome finally adopts the conformation of a covalently, closed, circular DNA (cccDNA) (Figure 9b). The cccDNA serves as a template for synthesis of 5 mRNA, and acts as a minichromosome. It has been postulated that cccDNA will remain in the nucleus during the life of the infected hepatocyte (44). The life span of liver cells is estimated to be from 10 to 100 days (45). This long-lasting establishment of cccDNA in the hepatocyte nucleus is closely related to the

difficulty of eradicating HBV disease and the development of hepatocellular carcinoma (46). It must be kept in mind that viral antigen expression will continue even in the absence of HBV genome replication due to cccDNA presence. Therefore, these antigens (eg, HBsAg or even HBeAg) will be detected, and HBxAg will maintain its multiple roles over cell functions including its oncogenic risk.

The mRNAs are transported to the cytoplasm, where their translation by host ribosomes yields the viral proteins, including the three types of HBsAg (SHBs, MHBs, LHBs), HBxAg, HBeAg, polymerase, and HBcAg (47) (Figure 9c). These last two proteins are coded from pgRNA, which adopts a secondary loop structure in both the 5' and 3' redundant ends (Figure 10). This structure, which is called epsilon (ε), is recognized by the viral polymerase, and once it is associated with pgRNA, induces addition of HBcAg around the complex, driving construction of new HBV nucleocapsids (48). Inside the nascent capsid, pgRNA is retrotranscribed to initially yield the negative HBV-DNA strand (reaction primed by the OH corresponding to Tyr 63 from the terminal protein region of the polymerase) and, from the negative strand, the positive strand is synthesized. A detailed description of the pgRNA replication strategies adopted by HBV is provided in section 1.1.7.1, Encapsidation Signal.

Capsids bearing *de novo* synthesized rcDNA are identical to those coming from the blood that originally infected the cell. Therefore, some of the *de novo* capsids can reinfect the cell nucleus (capsid recycling process), thereby increasing the nuclear cccDNA pool and cccDNA viral variability. The variability of this intrahepatic form might be increased due to viral polymerase errors produced during retrotranscription and other potentially DNA-editing steps, such as activity of apolipoprotein B RNA-editing catalytic polypeptide-like 3 (APOBEC3) family of cytidine deaminases host enzymes, recently described to act upon HBV (49). However, most of the *de novo* capsids are directed to the endoplasmic reticulum, where they acquire the lipoprotein envelope containing HBsAg, yielding new viral particles that are exported from the hepatocyte (Figure 9d).

Figure 9 Steps in the HBV replication cycle (a) Entry of the HBV particle inside the hepatocyte, by an endosome. After releasing the outer envelope, the nucleocapsid is directed to the nucleus
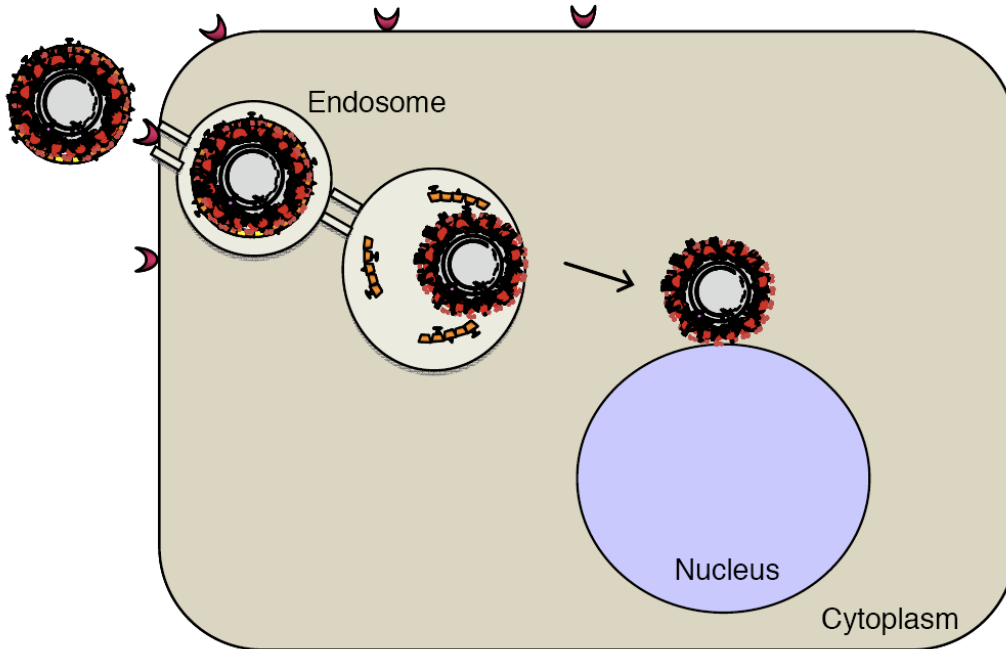
b) Synthesis of cccDNA, which serves as a template for synthesis of the five mRNAs.
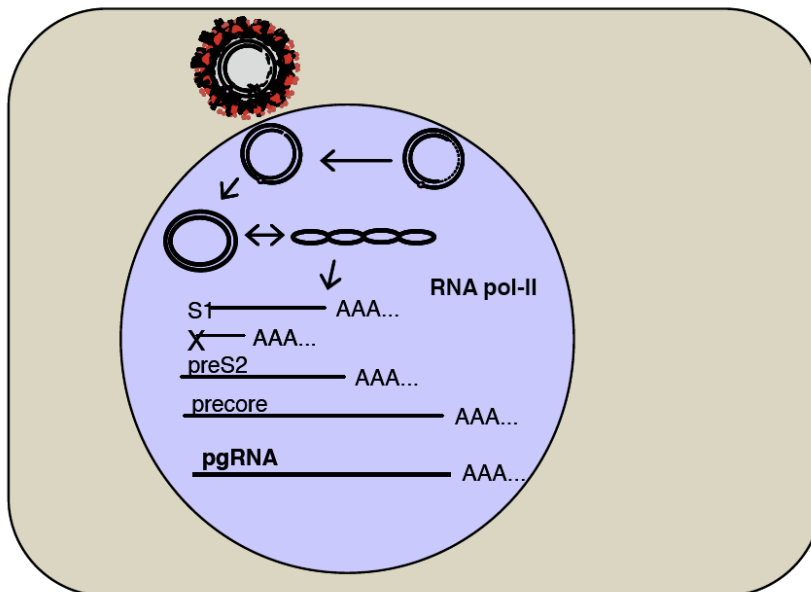
c) Production of seven viral antigens; the polymerase and the HBcAg interact with pgRNA and a new nucleocapsid is synthesized.
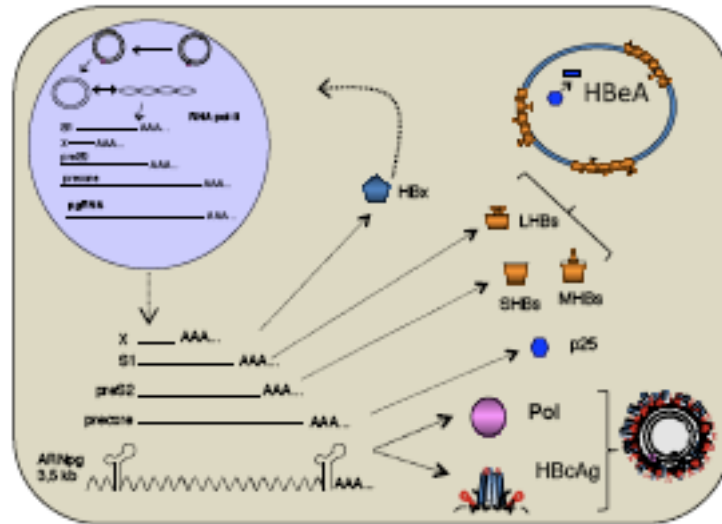
d) The new nucleocapsid can reinfect the nucleus or be exported from the hepatocyte.
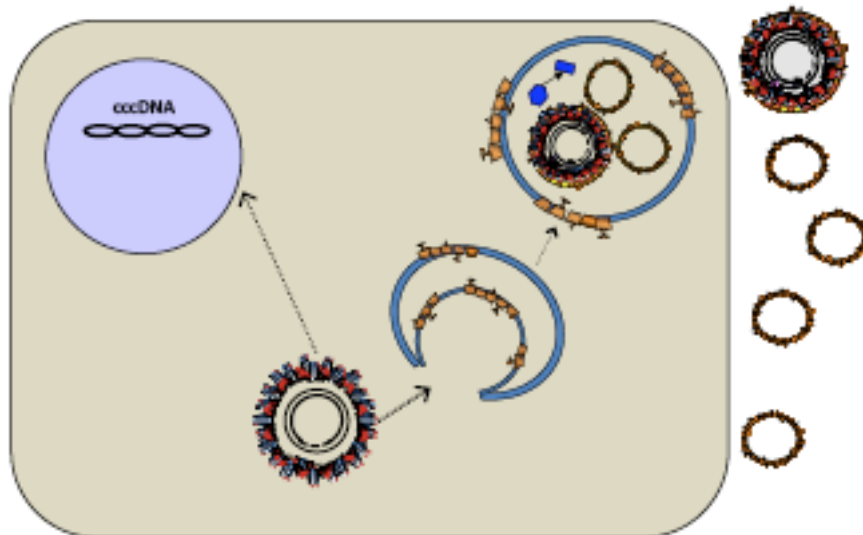
Figure 10 Adopted structure of pgRNA at the 5′(a) and 3′(b) ends.

(a) The 5' end comprises two stems, the basal and apical (in green) and two bulges, the inner bulge and the apical loop (in orange). The annealing region of the 4-nucleotide primer (4NT primer, in blue) is also indicated. The presence of an OH group in the tyrosine at position 63 (Tyr63) of the terminal protein of the polymerase (TP of Pol) enables synthesis of the 4NT primer.
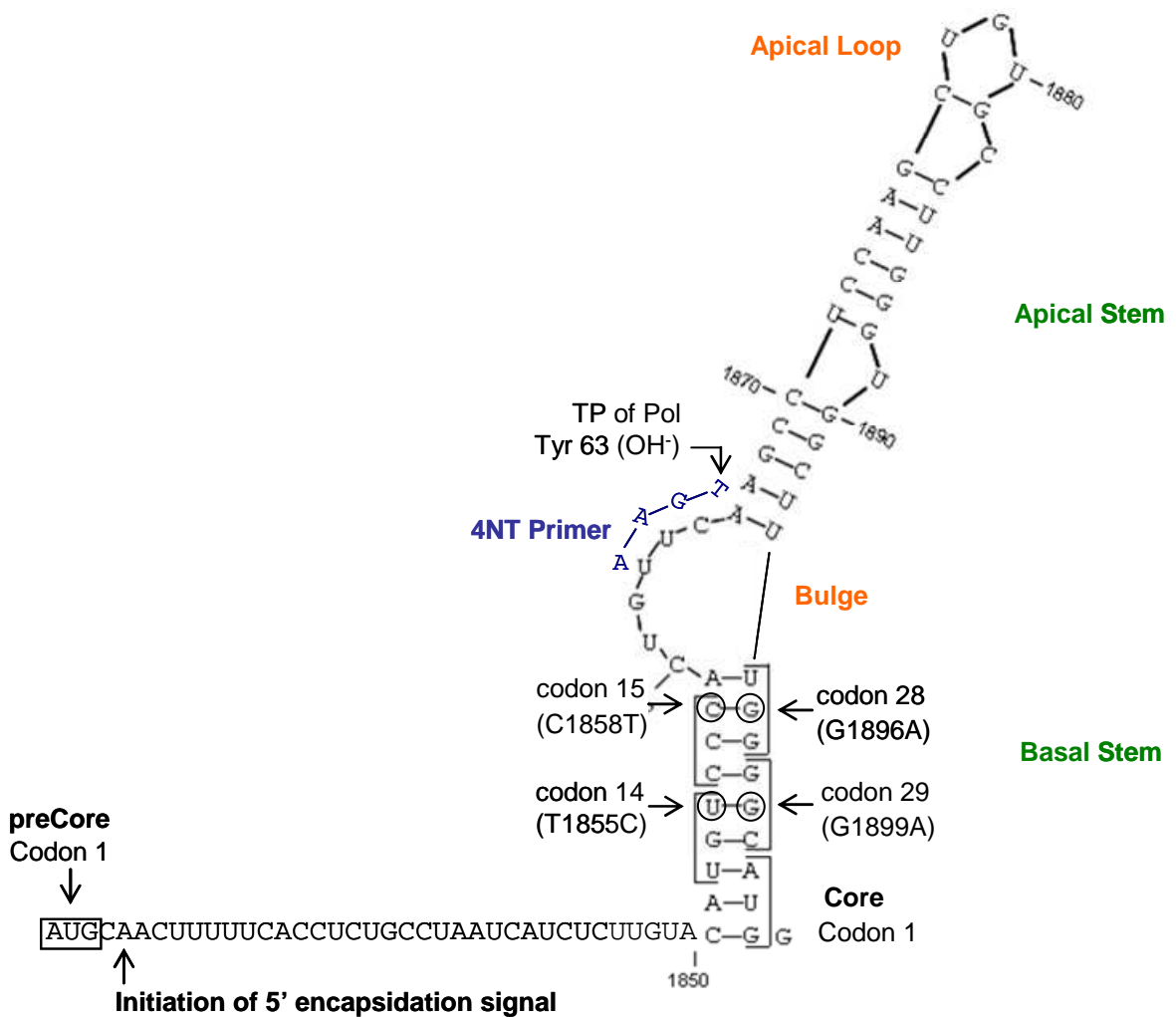
(b) The 3´end of the encapsidation signal contains direct repeat region 1 (DR1), which includes the acceptor site (AS) region. The four TATA-like boxes are highlighted in orange.
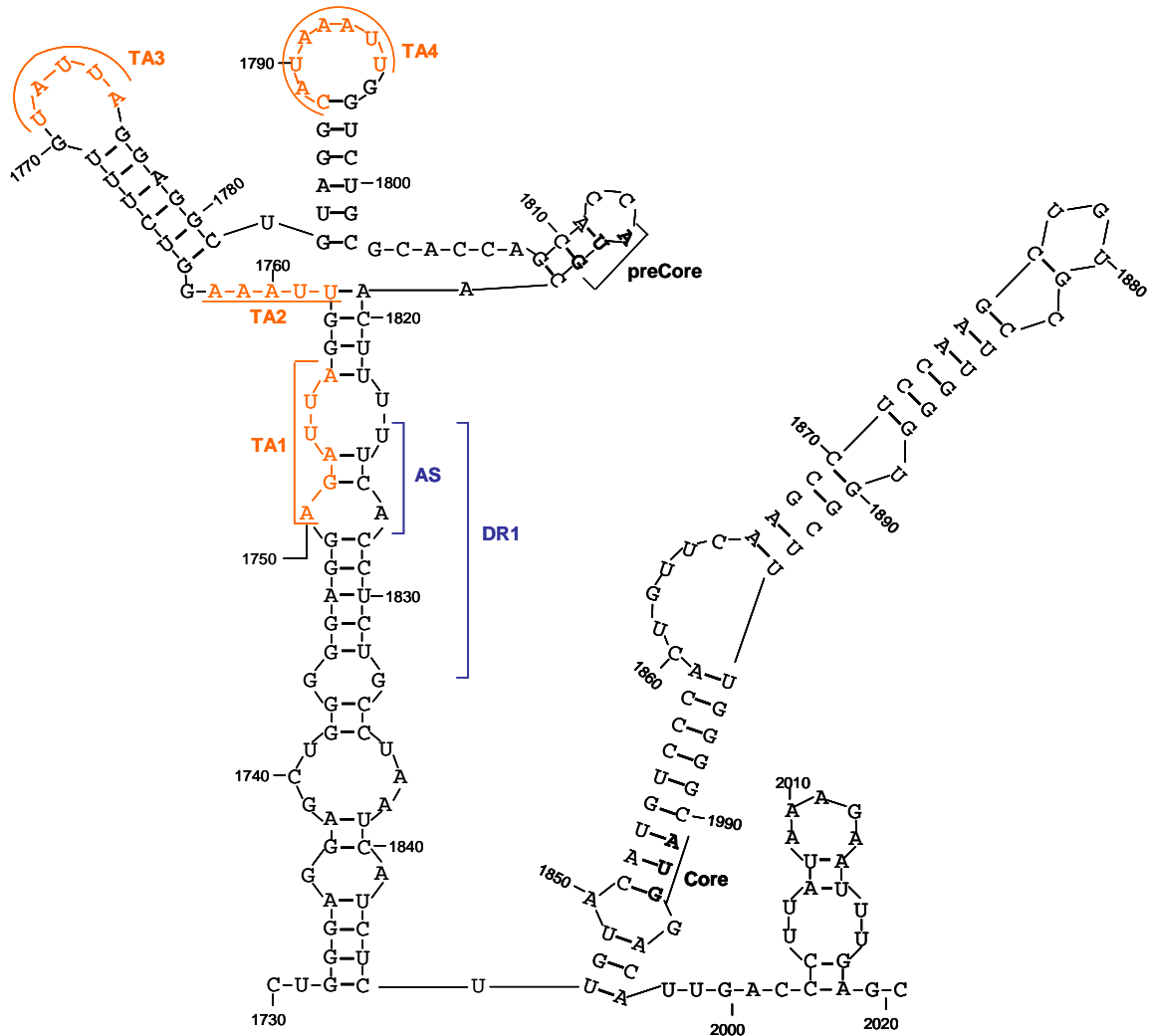
1.1.7 Regulatory elements of the HBV genome

cccDNA serves as a template for unidirectional synthesis of the seven HBV transcripts. All of them are modified by addition of a 5' cap at each promoter site and by a 3' polyadenylation between positions 1916 and 1960, which is essential for the ending at the same position of the transcripts. Four promoters initiate transcription: preCore/pg, S1, S2, and X. In addition, two enhancers (EN-1 and EN-II) as well as *cis*-acting negative regulatory elements play important roles in regulation of viral gene transcription (50). In the HBV genome, there are also two short, highly conserved 11-bp sequences involved in viral replication, defined as direct repeat regions (DR). DR1 (UUCACCUCUG, positions 1824-1834) is located at the 5' end of the negative

strand and DR2 (UUCACCUCUGC, nt position 1590-1600) at the 3' end of the positive strand (50).

pgRNA and preCore RNA are both transcribed from the same HBV region. These two transcripts have different functions and are subject to a combination of coordinate and differential transcriptional regulation. The Core promoter (nt positions 1591-1822) initiates transcription of both preCore RNA and pgRNA, and consists of the basal core promoter (BCP, nt positions 1742-1849), and an upstream regulatory region (nt positions 1636-1742). BCP directs precise initiation of the preCore and pgRNA transcripts. BCP lacks canonical TATA boxes, but contains four TA-rich regions (TATA-like sequences, indicated in Figure 10b). Three types of TATA-like boxes are required to control transcription of preCore RNAs: TA1 (nt positions 1750-1755), TA2 (nt positions 1758-1762) and TA3 (nt positions 1771-1775). TA4 (nt positions 1788-1795) controls transcription of pgRNA and can regulate preCore RNA transcripts (51). The regulatory sequences upstream of the BCP contain *cis*-elements that may positively modulate promoter activity, referred to as core upstream regulatory sequences (CURS, nt positions 1636-1742) and negative regulatory element (nt positions 1613-1636).

1.1.7.1 Encapsidation signal

pgRNA contains identical 143-bp sequences (nucleotide 1818 to 1960) at both the 5' and 3' ends (Figure 10) that act as *cis*-acting motifs. These redundant sequences adopt a hairpin stem-loop structure named epsilon (ϵ). Secondary structure analysis (2D) has provided a description of the base-pairing pattern in the structure (52), and nuclear magnetic resonance analysis has confirmed the structure of the upper stem (52-54). The ϵ is also highly conserved in other mammalian Hepadnaviruses, as well as HBV isolates, indicating the importance of these structures (55).

The 5' end ϵ signal contains two stems, basal and apical, and two bulges. Base pairing characterizes the stems, while the apical loop has been recently described to contain 3 nucleotides, as is indicated in Figure 10a (54). Thermodynamic calculations in this region indicate that the entire signal is highly stable. The high structural stability induces strong affinity of the polymerase to recognize the structure, and increases viral replication (55). For this reason, most changes selected in this region maintain base pairing and the energy of stabilization.

The preCore variants described to abolish HBeAg expression are located in the ε signal (see section 1.1.5.2, The preCore and Core Gene). The mutants abolishing HBeAg synthesis found in nature also maintain a stable ε secondary structure. For example, the main preCore mutation, G1896A, increases the loop stability in genotype D. This genotype has a characteristic polymorphism with a T in position 1858; in consequence, a change to A in position 1896 increases stability of the ε signal. For this reason, selection of the main preCore mutation is more common in genotype D than in genotype A, in which the polymorphism in position 1858 is a C (Figure 10a) (12,55).

The ε signal at the 5' end directly interacts with the polymerase and constitutes the first step in the initiation of reverse transcription. This interaction induces recruitment of 180 to 240 HBcAg monomers around the polymerase-pgRNA complex (56). The interaction of Pol at the 5' end of pgRNA induces synthesis of the (-) DNA strand. Initially, the polymerase anneals to the bulge of ε at the 5' end and synthesizes a sequence of four nucleotides (4NT, Figure 11a). The terminal protein of the polymerase contains tyrosine at position 63; presence of the OH group of this amino acid enables synthesis of the 4NT primer (Figure 10b).

The 4NT primer translocates to the DR1 sequence located in the 3' end of pgRNA (Figure 11b) and acts as a primer for synthesizing the (-) DNA strand by retrotranscriptase activity of the polymerase (Figure 11c). While (-) DNA synthesis occurs, RNAaseH activity of the polymerase degrades pgRNA. However, a short RNA sequence of approximately 15 NT is not degraded and translocates to the DR2 located at the 5' end of the (-) DNA strand (Figure 11d). 15NT serves as the primer for synthesis of the (+) DNA strand, by polymerase activity (Figure 11e). The nascent (+) DNA fragment also translocates to the 3' end of the negative strand (Figure 11f), and synthesis of the (+) DNA continues (Figure 11g). However, this last step is limited to the moment the capsid is released, when nucleotides become unavailable for replication. For this reason, during capsid maturation, the positive strand remains incomplete, and the HBV genome adopts a partially double-stranded DNA (rcDNA) conformation (Figure 11h).

Figure 11 pgRNA Retrotranscription Steps.

a) The polymerase anneals to the bulgue of ε

Synthesis of the
primer (4NT)

Pg RNA (3.5 kb)   5'   3'

DR1   DR2   DR1
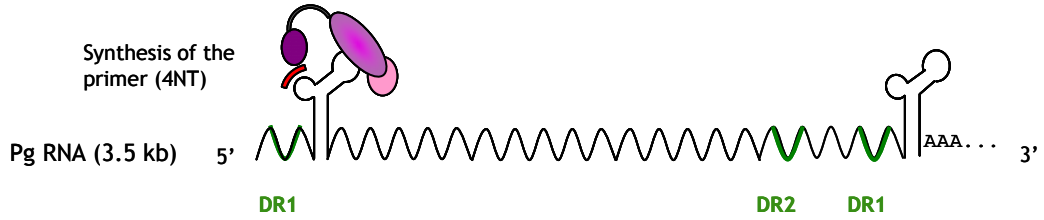
b) Translocation of the 4NT primer and the polymerase to the DR1 sequence

Synthesis of the (-) DNA

Pg RNA (3.5 kb)   5'   AAA...   3'

DR1   DR2   DR1
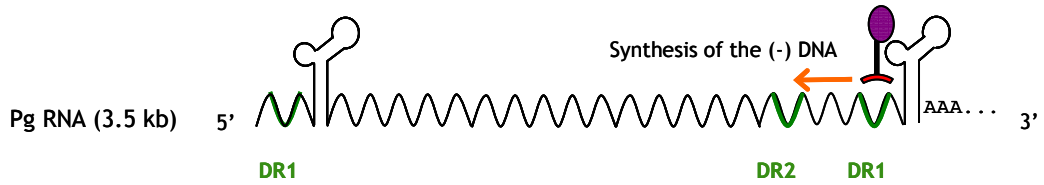
c) (-) DNA synthesis the Retrotranscriptase activity

5'   Degradation of the RNA
strain by the RNAseH activity

(-) DNA   3'   5'

DR1   DR2   DR1
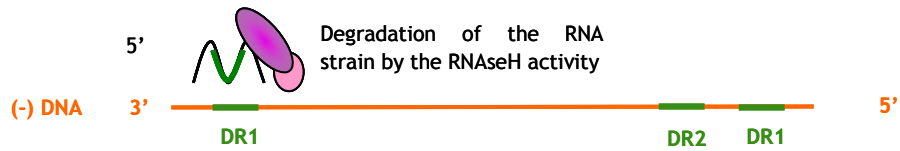
d) Translocation of the complex Pol+RNA to the DR2 sequence at 5' end of the (-) DNA

(-) DNA   3'   5'

DR1   DR2   DR1

e) Initiation of the (+) DNA synthesis

3'

(-) DNA   3'   5'

DR1   DR2   DR1

f) Translocation of the (+) DNA to the 3' end of the (-) strand

3'

(-) DNA   **3'**   DR1   DR2   DR1   **5'**

g) Synthesis of the (+) DNA

5'

(-) DNA   **3'**   DR1   DR2   DR1   **5'**
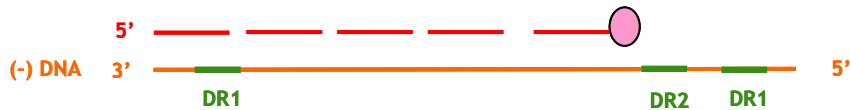
h) rc DNA genome conformation (3.2-kb)

1.1.8 Viral variability: Quasispecies structure of HBV

The HBV retrotranscription process includes the retrotranscriptase activity of the viral polymerase. This enzyme is similar to other reverse transcriptases or viral RNA polymerases that have no proofreading activity, such as HIV or HCV. Therefore, mutations can accumulate throughout the genome. However, the high mutation rates ($1.4\text{-}7.9\text{x}10^{-5}$ substitutions/site/year) [57] and replication ($5\text{x}10^{12}$ virions/day in HBeAg-positive and $6\text{x}10^{10}$ virions per day in HBeAg- negative infection) [58] are constrained by the restrictions imposed by the highly overlapped HBV genes: there is a high possibility that one nucleotide substitution will affect different ORFs. That is, a mutation in an overlapped region might alter the functionality of two different encoded proteins. The two different evolutionary trends observed in HBV infection yield to a complex viral composition called quasispecies, which means that in an HBV-infected patient, the virus circulates as a complex population of different, but closely related, viral genomes [59]. Recently, an increase in HBV complexity has been attributed to the effect of innate antiviral defense mechanisms mediated by the activity of host enzymes belonging to APOBEC3 family of cytidine deaminases. These enzymes are able to cause extensive deamination of cytidine bases to uridine in the negative DNA strand,

resulting in a G to A hypermutation in the positive strand. G to A hypermutations have been identified in HBV UDPS reads (60). This activity was first reported in HIV (61), but it has also been demonstrated in HBV (49) and other retroviruses and retrotransposons (62).

The quasispecies structure confers a relevant plasticity to viruses, allowing fast and easy adaptation to changes in the viral environment. Host immune pressure and exposure to antiviral treatment are actually evolutionary factors for these viruses. The selection of variants that confer higher adaptability to new environmental conditions results in remodeling of the viral population composition. Therefore, members of the quasispecies present at baseline infection (eg, before antiviral treatment or vaccination) will be selected according to evolutionary pressure, such as that occurring from the immune system or antiviral treatment (63) (Figure 12).

In the complex composition of the HBV quasispecies, some variants with specific mutations are especially relevant to the clinical course of HBV infection. For example, antiviral treatment might induce selection of members of the viral population that carry amino acid changes conferring resistance to therapy. Most of these mutations might be located in the viral polymerase. The quasispecies members that survive under antiviral treatment, labeled as antiviral-resistant variants, maintain replication despite the presence of an antiviral drug. Thus, selection of variants conferring resistance to oral antiviral treatment has clinical implications on the course of HBV infection (37,38). In addition, variants in the surface ORF that induce changes in the antigenic loop have been associated with vaccination and prophylaxis escape mechanisms (64).

However, mutations in the preCore and Core promoter regions are the most common naturally occurring changes in the HBV genome (29). HBV genes are highly overlapped, but the region showing the least overlapping is the PreCore/Core region. Specifically, only 24 bp of the preCore 5' end (position 1814 to 1838) overlap the X ORF, and 95 bp of the 3' end of the preCore/Core region (2357-2452) overlap the polymerase (Figure 3). Thus, the region from position 1839 to 2358 can accept mutations without affecting any other gene; indeed, the HBV Core gene (positions 1901 to 2452) can concentrate most of the mutations selected in the HBV genome. As was described in section 1.1.5.2., most Core gene changes are clustered in epitopic regions (22,25,29,65,66), but the distribution of these mutations has not been completely defined. Their high frequency and stable phenotype imply that these variants are the result of biological selection, possibly one that enhances HBV replication or evasion of the immune

response. Mutations in the preCore region can also be selected and can inhibit HBeAg expression (very common in the Mediterranean population) (11).

As was previously discussed (section 1.1.7.1, Encapsidation signal), mutations in the PreCore region are related with abolishment of HBeAg expression and affect the secondary structure of the encapsidation signal (12). The preCore/Core region is closely related to the immune response: the preCore codes for the immunotolerogen HBeAg antigen, and the Core contains epitopes. Selection of mutations in both regions has implications on evasion of the host immune response, and ultimately, on the outcome of HBV infection.

Figure 12 Representation of the variability of the quasispecies population. Possible selection of mutations under different evolutionary pressures, such as antiviral treatment or the immune response.
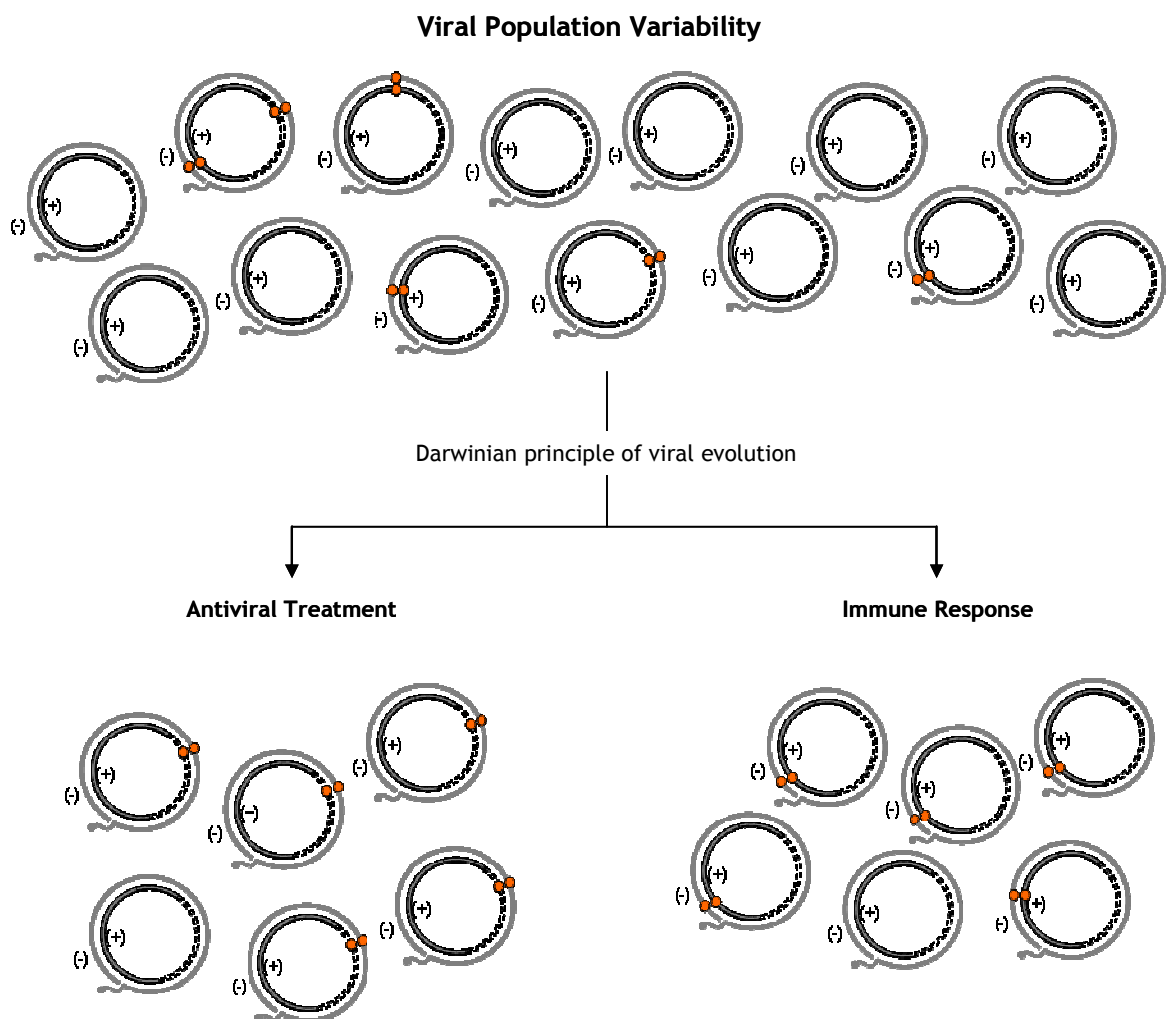
HBV variability has led to the definition of 9 (A to I) genotypes, which differ by an 8% difference in the nucleotide sequence (67). These genotypes have a defined distribution, in which genotypes A and D are the most prevalent in Europe, genotype B and C in Asia, genotype E in Africa, genotype F in Central and North America, and G in North America and Europe. In the Mediterranean basin, genotypes D and A are the most common and are present in similar proportions (genotype D, 48.1% and genotype A 39.5% of cases). As expected, genotype A is predominant in HBeAg-positive cases (68.6%) and genotype D in HBeAg-negative infections (60.7%) (68).

Genotype also has influence on the course of the infection (69). In general, it has been found that genotype C induces a more severe liver disease than B, F, D or A, in that order. On the other hand, progression to chronic hepatitis after acute infection seems more likely and faster with genotypes B and C than A, and genotype A seems to be faster than D. In addition, the antiviral response to interferon therapy is associated with HBV genotype: achievement of sustained viral response is clearly more probable in genotype A cases than genotype D (70). A detailed explanation of the relationship between HBV genotype and antiviral therapy is described in section 1.2.3, Antiviral Treatments.

1.2 Clinical Implications of HBV Infection

1.2.1 Natural history

The clinical expression of HBV is very diverse. Primary HBV infection can be symptomatic or asymptomatic. After initial exposure to HBV, some individuals experience acute infection showing ALT flares, and test positive to HBsAg and anti-HBc IgM. After 3 to 6 months, 95% of adults, but less than 10% of children, clear HBsAg and develop anti-HBs seroconversion. Individuals in whom infection does not resolve (5% of adults and around 90% of children) develop chronic liver disease (Figure 13). In such cases, viral replication continues in the liver, although the titers of the virus are variable. Persistent HBV infection can be symptomatic or asymptomatic. People with subclinical persistent infection, normal serum ALT levels, and normal or nearly normal findings on liver biopsy are termed inactive chronic HBV carriers. Those with abnormal liver function and histologic features are classified as having chronic hepatitis B (CHB). This last group of patients has a higher risk of developing liver-related

complicatons, such as cirrhosis, decompensated cirrhosis, and/or hepatocellular carcinoma (71).

Figure 13 Natural history of HBV infection
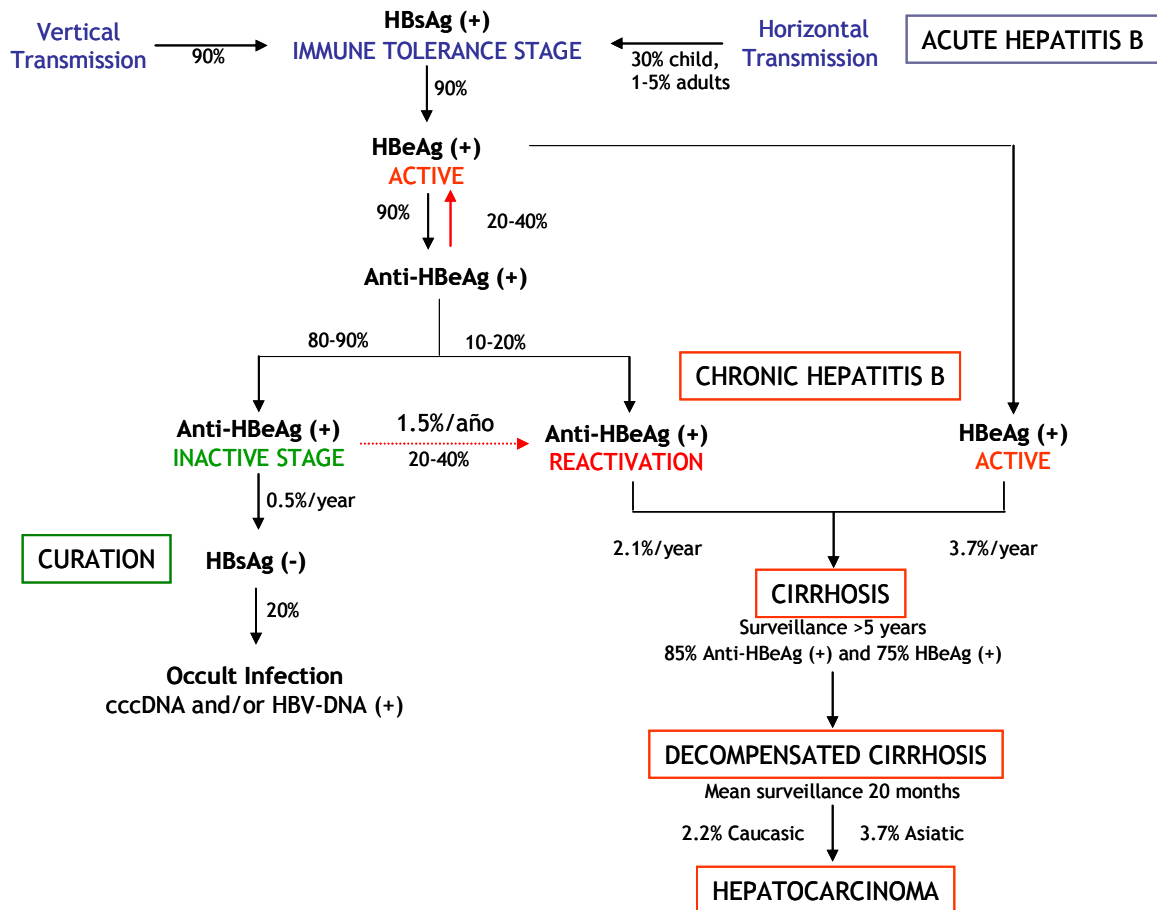
Chronic hepatitis B infection has four natural stages: the immune tolerant stage, immune clearance stage, inactive HBsAg carrier stage, and reactivation stage. However, not all CHB patients go through all the four stages. The immunotolerant stage is typical in patients with perinatal or early childhood-acquired infection. Patients are HBeAg (+) and have active viral replication, but ALT levels are normal and there is minimal or no inflammation on liver biopsy. After this stage, the condition passes to the immune clearance stage, which is characterized by maturation of the host immune system and an increase in hepatic inflammation. HBV-DNA and ALT levels are high during this phase of infection, reflecting active HBV replication. After immune clearance, some patients achieve the inactive HBsAg carrier stage, in which biochemical and virological parameters normalize and there is a lower risk of developing

cirrhosis or HCC. Some of these inactive patients experience intermittent virological activity and enter the reactivation stage (72).

1.2.2 Immune system activation in HBV infection

HBV is an obligate intracellular pathogen that primarily infects hepatocytes. The virus is not considered directly cytotoxic. However, persistent production of viral antigens due to the presence of cccDNA, even in inactive replication periods (eg, under effective antiviral therapy), may justify some cell toxicity, for instance, due to the multifunctional and potentially carcinogenic role of HBxAg.

Moreover, the presence of HBV viral antigens expressed by HLA (mainly type I, but also type II) on the surface of hepatocytes strongly activates the host immune system. The main purpose of the immune response is to protect the host from infection, and this implies clearance or elimination of the virus. Both the cellular and the humoral arms of the immune response are essential for viral clearance, but the cellular response, which results in lysis of infected hepatocytes, is thought to be the main factor responsible for disease pathogenesis and the clinical manifestations of HBV infection (73,74).

Various studies based on animal and human models have shown that after the initial inoculation, HBV does not immediately begin to replicate efficiently (75). HBV-DNA and HBV antigens are not detectable in serum or liver until 4 to 7 weeks later, and after this period, a logarithmic expansion phase begins (76,77). The elements of innate and adaptive immunity seem to be activated during this initial phase of HBV infection, as has been seen in chimpanzee models (78,79).

1.2.2.1 Early events: Innate immunity

Innate immunity generally plays a role in limiting HBV spread and initiating an efficient adaptive immune response. Innate responses in the early phases of infection are mainly characterized by production of type-1 IFN alfa/beta cytokines and activation of natural killer (NK) cells. NKs are activated by recognition of stress-induced molecules and/or modulation of the quantity of HLA-I molecules on the surface of infected hepatocytes (80). Thus, activation of the elements of innate immunity leading to production of large quantities of IFN α/β seems to be the key to activating adaptive immunity (81).

However, several characteristics of HBV infection apparently make the innate immune response ineffective. First, the peculiar kinetics of HBV replication, which shows an initial lag phase and an exponential phase at week 4 to 5 after infection (77). The initial lag phase is not a consequence of innate immune activation. As has been observed in experimental infection in woodchucks (82) and chimpanzees (79), activation of IFN-gamma, IL-2, and TNF-alpha, and intrahepatic recruitment of inflammatory cells is delayed until logarithmic expansion. Furthermore, in infected chimpanzees, activation of cell genes was not detected (78).

Another aspect of HBV that induces ineffective immune response activation is IFN alpha and beta production. These two types of IFNs can efficiently limit HBV, but data in HBV-infected chimpanzees suggest that antiviral cytokines are not triggered by HBV replication (78). Production of IFN and other cytokines depends on the presence of free DNA or double-stranded RNA molecules in cytoplasm, which are recognized by toll like receptors (TLR). However, HBV has developed a "hidden" replication mechanism to escape this initial antiviral defense: HBV replicates inside nucleocapsids, and therefore, it evades contact of pgRNA or rcDNA with TLR. Moreover, cccDNA always remain inside the hepatocyte nucleus.

All these observations regarding immune activation have been evaluated in chimpanzees, whose expression of hepatitis after HBV infection is mild compared to humans. This might be the reason why gene expression is absent in chimpanzees. Furthermore, after HCV infection and HBV infection, there is a large difference in IFN production in chimpanzees. This difference might support the observations in the chimpanzee model that HBV evades the first line of host defense mechanisms. The detection of clinical symptoms in patients, generally 10 to 12 weeks after infection (83) might be indirect evidence of the early phases of HBV infection.

After the exponential phase of HBV expansion, chimpanzees able to control the virus show activation of IFN-gamma, TNF-alpha, and genes linked to a Th-type cellular response (78,79). IFN expression, together with activation of NK cells by expression of stress signals on hepatocytes or liver dendritic cells might be determinant factors of subsequent activation of adaptive immunity, the main factor responsible for viral clearance.

1.2.2.2 Adaptive immune system

The adaptive immune response is comprised of a complex chain of effector cell types, all of which play key roles in developing immunity against the virus (Figure 14). HBV replicates in hepatocytes to produce HBsAg subviral particles and virions. Both types of particles can be

taken up by antigen-presenting cells, which degrade the viral proteins to peptides that are then presented on the cell surface, bound to HLA class I or II molecules. Antigen-presenting cells can also display viral antigens, taken up by phagocytosis of the killed infected hepatocytes. CD8+ or CD4+ T cells can recognize these peptide antigens.

The importance of activation of all these cell types is demonstrated by resolution of acute HBV infection, which is associated with a vigorous, polyclonal, multispecific Th and CTL cell response. In addition to lymphocyte activation, the mononuclear phagocytic system (monocytes and macrophages) and Kupffer cells are involved in the immune response. Monocytes and macrophages are able to capture antigens and act as antigen-presenting cells to lymphocytes, whereas non-specific activation of Kupffer cells induces production of TNF-α and IL-12, both HBV inhibitors (78).

T cells play a central role in the mechanisms of action and control of the immune response (84), by responding with specificity and memory against an antigenic stimulus. T cells recognize the antigen as a peptide fragment linked to the HLA molecule through its T-cell receptor. Th cells express the CD4 + molecule that activates cytokines, is able to regulate activation of macrophages, controls antibody synthesis, and participates in CTL activation and expansion. T cells may also have a cytolytic function, in which they recognize peptides expressed in the membrane of antigen-presenting cells through HLA type II.

Cytotoxic T lymphocytes express CD8+ molecules and can recognize and destroy infected cells through cell lysis or through a non-cytolytic pathway, depending on the cytokine activated (19). CTLs recognize infected cells by the peptides bound in the T-cell receptor presented through HLA-I. These peptides, which are presented in the membrane of antigen-presenting cells or infected hepatocytes, come from the HBV proteins, such as HBcAg, HBeAg, and HBsAg. As was described in section 1.1.5.2 (The preCore and Core gene), the most immunogenic HBV antigen is HBcAg, which presents different epitopes that can activate the immune system (Figure 7 and Table 3).

The cellular immune response seems to be the major contributor to HBV clearance; however, humoral responses also play a role in controlling HBV infection. In fact, B-cell antibody production neutralizes free viral particles and can prevent (re)infection. Therefore, integrated activation of both the cellular and humoral arms enables control of infection. The different immune response components are interconnected and failure of one of them clearly affects the expansion and protective efficacy of the others. A lack of CD4+ Th cell response can impair

CD8+ T-cell activity and antibody production (85), whereas an inability to establish a virus-specific CD8+ T-cell response results in a level of circulating virus that cannot be cleared by antibodies alone (86).

Figure 14 Hepatocyte immune response developed in the presence of HBV (9).
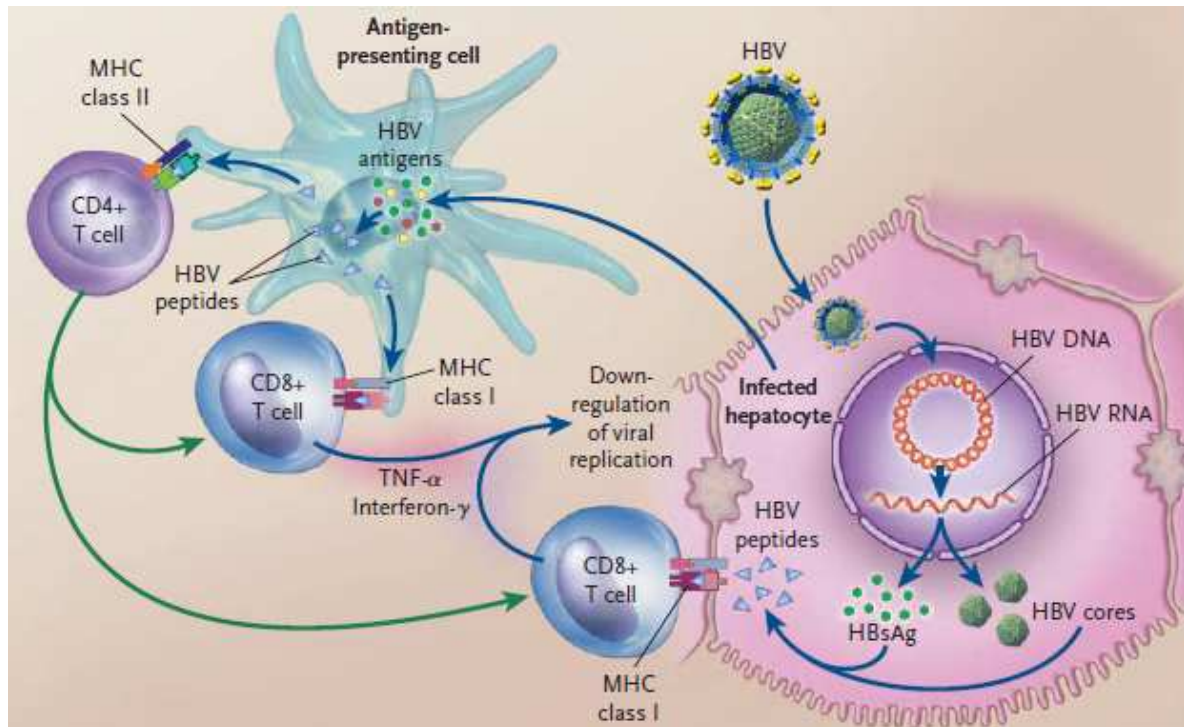
1.2.3 Antiviral Treatments

Progression of liver disease in HBV chronic carriers is directly related to the intensity of viral replication (87). However, progression also occurs in the absence of detectable viral replication (87). The goal of antiviral therapy for hepatitis B is to improve quality of life and survival by preventing progression of the disease to cirrhosis, decompensated cirrhosis, end-stage liver disease, HCC, and death. This goal can be achieved if HBV replication can be suppressed in a sustained manner. The accompanying reduction in histological activity of chronic hepatitis by successful therapy decreases the risk of cirrhosis and HCC in non-cirrhotic patients. However, HBV infection cannot be completely eradicated because of the persistence of cccDNA in the nucleus of infected hepatocytes. It must be kept in mind that successful therapy results from inhibition of the pgRNA reverse transcription step, but not mRNA transcription (including pgRNA). Hence, viral antigen production is maintained even in the absence of active viral

replication. As pgRNA transcription is not blocked, production of this mRNA is maintained under antiviral therapy. If antiviral therapy ends, reverse transcription activity and consequently, viral replication might be reactivated. For this reason, the duration of antiviral therapy for treating chronic hepatitis B infection currently remains undefined, and is mainly considered indefinite. However, some virologic and serologic parameters can be evaluated for stopping therapy. For example, in HBeAg-positive cases, when HBV DNA levels become undetectable and seroconversion to anti-HBe is achieved and maintained for at least 12 months, treatment can be stopped (88).
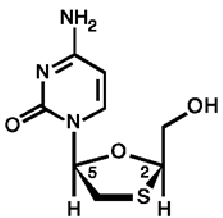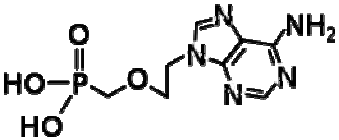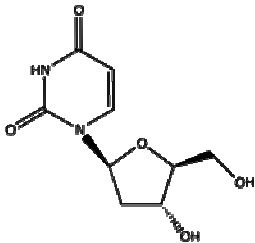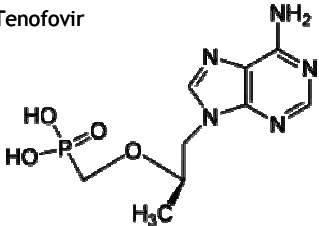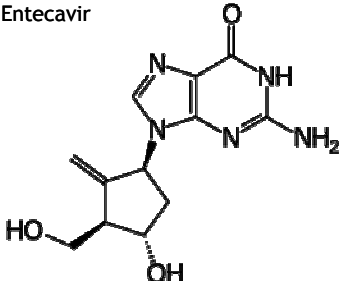
Currently, seven antiviral treatments are available for chronic HBV infection, divided in two groups: immunomodulators (interferons) and direct oral antivirals (viral polymerase inhibitors). The interferons (IFN αnd pegylated IFN) have immune-stimulatory properties, and this is the main mechanism of their antiviral activity against HBV. In general, HBeAg-negative patients present higher rates of response than HBeAg-positive ones (~20%) (Figures 15 and 16). The response to IFN is clearly conditioned by HBV genotype, being higher in genotype A than genotype D cases (70). Interferon administration has been associated with improved serological response, such as HBsAg loss and HBeAg seroconversion. However, the fact that administration is by intravenous route, the potentially severe associated side effects, and the genotype-dependent efficacy, limits this treatment to a small percentage of highly selected cases (ie, young patients with genotype A) (70).

Figure 15 Percentage of response to each antiviral after one year of treatment, attending to HBeAg status (88).
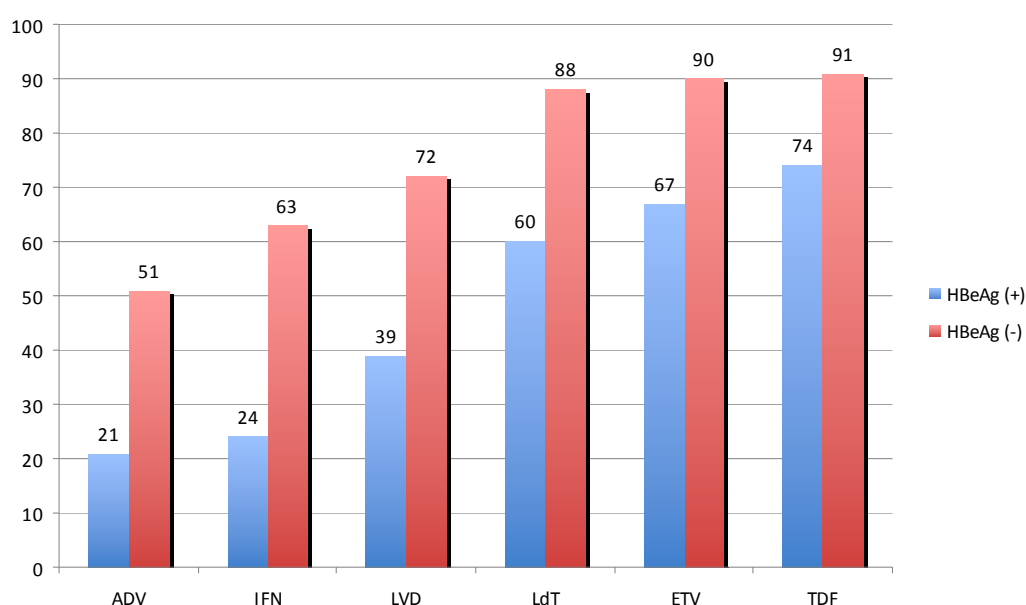
The other drug group is oral viral polymerase inhibitors, which include nucleoside and nucleotide analogs (NUCs): L-nucleosides (lamivudine [LVD] and telbivudine [LdT]), alkyl phosphonates (adefovir dipivoxil [ADV] and tenofovir disoproxil fumarate [TDF]) or D-cyclopentanes (entecavir [ETV]) (Table 4). Several clinical trials have assessed the efficacy of these treatments and have shown that ETV and TDF are the most potent currently available antivirals. The recently updated EASL guidelines recommend ETV and TDF as first line treatment, and IFN for young, genotype A males with low HBV DNA load (89). The response to treatment differs between HBeAg-positive and HBeAg-negative cases, as is summarized in Figure 15.

Table 4 Chemical formulation of the five types of nucleos(t)ide analogues.

The most important disadvantage of these treatments is selection of mutations in the HBV polymerase gene conferring resistance to NUCs in long-term therapy (section 1.1.5.3, The polymerase gene and section 1.1.8, Viral variability: Quasispecies structure of HBV). Mutations selected by NUC treatment can be split into 2 groups: those that cause resistance, sometimes leading to decreased viral fitness and conferring a capacity to replicate under the treatment

effect, and compensatory mutations, which partially or fully restore viral fitness. Treatment resistance has been found in 70% of patients after 4 years of LVD treatment (90), in 29% of HBeAg-negative patients after 5 years of ADV (91), in 25% of HBeAg-negative patients after 2 years of LdT (92), but in only 1.2% of treatment-naïve patients after 5 years of ETV (93). In contrast, no resistances have been associated with TDF after 5 years of therapy (Figure 16) (89,94). The main mutation patterns described in the polymerase gene, classified according to their sensitivity to treatment are presented in Table 5.

Table 5 Main mutations described in the HBV polymerase gene that are resistant (R), intermediate (I) or sensitive (S) to oral antiviral drugs.

Main mutaions	LVD	LdT	ETV	ADV	TDF
M204I	R	R	I	S	S
L180M+M204V	R	R	I	S	S
A181A/T	I	S	S	R	S
N236T	S	S	S	R	I
L180M+M204V/I±I169T±V173L±M250V	R	R	R	S	S
L180M+M204V/I±T184G±S202G	R	R	R	S	S

Figure 16 Cumulative incidence of HBV resistance to Lamivudine (LVD), adefovir (ADV), entecavir (ETV), telbivudine (LdT), and tenofovir (TDF) in pivotal trials in NUC-naive patients (88).
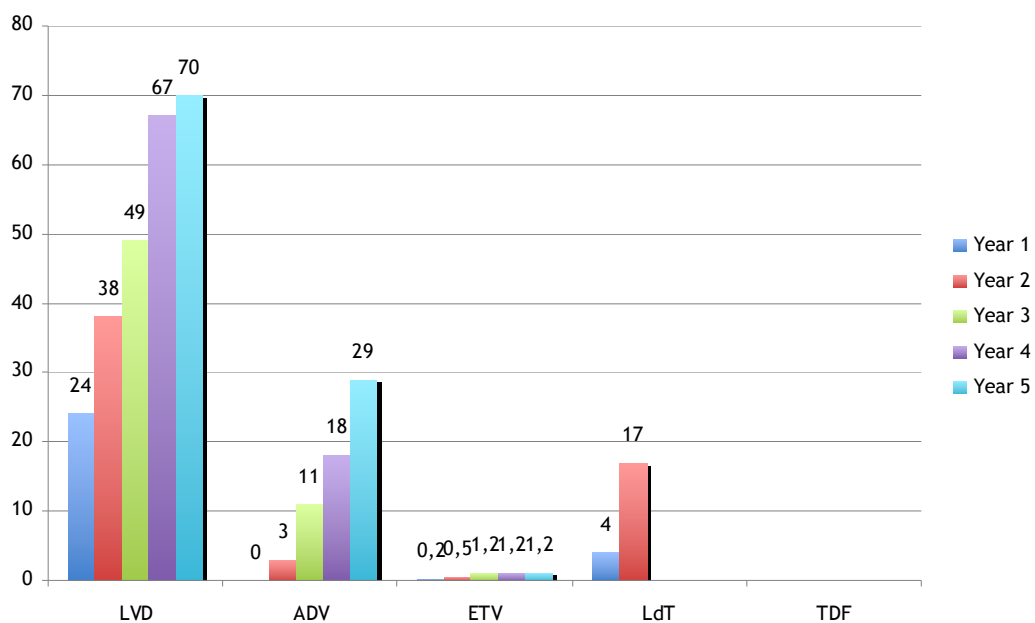
Selection of resistance mutations is usually characterized by a sudden, fast increase in viral activity, known as viral breakthrough (VBK). An increase in serum HBV DNA is seen in VBK, as well as an increase in liver damage, reflected by the subsequent elevation of liver enzymes (ALT). In theory, resistance can be prevented with a sufficiently potent antiviral drug or a combination of NUCs. Currently VBK is mainly prevented by use of the last-generation antivirals ETV or TDF (95). Before commercialization of ETV and TDF, the most commonly used antiviral treatment was LVD, which was FDA-approved in 1995. This drug presented high percentages of response, especially in HBeAg-negative patients (Figure 15). However, after 5 years of LVD use, more than 70% of patients presented HBV variants with polymerase mutations resistant to treatment. The main variant associated with LVD resistance is rtM204I/V, which occurs within the YMDD motif of the reverse transcriptase region of the polymerase. Although viral strains with rtM204I/V mutations (YVDD/YIDD motif) can replicate under LVD, these variants present lower replication rates without treatment than variants without mutations (YMDD motif), both *in vivo* (96) and *in vitro* (97).

Some mutations within the HBV polymerase gene act as compensatory mutants. This is the case of the variants rtL80I/V, rtL180M, and rtV173L. When these are present in the same HBV genome as the M204V mutation (YVDD), the replication efficiency of the resulting strains is similar to that of wild-type (YMDD) strains (97). LVD-resistant mutations also present resistance against L-nucleotide analogues drugs. For example, in patients treated with LdT, selection of rtM204I mutations (but not rtM204V) has been observed (98). Furthermore, ETV resistance patterns include the simultaneous presence of LVD-resistance mutations (rt L180M+rtM204I/V, commonly known as "LVD signature"), linked to an additional mutation in positions rtT184G, rtS202I/C, rtM250V, or rtI169T (99). Similarly, the mechanism mediating ADV resistance has major resistance mutations at rtN236T and/or rtA181T/V (100).

The complex patterns of mutations that can accumulate within the HBV genome over time, especially in the retrotranscriptase domain, may affect the efficacy of subsequent treatments Thus, identification of mutations will quickly become an important component in the management of patients with CHB. Because of early, widespread application of LVD treatment, most CHB patients have been treated with LVD. Although this drug is now less widely used, patients that have mutations against LVD are at a higher risk of reselecting these mutations when they are treated with ADV or even ETV (101,102,103). It is important to take into account that the rate of resistant mutations against ADV or ETV is higher in patients previously

treated with LVD, than in naïve patients. Specifically, 25.4% of LVD-treated patients reselect mutations after 2 years under ADV (101). In addition, only 1% of naïve ETV-treated patients present resistance to ETV compared to 40% patients previously treated with LVD (102).

As was mentioned above and is shown in Figure 15, HBeAg-negative patients present better treatment response rates than HBeAg-positive patients. The main virological difference between the two types of patients is presence of the main preCore mutation (codon 28), which inhibits expression of the antigen in HBeAg-negative cases. Some studies have evaluated the possible relationship between selection of mutations in the YMDD motif and HBV viral strains carrying the main preCore mutation (104,105), but no conclusive results have been obtained. In this line, seroreversion of HBeAg (from antiHBe positive to HBeAg positive) is quite common under LVD treatment, indicating a possible relationship between preCore mutations and treatment (105-107).

1.3 Ultra-deep pyrosequencing as a tool for studying HBV quasispecies

1.3.1 Sequencing methods

In 1975, Sanger and Coulson described the groundbreaking methodology of plus-minus sequencing (108), based on primed synthesis in the presence of polymerase combined with limited availability of deoxy NTPs, with visualization in acrylamide gels. This technique was optimized two years later. Maxam and Gilbert combined chemical cleavage of radiolabeled DNA at specific bases to determine the sequence of a long DNA fragment via Southern blotting (109). In another study, Sanger et al. described a plus-minus strategy that used dideoxy NTPs as chain-terminator inhibitors (109). This last method has been simplified and is now widely used.

Since 2005, next generation sequencing (NGS) platforms have changed genomic study (110). There are several commercially available NGS systems, all with a common characteristic: generation of thousands of millions of sequence reads in parallel within the same run. In contrast to Sanger sequencing, which is based on electrophoretic separation of chain-termination products and generates individual sequencing reactions, NGS techniques use

repeated cycles of polymerase-mediated nucleotide extensions or iterative cycles of oligonucleotides. The differences between platforms are related to template preparation and/or the sequencing chemistry (Table 6).

Table 6 Comparison of NGS platforms.

Characteristics	454 (Roche)	Illumina (Solexa)	Solid (Applied Biosystems)	CGAP	tSMS (Helicos)	SMRT (Pacific Biosciences)
Chemistry	Pyrose-quencing	Reversible dye terminators	Ligation	Ligation	Reversible dye terminators	Synthesis with stationary polymerase
Length (bp)	400-1000	50-100	50-100	50-100	50-100	400-1000
Template preparation	emPCR	Clonal bridge amplification	emPCR	Solution base	None*	None*

CGAP, Complete Genomics Analysis Platform. TSMS, True Signal Molecule Sequencing. SMRT, Single Molecule Real Time. *Avoids DNA template preparation by applying single molecule sequencing.

1.3.2 The 454 technology

Among the more consolidated NGS platforms, the 454 Sequencing System enables sequencing of the longest fragments. This is an advantage when analyzing simultaneous changes in the same viral genome to investigate possible complementary or synergic roles between distanced positions. For instance, study of ETV-related antiviral resistance requires analysis of a 241-bp fragment, because the possible mutations might be I169T, V173L, L180M, M204V/I, or M250V. To analyze long fragments with other technologies, a computational study based on overlapping of small fragments would be required, and this can be a very difficult, and in some cases, impossible task because of the extreme complexity of the HBV quasispecies.

The 454 technology is based on pyrosequencing, a DNA sequencing method based on the "sequencing by synthesis" principle, which relies on detection of pyrophosphate release on nucleotide incorporation rather than chain termination with dideoxy NTPs (111-113). The 454

Sequencing System supports analysis of samples from a wide variety of starting materials, including genomic DNA, PCR products, bacterial artificial chromosomes and cDNA. The system relies on fixing nebulized and adapter-ligated DNA fragments to small DNA-capture beads in a water-in-oil emulsion. The DNA fixed to these beads is then amplified by PCR. Each DNA-bound bead is placed on a PicoTiterPlate (PTP), a fiber optic chip. A mix of enzymes such as DNA polymerase, ATP sulfurylase, and luciferase are also packed into the well. The PTP is then placed into the GS FLX System for sequencing.

1.3.2.1 DNA library preparation and emPCR

The steps required for pyrosequencing are shown in Figure 17. The initial step is the generation of a library of template DNA, where short adaptors are included in the primers. The adaptors are added at the 5' and 3' ends of the amplicon (Figure 17a) and contain universal priming sites to simplify subsequent amplification, purification, and sequencing.

The DNA library is purified, titrated, and assessed for quality. It is then diluted to guarantee a ratio of one DNA molecule to one bead. DNA molecules are then denatured and the single-stranded template DNA (sstDNA) is obtained. The sstDNA is hybridized to individual beads, which contain sequences complementary to the adapter sequence included in the amplification primers. Thus, the amplicon library is immobilized onto the beads.

Once the bead-bound library is obtained (Figure 17b), it is emulsified with amplification reagents in a water-in-oil mixture (Figure 17c). The method has been designed to produce oil drops whose diameter limits introduction of one bead into each drop. With this strategy, each oil drop contains a single bead and a single DNA molecule, and acts as a micro-reactor where PCR amplification occurs. This process, known as emulsion PCR (emPCR) (Figure 17d), induces amplification of sstDNA molecules with primer coated beads in aqueous droplets within an oil phase, resulting in bead-immobilized, clonally amplified DNA fragments.

Figure 17 Steps required for emulsion PCR. (a) Template DNA generation by PCR amplification using primers that include the sequence of adapters A (forward) and B (reverse). After dilution and preparation of the DNA library (b), the diluted DNA library is emulsified in water-in-oil (c) and clonal amplification is then performed (d).

a. Template DNA generation

b. Bead-bound library

- Purification and Tritation
- Dilution of DNA
- Denaturation: sstDNA
- Hybridiyation to coated beads

A

B

Adapters A and B carring library DNA

One sstDNA molecule per coated bead

- PCR reagents
- Emulsion Oil

d. emPCR in each micro-reactor

c. Water-in-oil emulsion

- High speed shaker

Each drop acts as a micro-reactors

Clonal amplification of each sstDNA

After emPCR amplification, the emulsion is disrupted (Figure 18a) and beads are counted. Optimal recovery is between 65% and 85% (Figure 18b). The recovered beads are enriched by adding a primer containing a 5´-biotin tag that anneals to Adaptor B; therefore, beads containing clonal amplified DNA are selected. Streptavidin-coated beads are then added, and the DNA library is immobilized (Figure 18c). After the realising of the non-biotinylated strand, the beads remaining in the magnet are used as the single-stranded template DNA (sstDNA) library (Figure 18c). The optimal enrichment is between 5% and 20% (Figure 18d). After this step, and if recovery and enrichment percentages are optimal, the product is deposited in a PTP (Figure 19 a).

Figure 18. Steps prior to the sequencing reaction. First, emPCR is disrupted (a), the percentage of recovered beads is quantified (b), and the solution is enriched with biotin primers and streptavidin. The immobilized beads in a magnet (c) are quantified (d) to proceed with the sequencing.
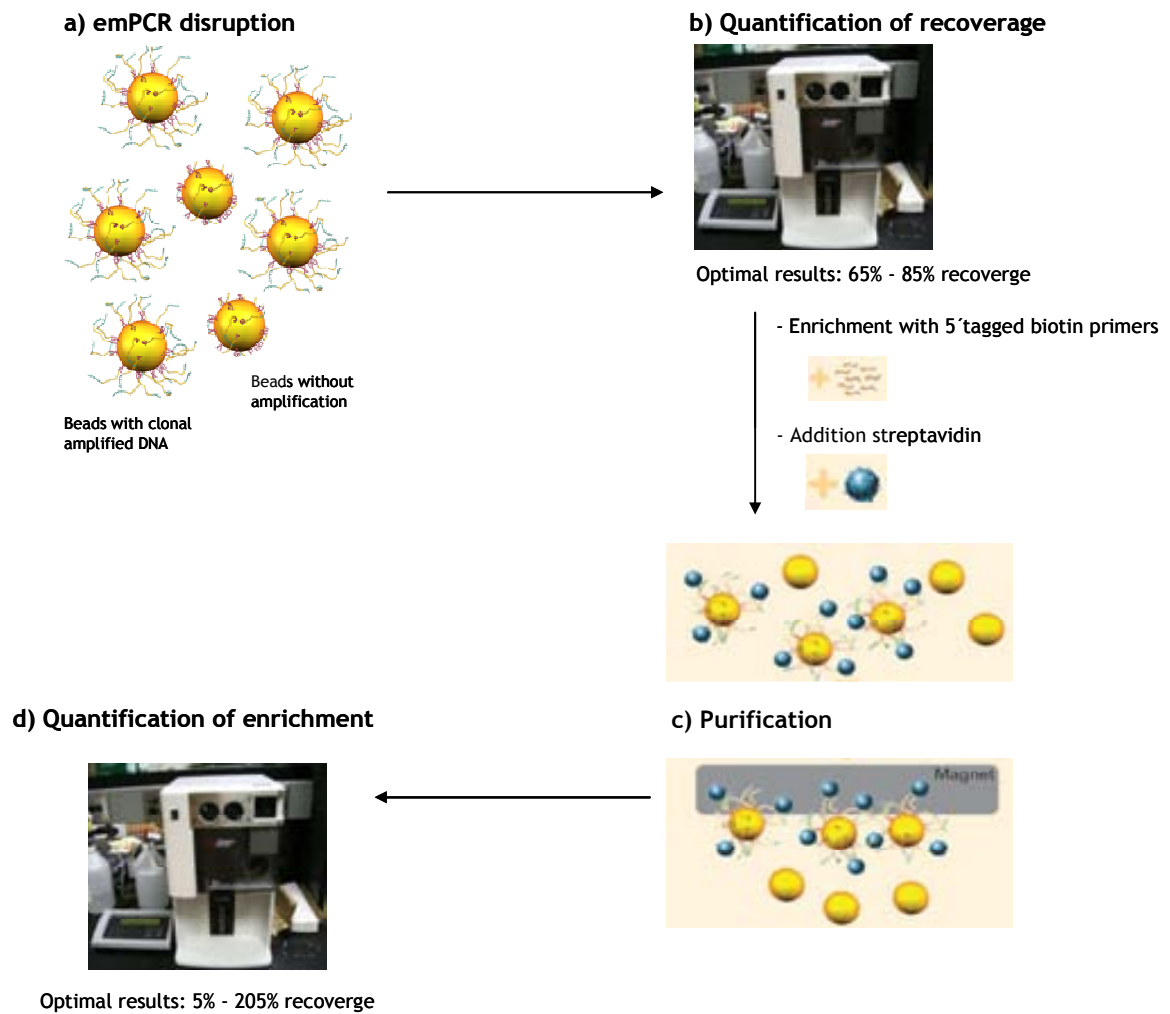
Figure 19 The treated emPCR product is deposited in the PTP (a). In theory, only one bead containing one clonally amplified DNA strand is incorporated in each PTP well.
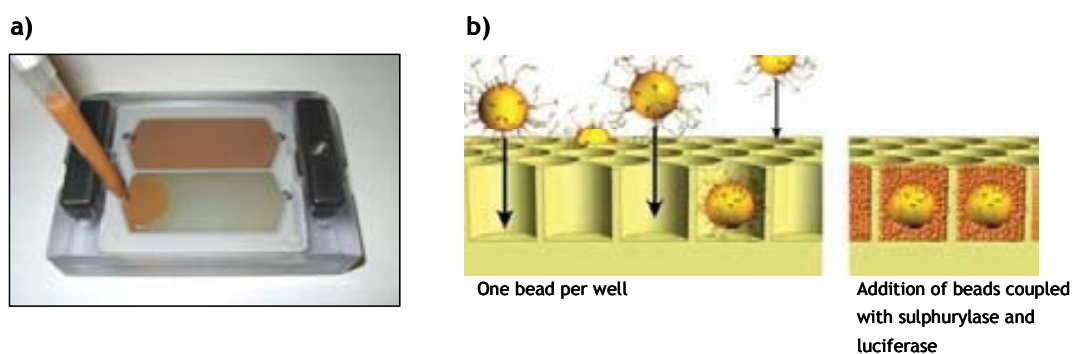
1.3.2.2 Sequencing

The diameter of the PTP wells allows addition of only one bead per well. sstDNA library beads are added to the DNA bead incubation mix (containing DNA polymerase) and are layered with enzyme beads (containing sulfurylase and luciferase) onto a PTP (Figure 19b). The 70-mm×75-mm plate is an optical device containing 1.6 million wells at a diameter of 29 μm per well. The device is centrifuged to deposit the beads in the wells. The layer of enzyme beads ensures that the DNA beads remain positioned in the wells during the sequencing reaction. The bead deposition process maximizes the number of wells that contain a single amplified library bead (avoiding more than one sstDNA library bead per well).

The PTP is placed in the FLX Genome Sequencer. The fluidics sub-system delivers sequencing reagents (containing buffers and dNTPs) across the wells. The four nucleotides are added sequentially in a fixed order across the plate during each sequencing run (Figure 20a). During the nucleotide flow, millions of copies of DNA bound to each bead are sequenced in parallel. When a nucleotide complementary to the template strand is added to a well, the polymerase extends the existing DNA strand by adding nucleotide(s). Addition of one (or more) nucleotide(s) generates a light signal (Figure 20b) that is recorded by a Charged-coupled Device (CCD) camera in the instrument (Figure 20c).

Figure 20. Sequencing includes the flow of the four nucleotides (a), always in the same order. When one nucleotide adds to the DNA template, a light signal is generated (b) and it is reordered by the CCD camera (c).
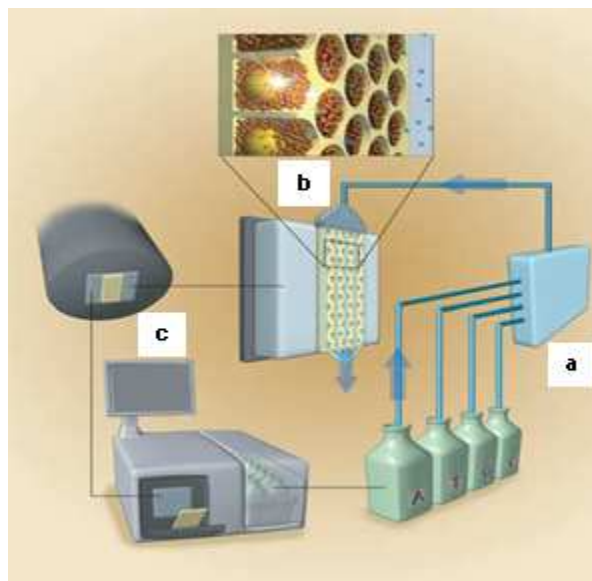
The signal strength is proportional to the number of nucleotides in a single-nucleotide flow; for example, homopolymer stretches, incorporated in a single nucleotide flow generate a stronger signal than single nucleotides. However, the signal strength for homopolymer stretches is linear only up to eight consecutive nucleotides, after which the signal rapidly falls off. Data are stored in standard flowgram format files for downstream analysis.

Pyrosequencing is based on chemiluminescent detection of pyrophosphate (PPi) released during polymerase-mediated dNTP incorporation. ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5´ phosphosulfate. The ATP acts as fuel for luciferase-mediated conversion of luciferin to oxyluciferin, which generates visible light in amounts proportional to the amount of ATP (Figure 21). The light produced in the luciferase-catalyzed reaction is detected by a camera and analyzed by dedicated software. GS-FLX analyses the signal-to-noise ratio of each dNTP, filters according to quality criteria, and subsequently algorithmically translates the output data into a linear sequence.

Figure 21. Pyrosequencing reaction (APS, adenosine 5´ phosphosulfate).
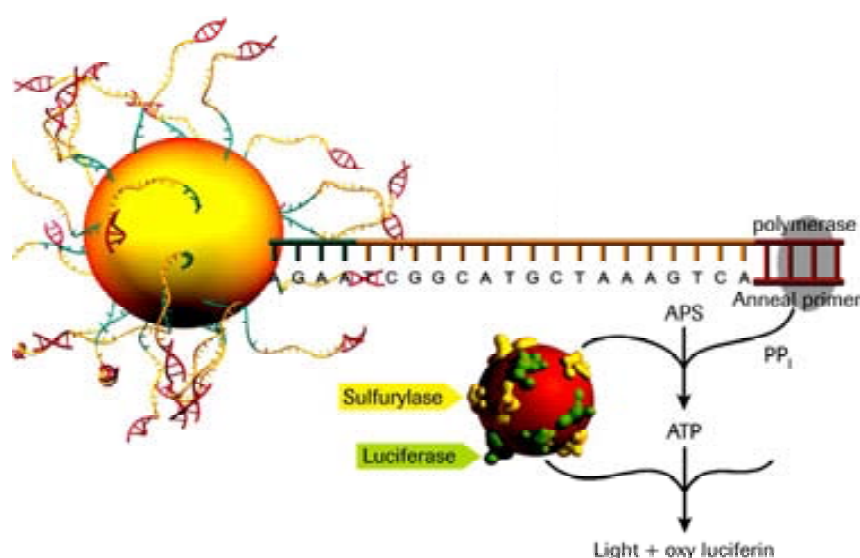
1.3.3 Data analysis and error rate

The NGS experiments generate high volumes of data requiring a complex, powerful data-pipeline system for storage, management, and processing. The main feature of the data pipeline is the computational conversion of image data into sequence reads, known as base calling. First, individual beads are identified and localized in image series. Image parameters, such as intensity, background and noise, are used in an algorithm to generate read sequences and error probability-related quality scores for each base. The flowgram obtained after an

experiment is similar to Figure 22. A four-NT calibrating sequence (TCAG) is incorporated in the amplification primers.

Figure 22 Flowgram obtained after a pyrosequencing reaction. According to the quality criteria and the algorithms, the data is translated into a linear sequence output.

Sequence output derived from the flowgram
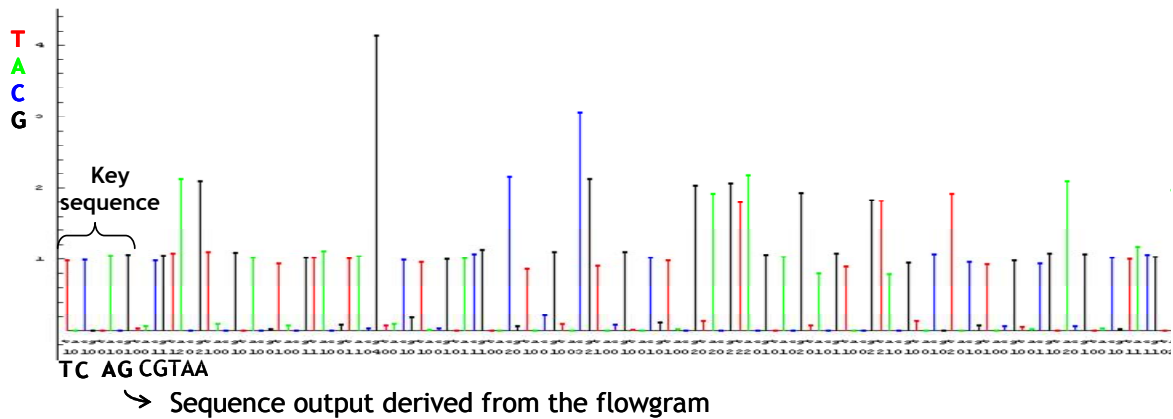
Specific 454-platform data-pipeline software is available (Amplicon Variant Analyzer, AVA, Roche). However, there is a need for more advanced software with specific statistical tests that include incorporation of ambiguous bases into reads, improve removal of poor-quality bases, and filtering of mismatch sequencing errors (more information about the filter applied in the present work is described in the Annex, third study). Recently, G to A hypermutations have been identified in HBV UDPS reads (60).

NGS error rates depend on several factors, such as signal-to-noise levels, cross-talk from nearby beads, and dephasing. Accuracy in NGS is achieved by sequencing a given region several times, thereby contributing to deep coverage. In the case of sequences longer than the maximal sequencing length allowed by 454 technology, an adequate number of overlapping reads is required to assemble, align, and analyze the data.

First Study

2 FIRST STUDY: HBV core region variability: effect of antiviral treatments on main epitopic regions

2.1 Hypothesis and Aims

2.1.1 Introduction

The first study was designed to analyze Core gene variability under different conditions. HBcAg contains various regions that can stimulate the host immune system. The most important epitopes are Th50-69 and B74-84, but the entire Core gene has regions able to activate the immune system (22-25,27,28,65,66,114). The Core shows the least overlapping of all genes in the HBV genome and this enables accumulation of mutations all along its sequence (115). The hypothesis of this study was that selection of mutations in the Core gene might be an alternative mechanism for HBV to maintain its viral fitness. Disease progression is related to Core gene changes (116-118), which have been particularly described during interferon (IFN) administration, indicating a possible treatment escape mechanism (119,120).

The study was focused on determining the effect of Nucleos(t)ide analogues (NUCs) under the assumption that they have little or no immune-stimulating activity (121). It must be kept in mind that when an optimal antiviral response occurs, viral antigen production is maintained. Therefore, immune stimulation may continue to act against these antigens. We postulated that selection of mutations in the Core gene to escape treatment might be a complementary or alternative mechanism to mutations in the polymerase gene that confer resistance to treatment. In addition, other viral factors might interfere with treatment response, such as mutations in the preCore region that abolish HBeAg expression, or the HBV genotype. The absence of HBeAg could induce alternative immune responses, such as lymphocyte activation against Core epitopes. At the time the study was designed, we postulated that an absence of HBeAg expression might induce greater Core variability.

The cases of HBV inactive carriers (ICs) from our area are of special interest. These are patients with persistently low rates of HBV replication and low ALT levels. We hypothesized that these cases might select mutant Core viruses with low HBV replication capacity also able to evade the host immune system.

2.1.2 Hypothesis

The virological, biochemical, and histological differences between ICs and chronic hepatitis B (CHB) patients might be reflected by differential selection of changes in the Core gene. Moreover, the effect of NUCs on Core gene variability might be different from that occurring with IFN treatment.

2.1.3 Aims

1.   Evaluate the selection of changes in the HBV Core gene in three epitopic regions (Th28-47, Th50-69 and B74-84) and in a conserved region (AA1-11) in baseline samples of HBV-infected patients.

2.   Establish a comparison between the number of changes selected in the Core gene in chronic hepatitis B patients and in inactive HBV carriers.

3.   Analyze changes in the Core gene under IFN treatment and compare them with the changes under NUCs.

4.   Study possible associations of main preCore variants and HBV genotype in the evolution of the Core gene.

2.2 Summary of the study

Amino acid variability in specific HBV core gene regions was assessed in HBV-infected patients. To evaluate this variability, a fragment of the Core gene from nt position 1800 to 2559 was directly sequenced. Four regions were considered for this study; a region delimited by amino acid 1 to 11 (AA1-11), reported as highly conserved (122), a minor Th epitope from amino acid 28 to 47 (Th28-47) (24), the main Th epitope (Th50-69) and B lymphocyte epitope (B74-84) (22-25). Only 4 inactive HBsAg carriers had HBV DNA levels that were high enough to amplify

the Core region. A total of 76 CHB samples were selected and several sequential serum samples were studied. In some patients, different samples were available and all were included in the study, in order to obtain data during treatment-free (TF) periods or following administration of IFN, lamivudine (LVD) or adefovir (ADV). The following samples were included: 23 in the TF group, 22 in the IFN group, 39 in LVD patients, and 17 in the ADV group.

Core gene variability was established by detection of amino acid changes in the serum samples analyzed. Each sample was compared with the previous sample, and the baseline sample was compared with the consensus of the corresponding HBV genotype. For example, in a genotype A CHB patient with 3 samples (baseline, TF, and LVD), the results of sequencing the LVD sample were compared with the TF results, sequencing of the TF sample was compared with the baseline sequence data, and the baseline results were analyzed according to the consensus sequence of HBV genotype A.

Detection of an amino acid change was scored with 1 point. The number of changes in each sample was expressed as the percentage of changes per group (%N) and the average of changes after a standardization of 12 months (Av12), attending to the time of infection or the period between the samples analyzed. The results of the standardization are presented in Figure 23.

Figure 23 Standardized results of the changes detected in each group analyzed.

48
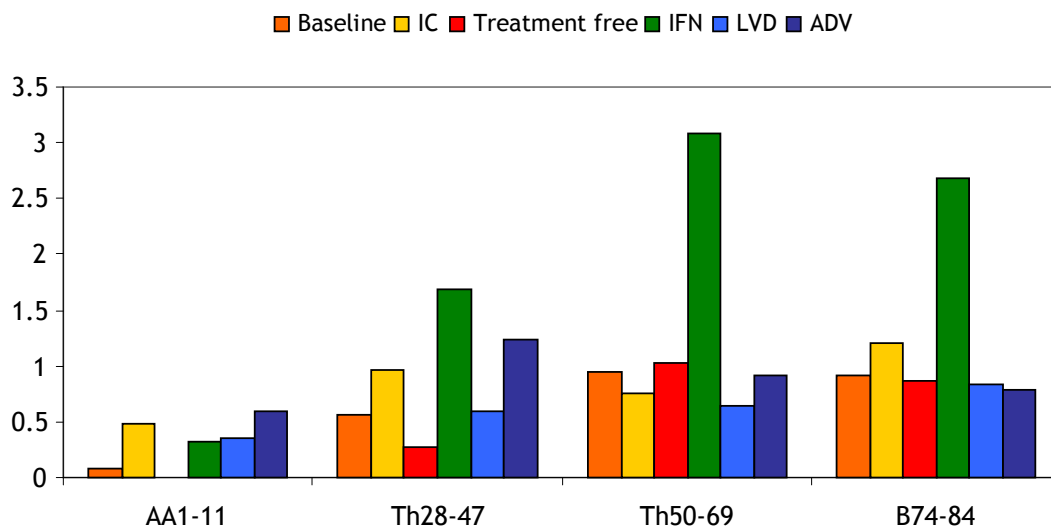
Baseline variability was assessed in 185 samples included in the study. At baseline, the %N and the Av12 detected in the AA1-11 region was significantly lower than in the three other regions analyzed (p<0.001). No differences in Core variability were observed between genotypes. In contrast, HBeAg-negative cases showed higher Av12 in Th50-69 and B74-84 than HBeAg-positive patients (Th50-69, p=0.04 and B74-84, p=0.06). Indeed, cases carrying the main preCore mutation (G1896A) presented higher variability than the wild type (Th50-69, p=0.07 and B74-84, p=0.02). Patients with changes in B74-84 had higher HBV DNA levels than patients in whom B74-84 was conserved (p=0.04).

TF samples were included to evaluate the natural evolution of the HBV Core gene. The changes detected in TF samples were compared with those in the previous baseline sample. The TF group presented the higher percentage of changes and higher Av12 variability in the Th50-69 and B74-84 regions than in Th28-47 or AA1-11. This high variability in Th50-69 and B74-84 was also detected in samples under IFN, LVD, or ADV treatment. However, the highest Av12 changes in epitopic regions were detected after IFN (%N=73%, Av12=3.1 in Th50-69 and %N=86%, Av12=2.7 in B74-84). Interestingly, changes in the Th28-47 epitope were significantly higher after LVD or ADV than in the treatment-free period (%N LVD vs. TF, p=0.04 and %N ADV vs. TF, p=0.01), possibly indicating an alternative immune system activation than that observed in natural evolution or even after IFN treatment. In addition, a comparative analysis was performed in patients with a treatment-free period followed by treatment with IFN, and it was confirmed that variability under IFN was localized in the main epitopic regions.

After TF, IFN, LVD or ADV, Core variability showed no differences according to HBV genotype (A and D). Nor did the main preCore mutation present significant differences, except under the effect of LVD: cases with the wild-type preCore mutation showed higher Av12 variability in Th50-69 (p=0.03) and B74-84 (p=0.08) than cases with the main preCore mutation. As to the specific mutated Core positions, codons 74 and 77 were the most polymorphic ones, and the E64D-N67T double change was significantly observed. Codon 84 substitutions were mainly associated with IFN treatment (p=0.05).

A comparison of CHB patients and ICs was performed, but the number of IC patients included was small because of their low viral loads. Despite this limitation, a higher variability was detected in the Th28-47 region in IC than TF samples and higher variability in Th50-69 and B74-84 in IFN than IC. These differences might indicate alternative mechanisms of Core variability to evade the immune response in the ICs.

Based on our results, it was concluded that baseline variability in the Core gene was mainly detected in the main Th50-69 and B74-84 epitopes. HBeAg-negative baseline samples contained most of the changes in Th50-69, indicating that this region may be enhanced by the host immune response during HBeAg loss. We also postulate that the high variability in the B74-84 epitope might confer a mechanism to escape the immune system because baseline samples with changes in the B epitope had higher HBV DNA levels than samples with a conserved B74-84 region.

The natural evolution of the Core gene showed highest variability in the main epitopes, but the AA1-11 region was the most highly conserved of all the different groups, baseline, TF, IFN, LVD and ADV. Treatment-induced substitutions in the HBV core protein were particularly frequent under IFN. The main preCore mutation showed no effect on Core variability, except for cases under LVD treatment, which presented lower variability than wild-type preCore, possibly explaining seroreversion of HBeAg during LVD treatment in some patients. We observed some immune-stimulating activity related to the minor Th28-47 epitope under NUC treatment, indicating possible maintenance of viral protein synthesis. This was also observed in the IC group, suggesting a possible alternative mechanism of immune escape.

2.3 Complete manuscript

Original article

HBV core region variability: effect of antiviral treatments on main epitopic regions

Maria Homs[1,2], Rosendo Jardi[1,2], Maria Buti[1,3], Melanie Schaper[1,2], David Tabernero[1,2], Pilar Fernandez-Fernandez[2], Josep Quer[1,3], Rafael Esteban[1,3], Francisco Rodriguez-Frias[1,2]*

[1]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto Carlos III, Madrid, Spain
[2]Department of Biochemistry, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
[3]Department of Hepatology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

*Corresponding author e-mail: frarodri@vhebron.net

Background: Amino acid (AA) changes in specific hepatitis B core antigen (HBcAg) regions were assessed in patients infected with chronic hepatitis B (CHB) after a 12-month untreated period and after receiving antiviral therapy (interferon, lamivudine or adefovir dipivoxil), and in inactive hepatitis B surface antigen–positive carriers.

Methods: Samples corresponding to different time points in 76 CHB cases (64 on-treatment) and 4 inactive carriers were included. The main precore mutation, T-helper immunodominant epitope at AA 50–69 (Th50–69), minor T-helper epitope (Th28–47), B-cell immunodominant epitope (B74–84) and a conserved region of HBcAg at AA 1–11 (AA1–11) were directly sequenced. For comparisons, the average number of AA changes in each region was standardized to 12 months (Av12).

Results: AA changes clustered mainly in immunodominant regions (69%). The highest percentage of cases (%n) with changes and highest Av12 changes were detected after interferon treatment (%n=73%, Av12=3.1 in Th50–69 and %n=86%, Av12=2.7 in B74–84). At baseline, immunodominant regions had higher Av12 changes in hepatitis B e antigen–negative patients and those with main precore mutations. Changes in the Th28–47 region were more frequent after nucleoside/nucleotide analogue treatment (40%) than before treatment (9%). Codons 74 and 77 were the most polymorphic, and the double change E64D–N67T was significantly observed. Codon 84 substitutions were mainly associated with interferon treatment (P=0.05).

Conclusions: Natural and treatment-induced substitutions in HBV core protein, occurring especially with interferon treatment, were characterized. Some immune-stimulating activity related to the minor Th28–47 epitope might be associated with nucleoside/nucleotide analogues; this activity was also seen in inactive carriers.

Introduction

The HBV DNA core region encodes the hepatitis B core antigen (HBcAg), the viral capsid component and hepatitis B e antigen (HBeAg), a secreted non-particulate form of HBcAg [1,2]. HBV is not directly cytopathic; the cellular immune response is believed to be responsible for the pathogenesis of the disease. Infected hepatocytes are recognized and destroyed by HBV-specific cytotoxic T-lymphocytes, which act against peptides bound to human leukocyte antigen (HLA) type-I that is expressed on the cell surface [3,4]. Resolution of acute HBV infection is associated with a multispecific CD4+ T-helper and CD8+ cytotoxic T-lymphocyte response [5]. Development of chronic hepatitis B (CHB) infection is associated with poor establishment of the immune response during the acute phase [4,6].

HBeAg and HBcAg, which share a large number of amino acids (AA) and some epitopes for B- and T-cells, are important targets for antiviral immunity [7–9]. HBcAg, which is more immunogenic than hepatitis B surface antigen (HBsAg) and HBeAg proteins, is a major target for immune-mediated viral clearance, inducing B-cell, T-helper and cytotoxic T-lymphocyte responses. The nucleocapsid-specific CD4+ T-cell response has been identified as the dominant helper response correlating with recovery of HBV infection [4,10–12]. Several epitopes have been identified in HBcAg [11]. An important one is the immunodominant CD4+ T-helper

epitope unrelated to HLA haplotype, which is located between AA 50 and 69 (Th50–69) [13]. The major antibody specificities against HBcAg lie on the outside of the capsid structure, particularly at the tip of spikes where the main core antigenic loop is located, including the immunodominant B-cell epitope between AA 74 and 84 (B74–84) [9,14,15]. Although HBcAg contains a large percentage of invariant AA, significant sequence diversity is seen, predominantly in the capsid spikes where the two immunodominant epitopes are located [15]. This variability might cause alterations in the immunogenic core epitopes by modifying their sequences, which would decrease HBV recognition by T-cells and result in a low antiviral immune response, or even immune escape [4,16,17]. HBcAg AA changes have been associated with viral persistence by influencing the host immune response and the natural course of HBV infection [2,9,18,19]. These changes mainly occur in immunodominant T- and B-cell epitopes, such as Th50–69 and B74–84 [10,11,16,19].

The primary goal of antiviral treatment for CHB infection is sustained suppression of HBV replication and remission of liver disease. The antiviral effect of interferon (IFN) is mainly based on enhancing the cellular immune response against infected hepatocytes by increasing expression of major histocompatibility class I antigens and stimulating the activity of T-helper cells and natural killer lymphocytes [11]. Selection of HBcAg changes caused by IFN pressure or during natural seroconversion mainly occurs in immunodominant T- and B-cell epitopes recognized by CD4⁺ T-helper cells or by B-cells, and is considered an immune escape mechanism [12,16,19]. A large number of nucleoside/nucleotide analogues that act as HBV reverse transcriptase inhibitors are currently available to treat CHB [20]. To date, there are no studies investigating the relationship between emergence of AA changes in HBcAg epitopic regions and nucleoside/nucleotide analogue treatment.

The aim of this study was to evaluate a selection of HBV core gene changes in three epitopic regions (Figure 1): the main CD4⁺ T-cell recognition site (Th50–69), the main B-cell site (B74–84) [13,21,22] and a minor T-helper cell recognition site between AA 28 and 47 (Th28–47) [21–23]. In addition, we examined a conserved region of HBcAg from AA position 1–11 (AA1–11) [15]. These core regions were analysed in CHB patients with persistent viral activity and in inactive HBsAg-positive carriers. In CHB patients, selection of changes in the core nucleotide sequence was examined at baseline and in several additional serum samples: after a treatment-free period, and after non-response to IFN, lamivudine (3TC) or adefovir dipivoxil (ADV) therapy. The role of the G1896A HBV precore variant and HBV genotypes in selecting HBV core changes was also analysed.

Methods

Patients

A retrospective study was carried out in 76 CHB patients diagnosed and followed-up in our centre (Vall d'Hebron Hospital, Barcelona, Spain). In addition, a group of 20 inactive HBsAg-positive carriers were analysed for inclusion in the study. Inactive carriers were defined as follows: HBsAg-positive, detection of antibody against HBeAg (anti-HBe), normal alanine aminotransferase levels and low rates of viral replication (HBV DNA≤4 log₁₀ copies/ml) over the entire duration of known infection. Because of the study requirement to analyse two samples at different time points, only four inactive carriers who had HBV DNA levels between 3 and 4 log₁₀ copies/ml after 12 months were included in the study.

A total of 185 serum samples from the 80 patients enrolled in the study were analysed. These included a baseline serum sample from the 76 CHB patients and serial samples corresponding to various time points, which were classified as the treatment-free period group, the IFN, 3TC and ADV treatment period groups, and the inactive carrier group.

To evaluate the natural evolution of the HBV viral quasispecies, 23 of the 76 CHB patients, all of whom had remained untreated during a period of at least 12 months, were selected. Samples obtained after this treatment-free period (that is, before any treatment was initiated) were analysed as a separate group. In total, 21 of these patients were later treated with one of the study drugs, and samples from these patients at the time of non-response to therapy were included in the patient samples for the various treatment groups.

Overall, 64 patients had received treatment or were under treatment, as sequential monotherapies in all cases. A total of 35 patients were HBeAg-positive, and 6 (17%) of them had a total loss of HBeAg during 3TC therapy. Both the HBeAg-positive and -negative patients had documented CHB and persistent HBV DNA serum levels ≥4 log₁₀ copies/ml over the entire study period. The 22 patients treated with IFN had been selected because all were non-responders to this therapy. Although all of these patients received various therapies after IFN, the subsequent treatments were analysed in only 5 cases; the remaining 17 patients were analysed only after IFN treatment. A total of 26 patients received only 3TC, and 8 were treated only with ADV. Another 13 patients were sequentially treated: 4 treated with IFN followed by 3TC, 1 treated with IFN followed first by 3TC and then by ADV, and 8 treated with 3TC followed by ADV. Thus, a total of 22 treatment periods corresponding to the IFN period, 39 corresponding to 3TC and 17 corresponding to ADV were analysed. The serum sample for the IFN treatment group was taken

Figure 1. Core protein sequence from the consensus sequence of genotype A
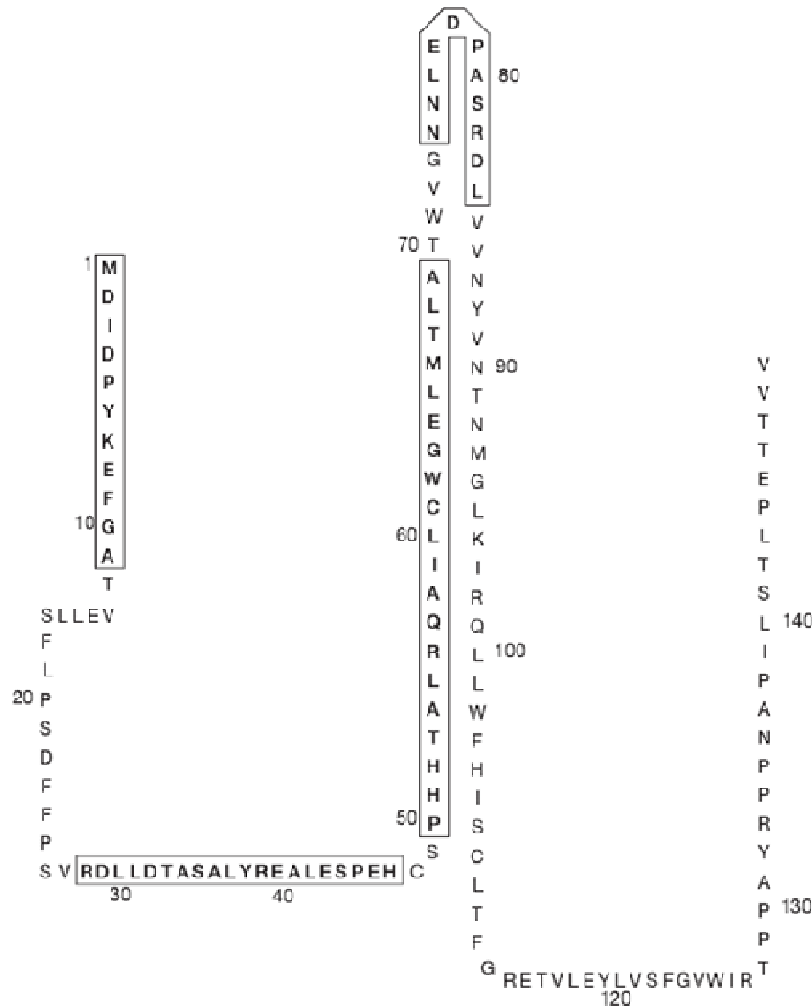
The four regions analysed (AA1–11, Th28–47, Th50–69 and B74–84) are shown in boxes.

6 months after starting therapy, and the samples for 3TC and ADV were drawn at ≥12 months after starting therapy or when a lack of response was evident. IFN-treated patients received IFN-α2b (Intron-A® Schering-Plough, Kenilworth, NJ, USA) at 5 million units/day for 24 weeks or pegylated IFN (PegIntron®; Schering-Plough) at 180 µg/week for 24 weeks, depending on the year when therapy started. Patients with nucleoside/nucleotide analogues were treated with 3TC 100 mg/day (Zeffix®; Glaxo Wellcome UK, Ltd, Uxbridge, UK) or ADV 10 mg/day (Hepsera®; Gilead Laboratories, Foster City, CA, USA). The main demographic,

biochemical, virological and histological data of the study patients are shown in Table 1.

HBV core region analyses

Four HBV core regions were analysed longitudinally. These included three epitopic regions of HBcAg, Th28–47, Th50–69 and B74–84 [12,13,20–23], and the conserved region, AA1–11, in the N-terminal region of HBcAg [15] (Figure 1). A total of 185 serum samples were studied. Baseline sequences were compared with consensus sequences established for each genotype by alignment of 185 complete GenBank sequences of HBV

M Homs et al.

Table 1. Demographic, biochemical, serological, virological and histological data of 4 inactive HBsAg carriers and 76 patients with CHB infection at baseline, and after the various periods studied

Characteristic	Baseline (n=76)	IC (n=4)	TF (n=23)	IFN (n=22)	3TC (n=39)	ADV (n=17)
Median age, years (range)	40 (15–70)	46 (42–53)*	43 (24–62)*	40 (24–66)*	44 (17–71)*	47 (27–66)*
Male sex, n (%)	61 (80)	2 (50)	17 (74)	19 (86)	31 (80)	15 (88)
Liver biopsy[b]						
CAH, n	–	3	22	6	8	5
LC, n	–	1	6	–	5	2
HCC, n	–	–	–	–	–	1
Genotype						
A, n (%)	3 (80)	13 (57)	37 (49)	8 (36)	22 (56)	10 (60)
D, n (%)	1 (20)	10 (43)	38 (50)	14 (64)	17 (44)	6 (35)
F, n (%)	0	0	1 (1)	0	0	1 (5)
HBeAg-positive, n (%)	35 (46)	0	12 (52)	4 (18)	10 (26)[c]	9 (53)
PC variants						
WT, n (%)	2 (50)	11 (48)	34 (45)	7 (32)	24 (61)	10 (59)
MT, n (%)	2 (50)	11 (48)	39 (51)	14 (63)	14 (36)	6 (35)
WT/MT, n (%)	–	1 (4)	3 (4)	1 (5)	1 (3)	1 (6)
Median ALT, IU/ml (range)	97 (61–700)	25.7 (14–40)	105 (54–452)	82 (53–513)	53 (45–280)	61 (47–200)
Median HBV DNA, log₁₀ copies/ml (range)	6.9 (4–12)	3.3 (3–4)	5.8 (4–8)	5.9 (5–8)	6.1 (4–9)	5.3 (4–7)
Median time of infection or treatment, months (range)	39 (12–108)	132 (84–204)	14 (12–30)	6 (6–12)	12 (12–20)	17 (12–18)

[a]Age at the beginning of the period. [b]Liver biopsy performed during the period. [c]Six (17%) cases lost hepatitis B e antigen (HBeAg) after lamivudine (3TC) treatment. ADV, adefovir dipivoxil; ALT, alanine aminotransferase; CAH, chronic active hepatitis; CHB, chronic hepatitis B; HBsAg, hepatitis B surface antigen; HCC, hepatocellular carcinoma; IC, inactive carrier; IFN, interferon; LC, liver cirrhosis; MT, mutated form of main precore mutation (A1896); PC, precore region; WT, wild-type form of main precore mutation (G1896); WT/MT, detection of both wild-type and mutation in main precore mutation (G/A1896); TF, treatment-free period.

genotypes A to H. Only sequences obtained from the genotypes identified in the study patients (A, D and F) were used in this analysis. Sequences obtained at the beginning and at the end of each study condition period in each patient (treatment-free, following IFN, 3TC or ADV treatment, and inactive carriers) were compared.

To enable proper comparisons, the average number of AA changes in each region and study condition period were standardized to 12 months (Av12) as follows: number of AA changes divided by months in the period and multiplied by 12. For the baseline sequences, the known time of infection to baseline was used in the Av12 calculation. The standardized results were also averaged for each group. Baseline samples and the 12-month treatment-free samples were used as a measure of the natural evolution of the virus. The rationale for this standardization step was based on the knowledge that viral evolution is a time-dependent process, and in this way more comparable results could be obtained. For example, an observation of 3 changes in 6 months is not the same as 3 changes in 12 months, and patients were treated for different periods of time. The selected patients receiving IFN were treated for only 6 months (as was recommended at that time), whereas samples from the remaining patients were collected at ≥12 months. Standardization to 12 months was done to enable observation

of AA substitutions even in periods when less variation was expected (for example, in the treatment-free periods) and in the most conserved regions. The Av12 changes were statistically analysed for each region and study condition period (treatment-free, following IFN, 3TC or ADV treatment, and inactive carriers). Results are expressed as the mean and range.

Serological and virological testing

HBsAg, HBeAg and anti-HBe were determined by EIA (Ortho Clinical Diagnostics, Amersham, Buckinghamshire, UK). HBV copy number was quantified by real-time PCR using the LightCycler system (Roche Diagnostics, Mannheim, Germany) with a detection limit of 5×10^2 copies/ml, as previously described [24]. HBV genotyping was performed by sequence alignment, as previously described [25].

Amplification and sequencing of the HBV core region

HBV DNA was extracted from serum by QIAamp microspin columns (QIAamp DNA Mini Kit, Quiagen, Hilden, Germany), according to the manufacturer's instructions. A 760 bp DNA fragment of HBV, located in the precore-core region of the HBV genome, was amplified by nested PCR using specific primers. First PCR: sense primer, nucleotides 1662–1681; 5'-(C/T)ATAAGAGGACTCTTGGACT-3'

and anti-sense primer, nucleotides 2912–2931; 5'-TGTTCCCA(A/G)GAATA(A/T)GGTGA-3'. Nested PCR: sense primer, nucleotides 1800–1822; 5'-CCC(A/C)GTAAA(A/G)TT(C/T)CCCACCT-TAT-3' and anti-sense primer, nucleotides 2540–2559; 5'-GGCTGTAGGCATAAATTGGT-3'. This fragment includes nucleotide 1896 of the precore region, in which a change from G to A abolishes HBeAg expression.

Sequencing was performed with the Big Dye Terminator version 1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and nested primers. Products were analysed with the ABI Prism 310 automatic genome analyzer (Applied Biosystems) and the consensus sequence of each patient was obtained.

Statistical analyses

Comparisons between quantitative variables were performed using the Student's *t*-test or Mann–Whitney U test (between groups), and the Wilcoxon signed-rank test (within group). Qualitative variables were compared with the Fisher's exact test and McNemar's test. Differences were considered statistically significant at P-values of <0.05. We also examined the possible significance of HBV genotypes and mutant forms of the precore region. All statistical analyses were performed with SPSS version 15.0 for Windows (SPSS Inc., Chicago, IL, USA)

Results

A total of 185 sequences were processed. The changes observed in each codon of the four core regions of the six groups analysed (baseline, treatment-free, after IFN, 3TC, or ADV, and inactive HBsAg-positive carriers) were standardized to 12 months (Av12). After standardizing, the mean number of changes in the four regions studied for each group was calculated, with the following results: 3.4 in the CHB baseline samples, 1.5 in the treatment-free period, 7.6 after IFN treatment, 2.3 after 3TC treatment, 3.2 after ADV treatment and 3.5 in inactive carrier patients (Figure 2).

AA analyses in baseline serum samples

Sequences from the 76 CHB patients obtained at baseline were analysed to assess the prevalence of AA substitutions in the three epitopic regions and conserved AA1–11 region. AA changes were compared with consensus sequences, relative to the known time of infection. The percentages of CHB cases with changes in epitopic regions (37% Th28–47, 55% Th50–69 and 59% B74–84) were higher than the percentage in the conserved region (9% AA1–11; $P<0.001$; Figure 3A). The Av12 changes were also lower in AA1–11 than in the three epitopic regions

($P<0.001$), with the minor Th28–47 region being less variable than the two immunodominant epitopes ($P<0.001$; Figure 3B). Six of the seven most variable codons (Av12≥0.1) were clustered in the immunodominant epitopes (Figure 2). Regarding HBeAg status, a higher number of Av12 changes in Th50–69 were observed in HBeAg-negative cases than in positive cases (mean [range] 1.34 [0–6] versus 0.49 [0–4]; $P=0.04$) and a trend towards more substitutions in negative cases was seen for B74–84 (mean [range] 1.21 [0–6] versus 0.8 [0–6]; $P=0.06$). No differences in Av12 values were found between genotypes A and D. Regarding the presence (mutant) or absence (wild type) of the main precore mutation G1896A, a significant difference in Av12 was observed in the B74–84 region (mean [range] 0.43 [0–2] wild type versus 1.35 [0–6] mutant; $P=0.02$) and a trend approaching significance was observed in the Th50–69 region (mean [range] 0.7 [0–6] wild type versus 1.13 [0–4] mutant; $P=0.07$). In both cases, the highest Av12 value corresponded to mutant cases. Patients with changes in the B74–84 region showed higher HBV DNA levels (mean [range] 6.9 [4–12] $\log_{10}$ copies/ml) than those without alterations in this region (mean [range] 5.8 [4–9] $\log_{10}$ copies/ml; $P=0.04$).

AA changes and their relationships with absence of treatment, antiviral therapy and inactive HBsAg carrier status

The 64 treated CHB patients (including 22 IFN-treated, 39 3TC-treated and 17 ADV-treated samples), 23 treatment-free period samples, and 4 inactive HBsAg-positive carriers were included in this analysis. Baseline and treatment-free samples were used to examine the natural evolution (that is, without drug administration) of the viral quasispecies during CHB infection. Percentages of cases with changes and statistical significance in the comparisons between groups and regions are shown in Figure 3A. In a significant percentage of treatment-free samples, changes were selected in the epitopic regions, mainly in the two immunodominant ones (52% in Th50–69 and 56% in B74–84), at values similar to those observed at baseline and after 3TC or ADV treatment, but lower than those recorded after IFN therapy (72% in Th50–69 and 86% in B74–84). Statistical differences between treatment-free and IFN cases were only found in B74–84 ($P=0.04$); however, the percentage of changes in the minor epitope Th28–47 was higher after 3TC and ADV (36% for 3TC and 47% for ADV) than in the treatment-free group (9%; $P=0.03$ and $P=0.01$, respectively). Percentages of changes in B74–84 were significantly higher after IFN therapy than after 3TC or ADV ($P=0.01$ and $P=0.03$, respectively).

Analysis after standardization to 12 months showed that in the natural evolution of the viral quasispecies

M Horns et al.

Figure 2. Av12 for each position at baseline and in each study condition period
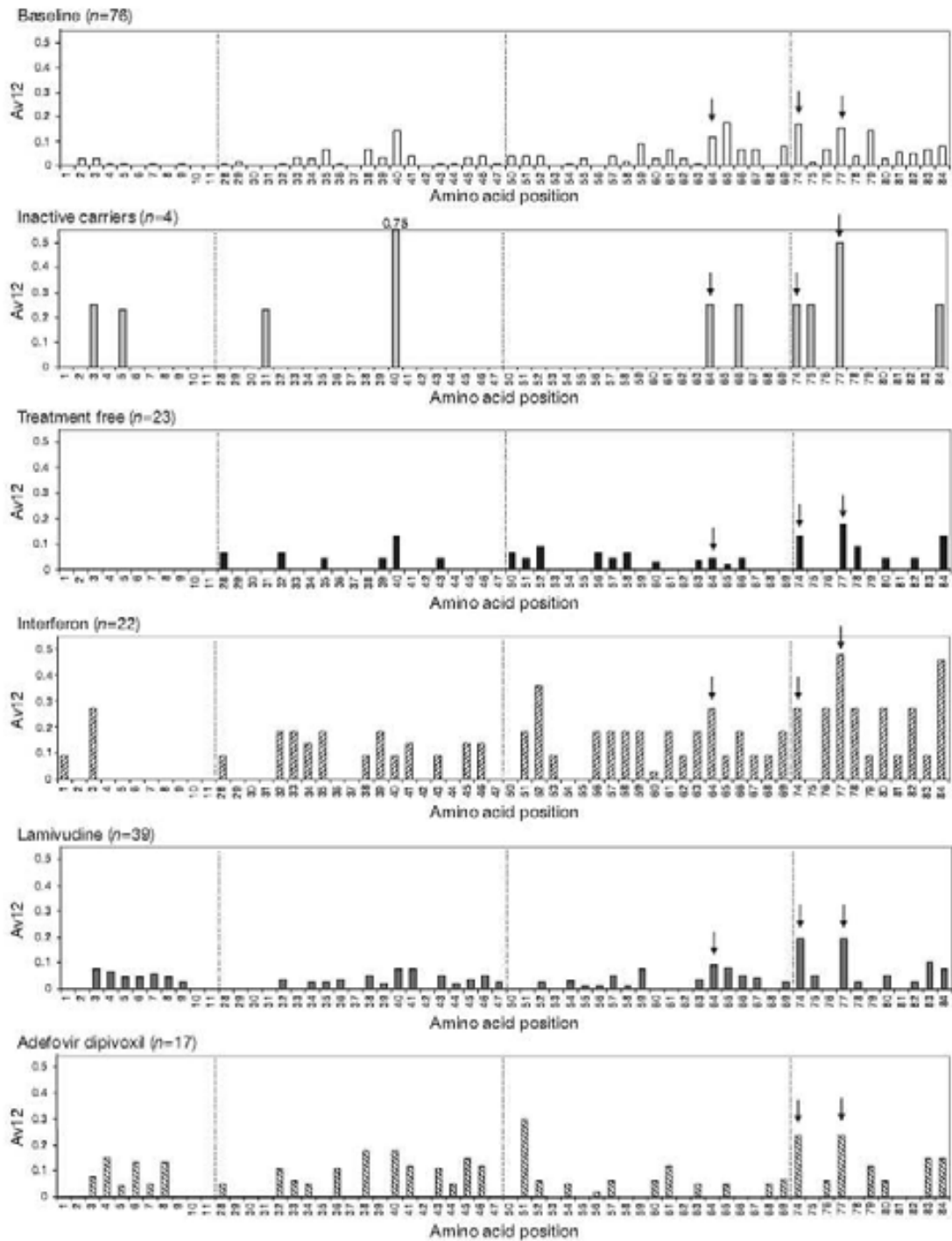
Arrows indicate the main amino acid position as described in *Results*. Av12, the average number of amino acid substitutions standardized to a 12-month period.

56

(treatment-free group), the AA1–11 region was conserved (no substitutions were found), whereas in the three epitopic regions, AA substitutions were selected. Differences between the AA1–11 region and the Th50–69 ($P=0.01$) and B74–84 ($P=0.001$) regions were significant. In relation to antiviral therapy, Av12 values after IFN treatment indicated substantial conservation of AA1–11 compared with the other three regions ($P<0.05$). Differences between Th28–47 and Th50–69 were also observed ($P<0.05$), but not between Th50–69 and B74–84. After 3TC treatment, a large number of substitutions were detected in the B74–84 epitope, with significant differences compared with the other three regions ($P<0.05$). After ADV treatment, there were no differences between Av12 changes for the four regions studied, although a difference approaching significance was found between B74–84 and AA1–11 ($P=0.07$). When the three antiviral therapies were compared (Figure 3B), the Av12 was higher in the immunodominant Th50–69 and B74–84 epitopes following IFN than after 3TC or ADV therapy ($P<0.05$). Similar results

were found when the IFN group was compared to the treatment-free group ($P=0.001$). After ADV treatment, the Av12 was significantly increased in the minor immunodominant Th28–47 region in comparison with the treatment-free samples ($P=0.04$; Figure 3B).

In the comparison between inactive carriers and CHB patients, differences approaching significance were found in the immunodominant regions Th50–69 ($P=0.05$) and B74–84 ($P=0.06$) between samples corresponding to IFN therapy, and in the Th28–47 region only in the treatment-free samples ($P=0.06$). A total of 9 of the 23 treatment-free patients presented reactivation episodes; however, no significant differences were found in the pattern or number of AA changes relative to patients in whom reactivations were not observed.

Patient-by-patient comparison of AA changes between the treatment-free period and following IFN
To confirm a greater selection of AA changes in the two immunodominant epitopes during IFN therapy, a patient-by-patient comparative analysis of these

Figure 3. Percentage of cases with amino acid changes and Av12 at baseline and in different study condition periods
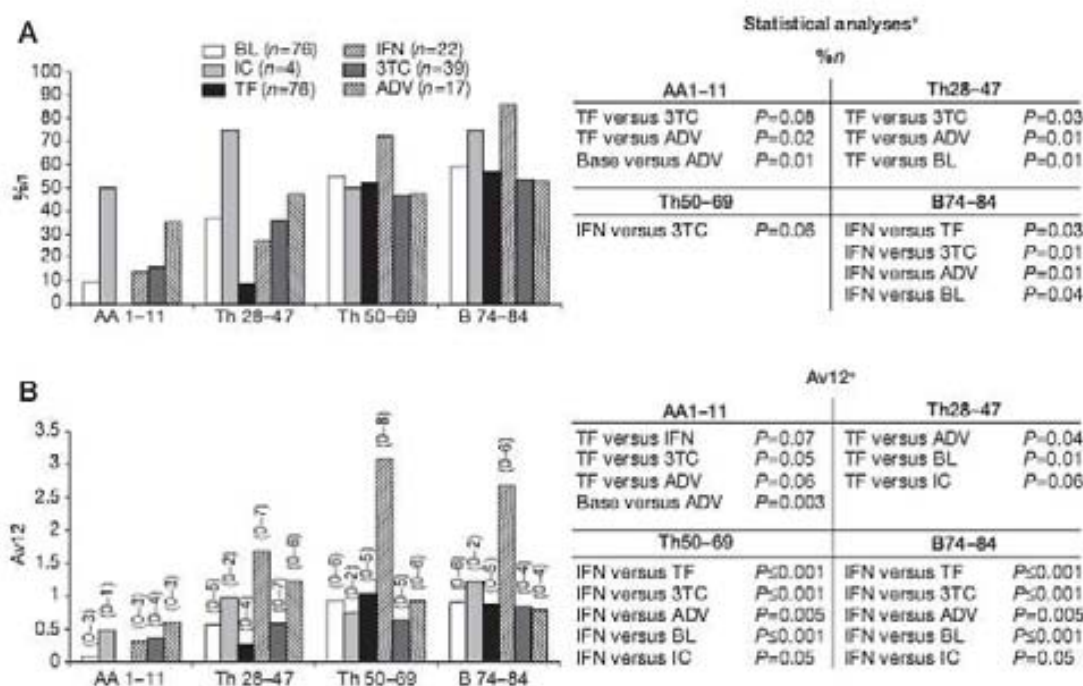
(A) Percentage of cases with amino acid changes (%n) and (B) average number of amino acid substitutions standardized to a period of 12 months (Av12) with range of values shown in parenthesis above the bars for the four hepatitis B core antigen regions analysed. Tables on the right show significant results obtained from the treatment comparison in each region. [a]P-values determined by the Mann–Whitney U test. ADV, adefovir dipivoxil; BL, baseline sample; IC, inactive carrier; IFN, interferon; TF, treatment-free period; 3TC, lamivudine.

57

regions was performed in 16 patients who were treatment-free for ≥12 months before starting IFN. The substitutions observed in the absence of treatment were compared with those detected after IFN therapy (Figure 4). As was seen in the overall longitudinal study, this analysis showed significant differences in the Av12 number of AA changes between the IFN-treated and treatment-free period in the Th50–69 ($P=0.002$) and B74–84 ($P=0.01$) epitopes. Differences were also significant relative to baseline ($P=0.005$ for Th50–69 and $P=0.04$ for B74–84).

Most common AA changes and their relationships with absence of treatment, antiviral therapy and inactive HBsAg carrier status

In the total population, eleven codons showed ≥10 absolute AA substitutions. The distribution of these changes and their stability in the various study conditions are shown in Table 2. More than one type of substitution was observed in each position, with codon 74 being particularly polymorphic (V/N mainly changed to A but also to D, K, H or I). The largest numbers of substitutions were in codon 77, with 36 changes, and codon 74, with 33 changes. A total of 31 of the 87 substitutions observed in these positions at baseline were seen to persist after the various study periods (three after the treatment-free period, IFN, and lastly, 3TC therapy). In the sequential study, 103 AA changes were selected: 16 (mean 0.8) after the treatment-free period, 10 of which remained after treatment (the N74A change was stable after IFN, 3TC and ADV therapy), 29 (mean 1.3) after IFN, 36 (mean 0.9) after 3TC and 22 (mean 1.3) after ADV.

The most prevalent AA substitutions in CHB patients were 24 E77Q, 13 N/V74A and 13 E64D; T67N was also observed in 6 (46%) of the patients with E64D changes. In 15 positions in the regions analysed, the differences in Av12 values between groups were significant or approached significance (Table 3). Interestingly, no differences in the highly variable codon 74 were observed. All positions in which Av12 changes were higher after IFN therapy than in the other study conditions were located in the main epitopic regions, Th50–69 and B74–84. Of note, the Av12 value for codon 84 (substitutions L84Q/R/H) was higher after IFN therapy (Av12=0.46) than in the other groups. Nucleotide substitutions that were significantly more common after ADV therapy than in the other groups were mainly located in the minor Th28–47 epitope (codons 38, 41, 45 and 46), or in the conserved AA1–11 region (codon 3 and 4). Only one position in the main epitopic regions (codon 51) was more variable following ADV (Av12=0.29) than 3TC (no changes; $P=0.04$). Changes in position 59 were the only ones found to be more common after 3TC than after the treatment-free period ($P=0.05$).

Relationships between emergence of AA substitutions and baseline precore variant at position 1896, HBV genotype and HBeAg status

Only one CHB patient showed changes in precore variant status: wild type was selected from a baseline wild type/mutant mixture after IFN treatment. With regard to HBV genotype, genotype A was selected in one patient after IFN therapy from a baseline A/D genotype mixture. No relationship between baseline detection of the main precore mutation (G1896A) and selection of changes in the four core regions was observed after the treatment-free period, IFN or ADV therapy. However, wild-type precore cases under 3TC treatment showed a significantly higher Av12 value in Th50–69 than mutant cases (mean [range] 0.72 [0–3] versus 0.18 [0–2]; $P=0.03$) and a nearly significant Av12 value in B74–84 (0.95 [0–4] versus 0.35 [0–2]; $P=0.08$). No associations were found with emergence of HBV substitutions for HBV genotype, HBeAg baseline status or HBeAg seroconversion in the different groups studied.

Discussion

An increase in the number of AA substitutions in HBcAg during chronic HBV disease has been described, mainly clustering in the CD4+, CD8+ and B-cell immunodominant epitopes [17,18]. Although these variants are associated with persistence of chronic infection and sensitivity to IFN therapy, their real significance remains controversial [7,17,21,23,26–28], and their distribution does not show specific patterns [3,18]. The present study investigated the natural occurrence and selection of HBV core variants in treatment-free CHB patients and those under treatment (IFN, 3TC or ADV) as the external evolutionary factor. The effect of IFN has been reported previously [17,24,29], but there are no studies investigating core variants in patients receiving nucleoside/nucleotide analogues. In the present study, HBV core substitutions were examined while active viral replication was occurring. Hence, possible associations of core nucleotide changes with response to antiviral therapy were not determined.

The study focused on three epitopic regions, including immunodominant Th50–69 and B74–84, and minor Th28–47, as well as the AA1–11 region. The Th18–27 epitope [30,31], which is also located in the 760 bp precore-core region analysed, was not studied because of the small number of changes detected (only seven cases; MH *et al.*, data not shown). The baseline analysis detected a high prevalence of HBV core variants, particularly in the immunodominant regions (>50% of cases), concurring with previous reports showing a high rate of core substitutions in HBeAg-negative cases, mainly in the Th50–69 region [16,25,28,29,32], and in cases with the main precore variant [18,23,27].

Figure 4. Patient-by-patient analysis of 16 individuals who were sampled after TF and IFN treatment

'The numbered columns correspond to amino acid positions with changes; the reference sequence is shown in the row below and amino acid changes in relation to the reference sequence are indicated in a single letter code. Dots indicate that the amino acid in this position is the same as in the reference sequence. 'Time in months. BL, sample obtained at baseline; Gen, HBV genotype; HBe, hepatitis B e antigen; IFN, sample obtained after 6 months of interferon treatment; MT, presence of main precore mutation; N, negative; P, positive; PC, main precore mutation (G1896) status; TF, sample obtained after the treatment-free period; WT, absence of main precore mutation; WT/MT, detection of both wild-type and mutation in main precore mutation.

M Homs et al.

Table 2. Amino acid positions with the highest variability observed in this study

Core position[a]						Amino acid change					
Wild type	Y38	E40	H51	H52	I59	E64	M66	N/V74	E77	P79	L84
Substitution	F/S	D/Q	Y/L/Q/N	L/P/Q	V/T/F	D/S	T/N/I	A/D/K/H/I	Q/N/K	Q/T/S	Q/R/H
Changes detected in one period											
Baseline	4	9	3	2	4	7	2	9	7	6	3
TF	–	–	–	–	–	–	1	2	–	1	2
IFN	1	1	2	2	2	3	2	3	5	1	5
3TC	2	1	–	1	2	4	2	7	8	–	3
ADV	3	3	5	1	–	–	–	4	4	2	–
Changes retained during consecutive periods[a]											
Baseline-3TC	–	2	–	–	3	2	1	3	3	3	2
Baseline-TF	–	–	–	–	–	–	–	–	2	–	–
Baseline-TF-IFN	1	2	–	–	–	–	1	1	–	–	1
Baseline-TF-IFN-3TC	–	–	–	1	–	–	1	–	–	–	1
Baseline-TF-3TC	–	–	–	–	–	–	–	–	–	1	–
TF-IFN	–	–	1	1	–	1	–	–	2	–	1
TF-IFN-3TC-ADV	–	–	–	–	–	–	–	1	–	–	–
TF-3TC	–	1	–	1	–	–	–	–	1	–	–
IFN-3TC	–	–	–	1	–	–	–	–	1	–	–
3TC-ADV	–	2	–	–	1	–	–	2	1	–	–
Total changes											
CHB	11	21	11	10	12	17	10	32	34	14	18
IC	–	3	–	–	–	1	1	1	2	–	1

[a]For each position the wild-type form is shown and all substitutions are indicated in the row showing change; the underlined amino acid indicates the most frequent change (for example, for E77Q/N/K, the most frequent change is E77Q). [a]Changes were detected in the first period indicated in the row of the first column and retained through the consecutive periods. For example, two E40D changes were observed at baseline and remained in the treatment-free period (TF) and the period under interferon (IFN) treatment (cases 13 and 15 in Figure 3), and one N74A change was selected during TF and remained after subsequent IFN, lamivudine (3TC) and adefovir dipivoxil (ADV) treatments (case 4 in Figure 3). CHB, chronic hepatitis B; IC, inactive carrier.

Table 3. Amino acid positions with significant differences in Av12 between the different study condition periods[a]

Codon position	First sample		Second sample		P-value[b]
	Period 1	Av12	Period 2	Av12	
84	TF	0.012	IFN	0.46	0.05
59	TF	0.04	3TC	0.09	0.05
3	TF	0	ADV	0.08	0.05
4	TF	0	ADV	0.15	0.02
41	TF	0	ADV	0.12	0.04
45	TF	0	ADV	0.15	0.04
46	TF	0	ADV	0.12	0.04
52	IFN	0.36	3TC	0.03	0.03
64	IFN	0.27	3TC	0.07	0.04
76	IFN	0.27	3TC	0.03	0.03
77	IFN	0.48	3TC	0.19	0.05
84	IFN	0.46	3TC	0.08	0.05
4	IFN	0	ADV	0.15	0.02
84	IFN	0.46	ADV	0	0.04
40	IFN	0.09	IC	0.75	0.01
4	3TC	0.06	ADV	0.15	0.06
38	3TC	0.05	ADV	0.18	0.05
51	3TC	0	ADV	0.29	0.03
31	3TC	0	IC	0.23	0.02
84	ADV	0	IC	0.25	0.04

[a]Period 1 versus period 2. [b]Mann–Whitney U test. ADV, adefovir dipivoxil; Av12, the average number of amino acid substitutions standardized to a 12-month period; IC, inactive carrier; IFN, interferon; TF, treatment-free period; 3TC, lamivudine.

All these findings suggest enhanced immune pressure against core epitopes by HBeAg loss, which we will discuss later. Interestingly, HBV DNA levels were higher in patients with AA changes in B74–84 than in those without, in keeping with the findings of other authors [32,33], but contrasting with a recent study by Sendi et al. [31], in which AA substitutions were more frequent at low HBV DNA levels, as well as in inactive HBsAg-positive carriers. We suggest that the presence of core AA changes enables HBV to evade immune clearance, thereby lengthening the life of the infected hepatocyte and increasing the viral population. This would provide the possibility of infecting more hepatocytes and ultimately raise the serum HBV DNA level, as occurred in our patients.

The natural evolution of the HBV core gene, evaluated in treatment-free serum samples, yielded the highest variability in the immunodominant epitopes and greatest conservation of the AA1–11 region, as has been described previously [15]. Interestingly, the percentage of changes in AA1–11, Th28–47 and B74–84 observed in the inactive carrier group was higher than in untreated CHB patients, in accordance with a recent report [31]. AA substitutions in epitopic regions in both untreated and inactive patients indicate persistent immune activity and suggest a viral escape mechanism [16,17,28], but the host response seems to differ in these two patient populations. Nonetheless, because of the small number of inactive carriers included in this study, these results should be interpreted with caution.

Regarding the effect of antiviral therapy on core gene evolution, we found a tendency to conserve the AA1–11 region, although variability was slightly higher than at baseline and in treatment-free periods. The percentage of changes in the immunodominant Th50–69 and B74–84 epitopes was higher after IFN therapy than in the remaining treated groups, in keeping with the expected immune-stimulating effect of IFN [34]. Core substitution selection during IFN therapy has been reported previously [17,24,29]; however, in the present study, we compared IFN-treated and treatment-free periods, as well as the treatment-free and IFN-treated period in each patient for whom the two measures were available. In both analyses, immunodominant regions under IFN stimulus showed higher variability than the other combinations of regions and conditions studied.

An increase in spontaneous or IFN-induced HBV core sequence diversity has also been reported after HBeAg seroconversion [28]. Among our IFN-treated patients, 80% were anti-HBe-positive; thus, the high rate of AA changes in core immunodominant epitopes observed after IFN therapy is consistent with these reported results. Apparently, changes selected under IFN can evade detection and elimination by the immune system, thereby allowing HBV escape and maintaining

active infection. We found no differences in the rate of AA changes between HBeAg-positive and anti-HBe-positive patients after IFN treatment, in contrast to the findings of Lim et al. [28]. This discrepancy might be related to the differing ethnic origin, viral genotypes, small number of seroconverters and the fact that only non-responders were included in our study.

The antiviral effect of nucleoside/nucleotide analogues is based on inhibition of viral replication by targeting HBV reverse transcriptase activity. No immunomodulator activity has been described. Thus, similar results in immunodominant epitopes were expected between untreated patients and those treated with 3TC or ADV. Surprisingly, minor Th28–47 epitope variability after nucleoside/nucleotide analogue treatment was higher than in the treatment-free period, suggesting stimulation of T-helper CD4+ activity against this epitope. A possible explanation for these results might be related to the effect of nucleoside/nucleotide analogue antiviral treatment, which decreases retrotranscriptase activity, but not pregenomic HBV RNA production or viral antigen expression. This could result in an accumulation of immature core particles, which might stimulate the proteosome, increasing production of viral epitopes and their expression on the hepatocyte membrane, and ultimately increasing immune pressure against these cells. We suggest that changes detected in Th28–47 of the core region are selected during nucleoside/nucleotide analogue treatment and could confer an escape mechanism from immune pressure; however, the reason why this theoretical phenomenon was more evident in Th28–47 than in the immunodominant epitopes remains unexplained.

HBcAg and HBeAg are highly cross-reactive at the T-cell level: the immunodominant loop, which includes the B74–84 epitope, is involved in the antigenicity of both of these proteins [9]. It is known that HBeAg can contribute to the outcome and pathogenesis of HBV infection, inducing tolerance in HBcAg-specific T-cells and reducing their potential to kill HBV-infected cells [35]. In this regard, HBeAg upregulates toll-like receptor expression in CD4+ CD25+ T-regulatory cells, an action associated with inhibition of the T-cell response [36,37]. This factor, and the existence of cross-reactive HBeAg/HBcAg epitopes would facilitate viral persistence [38]. It can be hypothesized that after the main precore variant selection (G1896A), the HBeAg immune-tolerant effect is lost and HBcAg epitopes expressed on the hepatocyte surface come under the pressure of the immune system. Thus, a higher number of HBV core substitutions would be detected in patients with a precore variant, as was seen at baseline and reported previously [18,23,27]; however, these results were not observed in the sequential analysis of the treatment-free, IFN and ADV periods.

Moreover and surprisingly, after 3TC therapy, the numbers of AA substitutions in the immunodominant Th50–69 and B74–84 regions were higher in wild-type cases than in precore mutant cases. Our findings could explain the reported cases of reselection of wild-type sequences after 3TC therapy [39] as the result of preferential immunoselection of precore wild-type sequences because they have a higher number of core substitutions than sequences with precore variants.

After IFN therapy, the most frequent AA substitutions were in the immunodominant regions, whereas after ADV, they were similarly distributed between immunodominant and minor Th28–47 regions. Also, specific changes in the minor Th28–47 epitope (codons 41, 45 and 46) were seen more often after ADV therapy than in the treatment-free period. Both these findings suggest differential immune stimulation during ADV therapy, more specifically associated with this minor epitope. Despite the finding that the substitution patterns detected were not specific, some changes were especially interesting. The E64D substitution in the Th50–69 region is notable, being observed in all the periods analysed except after ADV therapy. This change significantly reduces T-cell proliferation *in vitro* when linked to T67N, altering the $E_{64}LMT_{67}$ sequence, a common motif in T-cell epitopes [16]; thus, it is selected by the effect of immune pressure. In our patients, the double E64D-N67T change was observed in 46% of patients with an E64D substitution.

The main variants observed in the B74–84 region over the entire study were N74A/D and E77Q, which were similarly distributed between the baseline and sequential groups, indicating that their selection was not the result of specific evolutive pressures. Both codons are highly polymorphic [15], and their variants are often detected in CHB and inactive carrier cases [17,19,23,29,40]; however, changes in codon 77 have been associated with the HLA-B*4001 allele [41] and might be involved in viral escape mechanisms. Large studies are needed to determine the effect of these variants on immunological reactivity against HBV.

Interestingly, in the present report, selection of changes in codon 84 (mainly L84Q) was associated with IFN therapy. Although highly polymorphic positions, such as 74, 77 and 84 [15] have been observed with IFN therapy [17,29], there are no studies investigating their status after other therapies or the absence of treatment.

Despite the small number of HBsAg-positive inactive carriers, strong selection of E40D/Q was found in these patients, which were at levels similar to those observed at baseline in a recent study [31], suggesting differential immunological activity in the development of inactive HBsAg-positive carrier status and CHB status: immune activity specifically addressed to minor T-helper epitopes in cases of inactive HBsAg-positive carriers and activity addressed to changes in immunodominant epitopes in CHB. However, this observation must be interpreted with caution because of the small number of inactive HBsAg-positive carriers studied. Lastly, although AA1–11 has been described as a conserved region, some positions showed remarkable variability, particularly codons 3 and 4. The finding that region AA1–11 partially overlaps with the T-helper CD4+ epitope 1–25 [40] suggests that there might be some immunological pressure against this epitope, especially after ADV treatment.

The most important limitation of this study is the small number of inactive HBsAg carriers included. It was difficult to recruit eligible patients for this group because their low viral loads made HBV core analysis impossible. Another limitation involves the analysis of HBV evolution by using direct sequencing techniques, rather than more sensitive methods, such as clonal analysis. Nonetheless, the significant results obtained can be used as the basis for designing a clonal study with ultra-deep pyrosequencing methods focused on the most significant regions. It would be of particular interest to perform studies directed to specific positions, such as E64D-N67T, L84D (associated with IFN treatment) or to the highly polymorphic positions 74, 77 and 84.

In conclusion, this study reports that immunodominant regions Th50–69 and B74–84, as well as certain specific AA substitutions appear to be predominantly involved in natural and antiviral therapy-induced evolution in the HBV core protein. More changes occurred under the immune-stimulating effect of IFN than in treatment-free periods or during nucleoside/nucleotide analogue therapies. Some immune-stimulating activity, likely related to maintenance of viral protein synthesis, was associated with nucleoside/nucleotide analogues. This activity was mainly evident in the minor Th28–47 epitope, as was seen in inactive HBsAg-positive carriers.

Acknowledgements

Disclosure statement

The authors declare no competing interests.

References

1. Locarnini S, McMillan J, Bartholomeusz A. The hepatitis B virus and common mutants. *Semin Liver Dis* 2003; 23:5–20.

2. Seeger C, Mason WS. Hepatitis B virus biology. *Microbiol Mol Biol Rev* 2000; 64:51–68.

3. Pumpens P, Grens E, Nassal M. Molecular epidemiology and immunology of hepatitis B virus infection. *Intervirology* 2002; 45:218–232.

4. Bertoletti A, Gehring AJ. The immune response during hepatitis B virus infection. *J Gen Virol* 2006; 87:1439–1449.

5. Chisari FV, Ferrari C. Hepatitis B virus immunopathogenesis. *Annu Rev Immunol* 1995; 13:29–60.

6. Singh R, Kaul R, Kaul A, Khan K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J Gastroenterol* 2007; 13:1770–1787.

7. Milich DR, McLachlan A, Moriarty A, Thorton GB. Immune response to HBV core antigen (HBcAg): localization of T cell recognition sites within HBcAg/HBeAg. *J Immunol* 1987; 139:1223–1231.

8. Barros MF, Rodrigues PA, Matias SR, Carreira LM, Machado-Caetano JA. Detection of hepatitis B virus core mutants by PCR-RFLP in chronically infected patients. *J Viral Methods* 2004; 120:125–130.

9. Belnap DM, Watts NR, Conway JF, *et al.* Diversity of core antigen epitopes of hepatitis B virus. *Proc Natl Acad Sci U S A* 2003; 100:10884–10889.

10. Chisari FV. Rous-Whipple Award Lecture. Viruses, immunity, and cancer: lessons from hepatitis B. *Am J Pathol* 2000; 156:1117–1132.

11. Pumpens P, Grens E. HBV core particles as a carrier for B cell/T cell epitopes. *Intervirology* 2001; 44:98–114.

12. Kreutz C. Molecular, immunological and clinical properties of mutated hepatitis B virus. *J Cell Mol Med* 2002; 6:113–143.

13. Vanlandschoot P, Cao T, Leroux-Roels G. The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure. *Antiviral Res* 2003; 60:67–74.

14. Harris A, Belnap DM, Watts NR, *et al.* Epitope diversity of hepatitis B virus capsids: quasi-equivalent variations in spike epitopes and binding of different antibodies to the same epitope. *J Mol Biol* 2006; 355:562–576.

15. Chain BM, Myers R. Variability and conservation in hepatitis B virus core protein. *BMC Microbiol* 2005; 5:33.

16. Torre F, Cramp M, Owsianka A, *et al.* Direct evidence that naturally occurring mutations within hepatitis B core epitope alter CD4+ T-cell reactivity. *J Med Virol* 2004; 72:370–376.

17. Radecke K, Protzer U, Trippler M, Herrmann KH, Buschenfelde M, Gerden G. Selection of hepatitis B virus variant with aminoacid substitutions inside the core antigen during alfa interferon therapy. *J Med Virol* 2000; 62:479–486.

18. 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; 24:32–37.

19. Alexopoulou A, Baltayiannis G, Eroglu C, *et al.* Core mutations in patients with acute episodes of chronic HBV infection are associated with the emergence of new immune recognition sites and the development of high IgM anti-HBc index values. *J Med Virol* 2009; 81:34–41.

20. Tillmann HL. Antiviral therapy and resistance with hepatitis B virus infection. *World J Gastroenterol* 2007; 13:125–140.

21. Ferrari C, Bertoletti A, Penna A, *et al.* Identification of immunodominant T cell epitopes of the HBV nucleocapside antigen. *J Clin Invest* 1991; 88:214–222.

22. Salfeld J, Pfaff E, Noah M, Schaller H. Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. *J Virol* 1989; 63:798–808.

23. Carman WF, Boner W, Fattovich G, *et al.* Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive disease and in T helper cell epitopes during clinical remission. *J Infect Dis* 1997; 175:1093–1100.

24. Jardi R, Rodriguez F, Buti M, *et al.* Quantitative detection of hepatitis B virus DNA in serum by a new rapid real-time fluorescence PCR assay. *J Viral Hepat* 2001; 8:465–471.

25. Jardi R, Rodriguez F, Buti M, *et al.* Mutations in the basic core promoter region of hepatitis B virus. Relationship with precore variants and HBV genotypes in a Spanish population of HBV carriers. *J Hepatol* 2004; 40:507–514.

26. Naoumov NV, Thomas MG, Mason AL, *et al.* Genomic variations in the hepatitis core gene: possible factor influencing response to interferon alfa treatment. *Gastroenterology* 1995; 108:505–514.

27. Rodriguez-Frias F, Buti M, Jardi R, *et al.* Hepatitis B virus infection: precore mutations and its relation to viral genotypes and core mutations. *Hepatology* 1995; 22:1641–1647.

28. Lim SG, Cheng Y, Guindon S, *et al.* Viral quasi-species evolution during hepatitis B e antigen seroconversion. *Gastroenterology* 2007; 133:951–958.

29. Fattovich G, McIntyre G, Thursz M, *et al.* Hepatitis B virus precore/core variation and interferon therapy. *Hepatology* 1995; 22:1355–1362.

30. Bertoletti A, Costanzo A, Chisari FV, *et al.* Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope. *J Exp Med* 1994; 180:933–943.

31. Sendi H, Mehrab-Mohseni M, Shahraz S, *et al.* CTL escape mutations of core protein are more frequent in strains of HBeAg negative patients with low levels of HBV DNA. *J Clin Virol* 2009; 46:259–264.

32. Tanaka H, Ueda H, Hamagami H, *et al.* Mutations in hepatitis B virus core regions correlate with hepatocellular injury in Chinese patients with chronic hepatitis B. *World J Gastroenterol* 2005; 11:4693–4696.

33. Mohamadkhani A, Jazii FR, Poustchi H, *et al.* The role of mutations in core protein of hepatitis B virus in liver fibrosis. *Virol J* 2009; 6:209.

34. Perrillo R. Benefits and risks of interferon therapy for hepatitis B. *Hepatology* 2009; 49 Suppl 5:S103–S111.

35. Rehermann B. Immune responses in hepatitis B virus infection. *Semin Liver Dis* 2003; 23:21–38.

36. Zhang Y, Lian JQ, Huang CX, *et al.* Overexpression of Toll-like receptor 2/4 on monocytes modulates the activities of CD4(+)CD25(+) regulatory T cells in chronic hepatitis B virus infection. *Virology* 2010; 397:34–42.

37. Peng G, Li S, Wu W, Sun Z, Chen Y, Chen Z. Circulating CD4+ CD25+ regulatory T cells correlate with chronic hepatitis B infection. *Immunology* 2008; 123:57–65.

38. Guidotti LG, Chisari FV. Immunobiology and pathogenesis of viral hepatitis. *Annu Rev Pathol* 2006; 1:23–61.

39. Suzuki F, Suzuki Y, Tsubota A, *et al.* Mutations of polymerase, precore and core promoter gene in hepatitis B virus during 5-year lamivudine therapy. *J Hepatol* 2002; 37:824–830.

40. Jung MC, Diepolder HM, Spengler U, *et al.* Activation of a heterogeneous hepatitis B (HB) core and e antigen-specific CD4+ T-cell population during seroconversion to anti-HBe and anti-HBs in hepatitis B virus infection. *J Virol* 1995; 69:3358–3368.

41. Abbott WG, Tsai P, Leung E, *et al.* Associations between HLA class I alleles and escape mutations in the hepatitis B virus core gene in New Zealand-resident Tongans. *J Virol* 2010; 84:621–629.

63

Second Study

3 SECOND STUDY: Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome

3.1 Hypothesis and Aims

3.1.1 Introduction

HBV has a highly overlapping genome. In our first study, we evaluated the Core gene, the HBV gene showing the least overlapping and therefore, a region whose sequences can theoretically accumulate changes. For this reason, the Core is suggested to be the most useful region to describe the HBV quasispecies (115). The preCore region contains the encapsidation signal, which is recognized by the polymerase and is essential for viral replication. Mutations in the preCore region have several implications besides inhibition of HBeAg expression. They also stabilize the encapsidation signal, (12,48) thereby enhancing viral replication (55). The preCore region is mainly non-overlapped, and mutations in this region can occur without affecting other HBV genes. Therefore, preCore mutations are of interest because of their clinical and virological implications.

In contrast, the HBV polymerase gene overlaps the surface gene and this may restrict selection of mutations in both these genes. However, in patients under NUC treatment, mutations that confer resistance to antiviral therapy can be selected in the polymerase gene, mainly in the catalytic motif (YMDD) (37,38). The rate of response to treatment is closely related to HBeAg expression: anti-HBe-positive patients have a significantly greater response than HBeAg-positive patients (123). The causes of this differential response are not clear, but one theory is that it may be due to higher rates of viral replication in HBeAg (+) than HBeAg (-) patients; nonetheless, the virological characteristics have not been analyzed. Our aim in the present study was to analyze the preCore region and the polymerase gene from the same viral genome to evaluate possible associations between the main variants in both regions.

Ultra-deep pyrosequencing (UDPS) seems to be the most powerful tool to analyze the variability of viral quasispecies. The main limitation of this technique at the time of the study was that the length of the region to be analyzed was limited to 250 bp. Thus, UDPS analysis of the preCore region and the main motif of the polymerase gene from the same viral genome

was not feasible because the two regions are separated by more than 1 kb. Therefore, we designed a technique to obtain an HBV DNA product containing both regions of interest from the same viral genome in a 250-bp length fragment.

3.1.2 Hypothesis

UDPS technology may be useful to study the HBV quasispecies composition. It is hypothesized that HBV preCore region variability may be determined by the secondary structure adopted by the encapsidation signal. It is also hypothesized that in the same viral genome, preCore variability (main variants associated with HBeAg non-expression) is related with selection of mutations in the main catalytic motif of the polymerase gene.

3.1.3 Aims

1. Establish a technique to analyze the preCore regions and the YMDD motif of the polymerase gene from the same viral genome, so that both regions can be evaluated by UDPS.

2. Analyze the prevalence of mutations in the preCore region and define possible limitations in the secondary structure adopted in the encapsidation signal.

    a) Study base pairing between the amino acid positions of the loop structure.

    b) Study the variability/conservation of the different functional domains of this signal, mainly: the signal bulge, the 4-nucleotide primer target and the acceptor site in the DR1 sequence.

3. Evaluate the rate of mutations in the YMDD motif of the polymerase gene by UDPS.

4. Examine possible simultaneous presence of mutations in the preCore gene and YMDD motif.

3.2   Summary of the study

Attending to the first aim of the study, we designed a technique based on PCR of the HBV genome that included amplification from the preCore region to the YMDD motif of the polymerase (Figure 24). The PCR primers used included a HindIII restriction enzyme specific target sequence at both 5´ ends of the sequences. After PCR, a 2.1-kb PCR product included the HindIII sequences at the 5' and 3' ends, the complete preCore region and the YMDD main polymerase catalytic motif. After intramolecular ligation of the 5' and 3' ends of the PCR product, a circular molecule was obtained. The molecule contained both regions of interest separated by only 45 nucleotides (the length of the primers and HindIII sequences) (Figure 24a). Once the technique had been well established (Figure 24b), it was possible to analyze the preCore and YMDD motif of the same region by UPDS, despite the 250-nt limitation for clonal sequencing. Presence of HindIII was required for synthesis of the circular construct. Therefore, HindIII served as an internal control to establish the error rate of the UDPS method: any nucleotide difference in this sequence corresponded to a sequencing error, which was ultimately established at 0.03%. In the present study, 8 different samples from 7 patients were analyzed. Four samples were collected at baseline (treatment-naïve) and four after viral breakthrough (VBK) under LVD treatment. Patient 4 was sequentially studied; one baseline and one VBK sample were processed.

Figure 24 a) Technique designed to obtain the preCore region and YMDD motif of the same genome in a circular product to be analyzed by UDPS
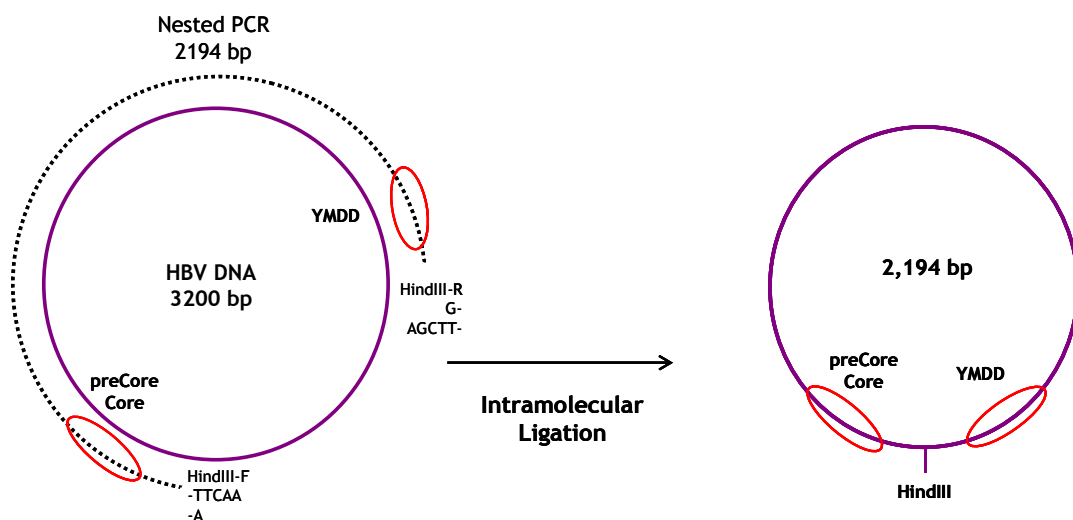
b) Master sequence obtained by conventional sequencing of the circular construct.
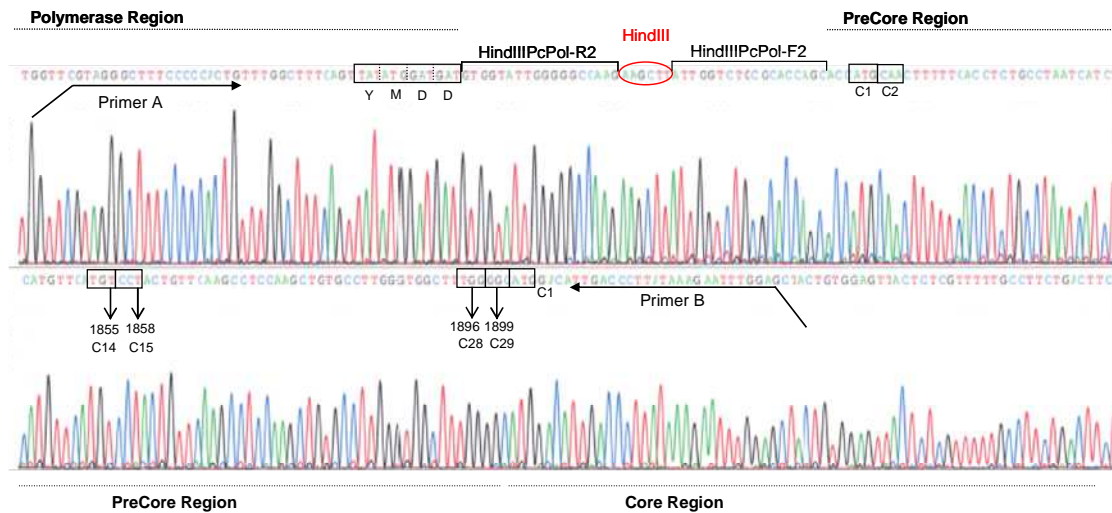
A total of 200 817 sequences were obtained. The preCore region was deeply analyzed, and the main percentages of changes in codons 1, 28, and 29 are presented in Table 7. The HBeAg-positive samples contained significant viral populations of genomes with the main preCore mutation, and HBeAg-negative samples also included viral strains with wild-type preCore sequences in the quasispecies.

Table 7 Percentage of changes detected in the preCore region and Core gene of the 8 samples analyzed.

Sample	Pt ID	HBe Ag	HBV Gen	preCore Region									Core Region	
				Codon 1		Codon 2			Codon 28			Codon 1	
				WT (M)	Other*	WT (Q)	Stop*	Other*	WT (W)	Stop*	Other*	WT (M)	Other
Baseline	1	N	D	99.73	0.27	99.80	0.03	0.17	0.14	99.80	0.06	99.26	0.74
Baseline	2	N	A2	0.13	99.87	99.82	0.02	0.16	99.64	0.25	0.11	99.53	0.47
Baseline	3	N	A2	99.88	0.13	99.94	0.02	0.04	0.02	99.96	0.02	99.89	0.11
Baseline	4	P	D	99.48	0.52	99.71	0.06	0.23	91.59	8.31	0.10	99.49	0.51
VBK	4	P	D	99.51	0.49	99.73	0.11	0.16	98.56	0.21	0.23	99.59	0.41
VBK	5	N	A2	99.54	0.46	99.65	0.14	0.21	0.10	99.85	0.05	99.09	0.91
VBK	6	P	A2	90.19	9.81	99.75	0.07	0.18	97.61	1.31	1.08	99.21	0.79
VBK	7	N	D	99.49	0.51	99.79	0.09	0.12	0.14	99.70	0.16	99.30	0.70

The nucleotide positions constituting the ε structure showed high conservation. However, the percentages of variability were similar between base-paired positions in the basal and apical stems of the structure. Interestingly, the changes in base-paired positions were sequence-

complementary, indicating that the variability in this region might be limited to the base pairing, and therefore, also related to the thermodynamic stability of this structure, as was suggested by conventional Sanger sequencing (12,55). The most frequent changes were in the basal stem, and the most variable were G1896A and G1899A, coinciding with their involvement in HBeAg inhibition. The most conserved position among all the samples analyzed was T1855. This was a striking finding considering that T1855 is base paired with G1899 in the stem of the encapsidation signal. The G1899A change, which was highly selected, stabilized the signal and resulted in a Gly to Asp substitution in codon 29. In contrast, the T1855C change was extremely rare although it conferred the same stabilizing effect (T1855C induced the lowest free energy value of ε for all sequences, 30.1 kcal/mol in genotype A2) and induced a "silent" change in codon 14 (TGT_{1855} and TGC_{1855}, both coding a Cys). All these data suggest a possible essential role of T1855, even more essential than thermodynamic stability. To test the possible involvement of T1855 in viral replication activity, we mimicked the T1855C change by site-directed mutagenesis in Huh7 cells infected with wild-type (T1855) and mutant (C1855) HBV strains. The presence of mutations (C1855) yielded lower HBsAg, HbeAg, and HBV DNA levels than wild-type sequences, suggesting the essentiality of T1855 and justifying its high degree of conservation (it is remarkable that this position is present in both the 5' and 3' ends of pgRNA).

The degree of conservation of several motifs of the encapsidation signal was evaluated to analyze their functionality. For example, the hexanucleotide sequence located at the top of the apical stem (from 1877 to 1882) is described to be highly conserved (54), and the pseudo-triloop (nt positions 1878-1880) may be particularly involved in pgRNA encapsidation. High conservation of the T1878, G1879, and T1880 nucleotides was observed in our study, particularly G1879, which suggests that they have a predominant role in the epsilon signal-capsid interaction.

The 4 nucleotides of the bulge were analyzed because they have a role as a template for synthesis of the 4-nucleotide primer and initiation of viral replication. The 4 nucleotides were highly conserved (98.53% of the sequences) showing to be essential. However, the remaining 1.47% of sequences presented mutations in the bulge and/or acceptor site (AS). These cases showed no homology between the 4NT bulge and the AS in DR1, but different patterns of alternative AS were described, mainly located in nt positions 1849-1852.

The variability of the YMDD region is presented in Table 8. In baseline samples, mutations conferring resistance to LVD (YVDD/YIDD) were detected as minor populations (0.07% to 0.54%) and in samples corresponding to VBK; wild-type YMDD viral populations were also detected (0.07% to 1.06%). Linkage analysis of preCore and YMDD mutations showed that simultaneous presence of mutations in both the preCore and polymerase regions of the same viral genome was allowed, as occurred in patient 5 (99.46% of M204I and 99.85% of G1896A) and Patient 7 (99.69% M204V and 99.70% of G1896A).

Two samples (baseline and LVD non-response) from Patient 4 were analyzed. Interestingly, rt204I was the main YMDD mutation detected in the initial sample (0.47%); however, rt204I was not selected in the viral breakthrough sample, whereas rt204V was selected. This result suggests that the baseline polymerase resistant variants present in low levels (>0.5%) do not determine the selection at VBK. In addition, in this single longitudinal study, the percentage of HBeAg-negative variants significantly decreased after VBK (8.37% to 0.31%), indicating that HBeAg-negative strains might be more sensitive to LVD treatment than HBeAg-positive variants. This concept is in accordance with the better response to antiviral treatment in HBeAg-negative than HBeAg-positive cases (89).

Table 8 Mutations in the YMDD motif in the 8 samples analyzed.

Sample	Pt ID	HBe Ag	HBV Geno	Amino acid at codon 204 of the polymerase (YMDD motif)			
				M	V	I	Other
Baseline	1	N	D	99.70	0.11	0.09	0.09
Baseline	2	N	A2	99.63	0.08	0.20	0.10
Baseline	3	N	A2	99.89	0.02	0.05	0.04
Baseline	4	P	D	99.33	0.08	0.47	0.12
VBK	4	P	D	0.25	93.90	5.63	0.23
VBK	5	N	A2	0.08	0.20	99.46	0.27
VBK	6	P	A2	1.06	98.69	0.01	0.24
VBK	7	N	D	0.07	99.69	0.00	0.24

The circular construct included the last 8 codons of the X gene and the first codon of the Core gene. A minor proportion of all the sequences (0.16%) contained a mutation that destroyed the X gene TAA stop codon. These mutants led to elongation of the X antigen by 51 amino acids, and a consequent alteration of the COOH terminal region of this antigen (39). Moreover,

significant percentages (0.11% to 0.91%) of mutations in the first codon of the Core gene were observed in the viral populations of all samples. This change might produce defective Core particles and their presence in the viral quasispecies could be explained by transcomplementation mechanisms, such as encapsidation of defective genomes with core proteins encoded by competent genomes (without these mutations).

In conclusion, this UDPS study confirms the feasibility of detecting mutations simultaneously occurring in the preCore region and YMDD polymerase motif. The presence of LVD-resistant variants in baseline naïve samples, preCore mutations in HBeAg-positive patients, and wild-type preCore variants in HBeAg-negative cases was also demonstrated. These findings show the complexity of HBV viral quasispecies, which contains reservoirs of the main clinically significant variants in HBV infection that can be selected when the corresponding evolutive factors are present (eg, in YMDD variants, administration or elimination of antiviral therapies and in preCore variants, the host immune response). The low percentages of defective Core genomes and mutated HBx strains whose presence might be allowed due to cooperation between viruses also illustrates the complexity of the HBV quasispecies.

The thermodynamic stability of the ε signal was confirmed to be the principal restrictive factor for selection of the main preCore mutation, which is responsible for abolishment of HBeAg expression. Furthermore, analysis of ε signal variability revealed the essential role of ε structural motifs and possible involvement of some nucleotides in ε signal function. Lastly, correct annealing between the 4NT primer, synthesized in the base bulge sequence, and the AS in the DR1 sequence does not seem to be an absolute condition for the 4NT primer, required for minus DNA strand synthesis. The variability observed in the bulge region has enabled definition of a putative cis-acting sequence as an alternative to the normal acceptor site in the 3' of pgRNA.

3.3   Complete manuscript

Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome

Maria Homs[1,2], Maria Buti[1,3], Josep Quer[1], Rosendo Jardi[1,2], Melanie Schaper[1,2], David Tabernero[1,2], Israel Ortega[4], Alex Sanchez[4], Rafael Esteban[1,3] and Franciso Rodriguez-Frias[1,2,*]

[1]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto Carlos III Corsega 180, 08036, Barcelona, [2]Department of Biochemistry, [3]Department of Hepatology, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona Passeig Vall d'Hebron 119–129, 08035, Barcelona and [4]Statistics and Bioinformatics Unit, Research Institut, Hospital Vall d'Hebron Passeig Vall d'Hebron 119–129, 08035, Barcelona, Spain

ABSTRACT

Hepatitis B virus (HBV) pregenomic RNA contains a hairpin structure (ε) located in the preCore region, essential for viral replication. ε stability is enhanced by the presence of preCore variants and ε is recognized by the HBV polymerase (Pol). Mutations in the retrotranscriptase domain (YMDD) of Pol are associated with treatment resistance. The aim of this study was to analyze the preCore region and YMDD motif by ultra-deep pyrosequencing (UDPS). To evaluate the UDPS error rate, an internal control sequence was inserted in the amplicon. A newly developed technique enabled simultaneous analysis of the preCore region and Pol in the same viral genome, as well as the conserved sequence of the internal control. Nucleotide errors in HindIII yielded a UDPS error rate <0.05%. UDPS study confirmed the possibility of simultaneous detection of preCore and YMDD mutations, and demonstrated the complexity of the HBV quasispecies and cooperation between viruses. Thermodynamic stability of the ε signal was found to be the main constraint for selecting main preCore mutations. Analysis of ε-signal variability suggested the essential nature of the ε structural motif and that certain nucleotides may be involved in ε signal functions.

INTRODUCTION

Hepatitis B virus (HBV) is a double-stranded DNA virus, whose genome (3.2 kb) contains four Open Reading Frames (OPFs) encoding seven proteins: surface antigens (preS1, preS2 and HBsAg), polymerase (Pol), core protein (HBcAg), X protein (HBx) and the soluble 'e' antigen (HBeAg). During the viral cycle, HBV-DNA conformation shows different phases: relaxed-circular, partially double-stranded DNA in circulating viruses and covalently closed circular DNA (cccDNA) within the hepatocyte nucleus. cccDNA acts as a mini-chromosome within hepatocytes and transcribes five messenger RNA: preS1, preS2, X, preCore and pregenomic RNA (pgRNA). pgRNA is a greater than full-length HBV genome (3.5 kb) that codes for the Pol and HBcAg proteins. Pol has two catalytic activities, reverse transcriptase (RT) and RNaseH, and contains a spacer sequence and a terminal protein (1).

In both the 5'- and 3'-ends of pgRNA, there is a redundant sequence of 60 nt that adopts a stem loop structure (Figure 1a and b) and acts as encapsidation signal (ε). The 5' sequence is crucial for Pol protein recognition and binding to the 5' ε structure (2). Pol triggers pgRNA encapsidation by core proteins, and initiates reverse transcription (3). Pol employs an N terminal tyrosine residue of the terminal protein domain as a primer to generate the first 4 nt (4 nt primer) of the DNA negative strand. This 4 nt primer is covalently attached to Pol and uses the bulge of the ε signal as template (Figure 1a) (4). The 4 nt primer and Pol protein complex switches from the 5'- to 3'-end of pgRNA and anneals to a 4-nt acceptor site within the 3' copy of direct repeat region 1 (Figure 1b) (5).

pgRNA and preCore mRNA share the same preCore region, but code for different antigens: HBcAg (starting at Position 1901 of the HBV genome) and HBeAg (starting at Position 1814), respectively. Several preCore region mutations abolish HBeAg expression. The most common is G
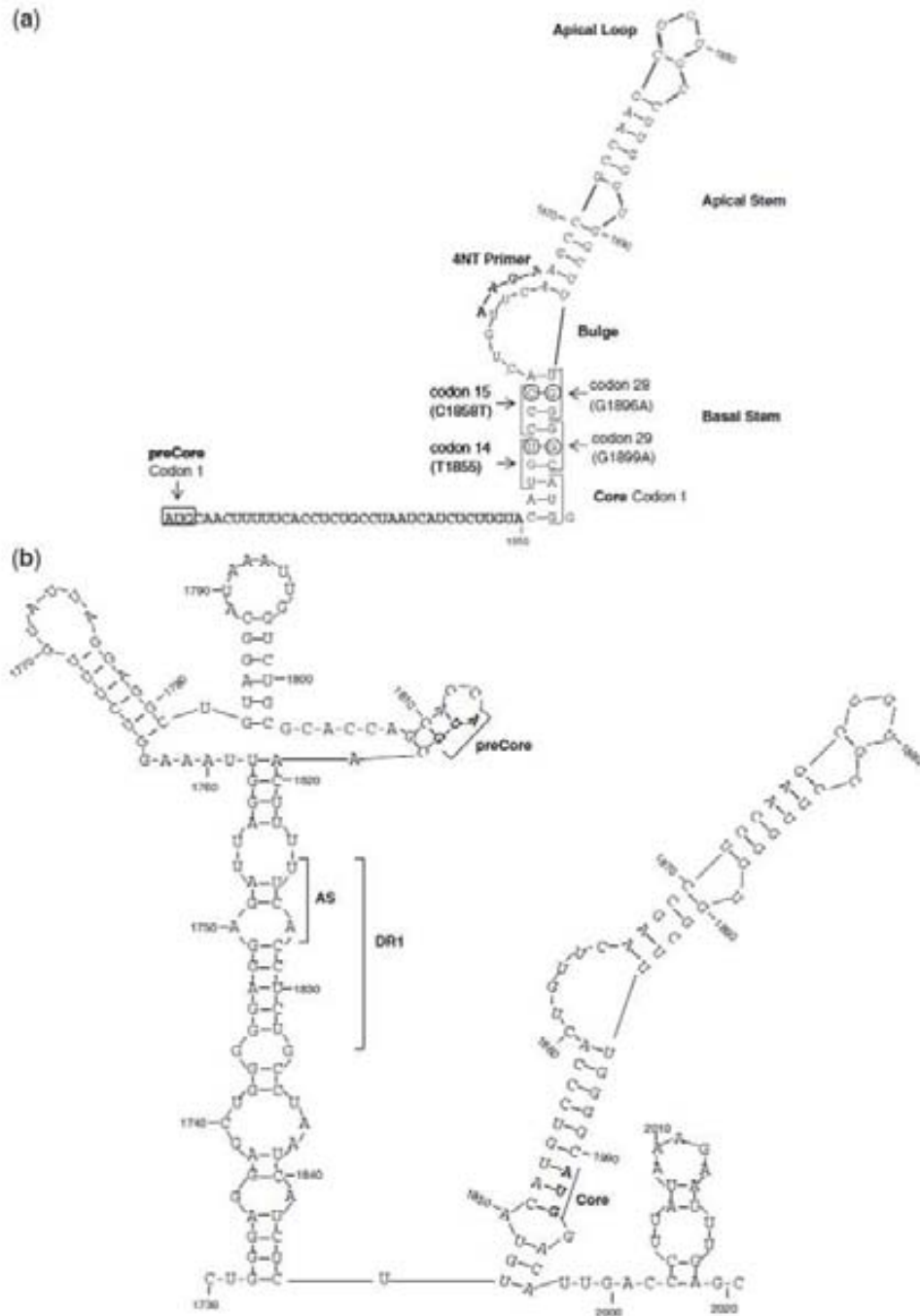
Figure 1. (a) Stem-loop structure of the 5′ HBV pgRNA. Main structural motifs (loops and stems), main codons studied (1, 14, 15, 28 and 29 of preCore and 1 of Core genes), and the location of 4-nt primer annealing in the 5′ pgRNA are indicated in the figure. (b) Stem-loop structure of the 3′ HBV pgRNA. The acceptor site (AS), direct repeat region 1 (DR1), and the preCore and Core start codons are indicated in the figure.

to A at Position 1896 (G1896A), which changes a tryptophan (W) in codon 28 ($TG_{1896}G$) to an amber stop codon ($TA_{1896}G$): pcW28stop. This substitution is often associated with the G1899A change in codon 29 (pcG29D) (6). Other minor mutations have been observed in codon 2 (CAA to TAA, Q2stop) and in codon 1 (7). The pcW28stop change is reported to be responsible for 95% of HBeAg-negative patients in our area, mainly in association with viral genotype D (8,9). The G1896A and G1899A mutations have been associated with secondary ε structure stability and therefore, with viral replication (8,10–12).

The main mutations conferring resistance to nucleoside analogues, such as lamivudine (LMV), are predominantly detected in the RT catalytic motif, $YM_{204}DD$, in which methionine (M) is changed to valine (V) or isoleucine (I) (rtM204V and rtM204I) (13,14). Prolonged LMV therapy is associated with a high probability (>70% after 5 years) of resistance to treatment (15). After the first year of therapy, anti-HBe-positive patients have a significantly greater response than HBeAg-positive cases; however, long-term response rates are compromised by the reappearance of HBV DNA in serum (16). Several factors can have an effect on the selection rates of LMV-resistant mutants: ethnic background, viral load, HBV genotype, core promoter variants and preCore variants (17).

Ultra-deep pyrosequencing (UDPS) by GS-FLX platform (Roche) has been used to detect clinically relevant minority drug-resistant variants in treatment-naïve and experienced HIV-1 (18,19) and HBV patients, thereby confirming its ability to identify individual, well-known mutations present in the Pol and HBsAg genes at frequencies of 0.1–1% of the HBV quasispecies (20,21). UDPS by GS-FLX platform is currently the only technology that allows simultaneous quantitative analysis of thousands of clonally amplified sequences of 200–250 nt in length for standard GS-FLX (longer fragments since last year).

The aim of the present study was to analyze by UDPS the HBV preCore region (thermodynamic structural restriction and variability of essential structural motifs of the ε signal) and its association with the RT catalytic YMDD motif variants, with analysis of both regions in the same sequence. However, these two regions are separated by a distance of 1 kb in the HBV genome,

making it difficult to simultaneously study both regions by classic cloning methodology (22,23). Therefore, our methodological hypothesis was based on executing a PCR that would encompass the preCore and YMDD regions, using primers that include the HindIII site on their 5'-ends. After induction of intramolecular ligation, a circular molecule was produced, which included the YMDD Pol motif and the preCore region flanking a HindIII sequence. The integrity of the restriction site sequence was used as the internal control to verify the fidelity of UDPS analysis.

PATIENTS AND METHODS

Patients

Seven chronically infected HBV patients, all males and with a mean age of 36 years (range 17–50), were included in the study. Patients were selected because they had complete clinical documentation and known viral characteristics by conventional techniques. They had been treated with LMV for a mean of 25 months because of high HBV DNA levels and elevated hepatic biochemical markers. Two patients were HBeAg-positive at baseline and e antigen loss did not occur during antiviral treatment. Genotype changes were not observed during the study period. Baseline biochemical and serological markers, and treatment characteristics are described in Table 1. All patients tested negative for HCV and HIV markers. HBV serology, HBV DNA and LMV-resistant mutations were tested every 3–6 months. The circular construct used to analyze the preCore and YMDD was obtained from four baseline samples (Patients 1–4) and from four samples corresponding to a time of viral breakthrough with mutations in the YMDD motif (Patients 4–7).

Serology and HBV DNA assays

Serological markers for HBV (HBsAg, HBeAg and anti-HBe), HCV (anti-HCV), hepatitis delta (anti-HDV) and HIV (anti-HIV) were tested by commercial enzyme immunoassays. HBV-DNA was quantified using real-time PCR with a detection limit of 20 IU/ml (COBAS TaqMan HBV V2.0, Roche Diagnostics GmbH, Mannheim, Germany). Main preCore mutations

Table 1. Baseline biochemical and serological markers, and treatment characteristics of the seven patients included in the study

ID	Baseline				Lamivudine treatment					
	HBV DNA (IU/ml)	HBeAg	ALT (IU/ml)	HBV genotype	Type response	Months	HBV DNA (IU/ml)	ALT (IU/ml)	Type mutation	Rescue treatment
1	>10^8	N	392	D	VBK	48	3.6×10^5	31	M204V	LAM+ADV
2	6.6×10^6	N	167	A2	SVR	24	<20	30	–	–
3	1.2×10^6	N	54	A2	SVR	24	<20	33	–	–
4	>10^8	P	52	D	VBK	12	6.2×10^6	281	M204IV	LAM+TDF
5	3.3×10^6	N	78	A2	VBK	34	8.3×10^6	401	M204MIV	LAM+ETV
6	>10^8	P	214	A2	VBK	24	1.3×10^7	93	M204V	LAM+TDF
7	2×10^6	N	117	D	VBK	12	2.2×10^6	103	L180LM	LAM+ADV

SVR, sustained virologic response.

were detected by Sanger sequencing, as previously described (8). HBV genotype and mutations conferring treatment resistance were detected with LiPA (INNO-LiPA, Innogenetics, Ghent, Belgium) (sensitivity 1–5%) and direct sequencing (sensitivity 20%) (24).

Generation of a circular construct containing the YMDD motif and preCore region

HBV DNA was extracted from 200 μl of serum with the QIAamp DNA MiniKit (QIAGEN, Hilden, Germany). With the aim of analyzing greatly separated regions in the same genome, a 2.7-kb HBV DNA fragment including the YMDD Pol motif and preCore region was PCR amplified (Figure 2a), using the forward primer F1 (Position 1703) and reverse primer R1 (Position 1018) (Table 2). The PCR reaction mix included 200 μM of dNTPs and 400 nM of each primer, and the method was performed as follows: 30 cycles at 94°C for 30 s, 55°C for 30 s and 72°C for 3 min, followed by 10 cycles at 94°C for 30 s, 58.6°C for 30 s and 72°C for 3 min and a final hold at 72°C for 7 min. The amplicon obtained was amplified by nested PCR with primers that included the HindIII restriction site sequence (*AAGCTT*) at the 5'-end (forward HindPcPol-F2, Position 1792, and reverse HindPcPol-R2, Position 764) (Table 2) under the same conditions as the first PCR run, yielding a 2194-kb product in genotype A2 samples and a 2155-kb product in genotype D samples. To minimize amplification errors, Pfu UltraII (Stratagene, Agilent Technologies, La Jolla, USA) polymerase was used in both PCR runs. The nested amplicon was purified through 0.9% agarose gel, digested for 1 h at 37°C with HindIII (NEB, Ipswich, USA), and purified again through 0.9% agarose gel. The next step involved ligation of both HindIII ends of the same fragment (intramolecular ligation) so that the two regions of interest would come close enough in the same molecule to be analyzed by UDPS (Figure 2a). Optimal results were obtained with 0.5 ng/μl of the HBV DNA-PCR fragment, and 1 U of T4 DNA ligase (Roche Diagnostics GmbH). The ligation product was loaded onto 0.9% agarose gel; the band pattern obtained is shown in Figure 2b. The band corresponding to the HindIII circular construct (apparent size ~1700 bp) (band a in Figure 2b) was cut from the gel and purified. The identity of the circular constructs was confirmed by sequencing two overlapping PCR amplicons that cover the complete circular construct sequence: a 1.2-kb amplicon obtained with primers A3 and B3 and sequenced with C3, D3 and E3 and a 1.8-kb amplicon obtained with primers A4 and B4 and sequenced with A4, C4, D4 and C3 (primers described in Table 2). The fragment obtained from sequencing the 1.3-kb amplicon, which includes the preCore region, the YMDD motif and the HindIII sequences with C3 is shown in Figure 2c.

Simultaneous UDPS analysis of the YMDD motif and preCore region

A 207-bp nucleotide fragment was PCR amplified from the DNA recovered from band a (Figure 2b). All HBV-specific primers included 5' adaptors as binding sites for the pyrosequencing reaction (left 5' GCCTCCC TCGCGCCA, right 5' GCCTTGCCAGCCCGC; both ending with the TCAG barcode). The complete sequence of UDPS-Primer A as forward and UDPS-Primer B as reverse is shown in Table 2. Briefly, PCR was performed using 1× Pfu UltraII polymerase buffer, 10 mM of each dNTP, 20 pmol of forward and reverse primers and 2.5 U of Pfu UltraII DNA polymerase (Stratagene, Agilent Technologies) in a final volume of 100 μl. After a single denaturation step of 2 min at 95°C, samples underwent 30 cycles for 30 s at 95°C, 30 s at 60°C and 2 min at 72°C, and a single final 10-min step at 72°C. Before UDPS, the quality and length of the HBV DNA amplicons was verified with the Agilent 2100 bioanalyzer (Agilent Life Science, Santa Clara, CA, USA). The 207-bp fragment, which included 163 bp of the HBV-DNA sequence, underwent UDPS with the 454 Life Science platform (GS-FLX, Roche Applied Science), according to the manufacturer's protocol. Base calls and quality scores were determined using the GS-Amplicon Variant Analysis software (included in the GS-FLX platform).

Algorithm for interpreting sequence data

Data accuracy was validated according to previously reported procedures (18,25,26). Briefly, reads with insertion-deletions, reads in which the primer did not match the reference primer by at least 75%, those that did not reach a fixed length and those containing ambiguous calls were discarded. A complementary reverse sequence was obtained from the reverse sequence. The program selected nucleotide sequences that passed the filter, and the sequences were translated to amino acids for the polymerase and preCore regions (both regions are in-frame in the circular construct). Several files were obtained and a matrix of the frequency of changes at each amino acid position in comparison with the master from each patient was used. As was mentioned above, the circular construct used as the target for UDPS analysis was the result of intramolecular ligation of the HindIII 5'- and 3'-ends. Therefore, the integrity of this 6-bp restriction enzyme sequence (AGGCCT) was a requirement for correct ligation, and any nucleotide substitution in this sequence would be a UDPS error. For this reason, instead of using a more complex Poisson filter (19,25,26), which needs UDPS processing of additional clonal sequences (external control), the changes detected in this 6-bp sequence were used to establish the cut-off fidelity of the UDPS process (internal control).

Analysis of RNA folding stability in the HBV pgRNA encapsidation signal

Computer models of RNA folding and structure representations were obtained from the RNA Structure Fold program (version 5.3), developed by Dr David H. Mathews (27,28). All calculations were performed at a temperature of 37°C. In keeping with Dr Mathews' kindly provided suggestion, a script adaptation of the RNA Structure Fold program (by command line
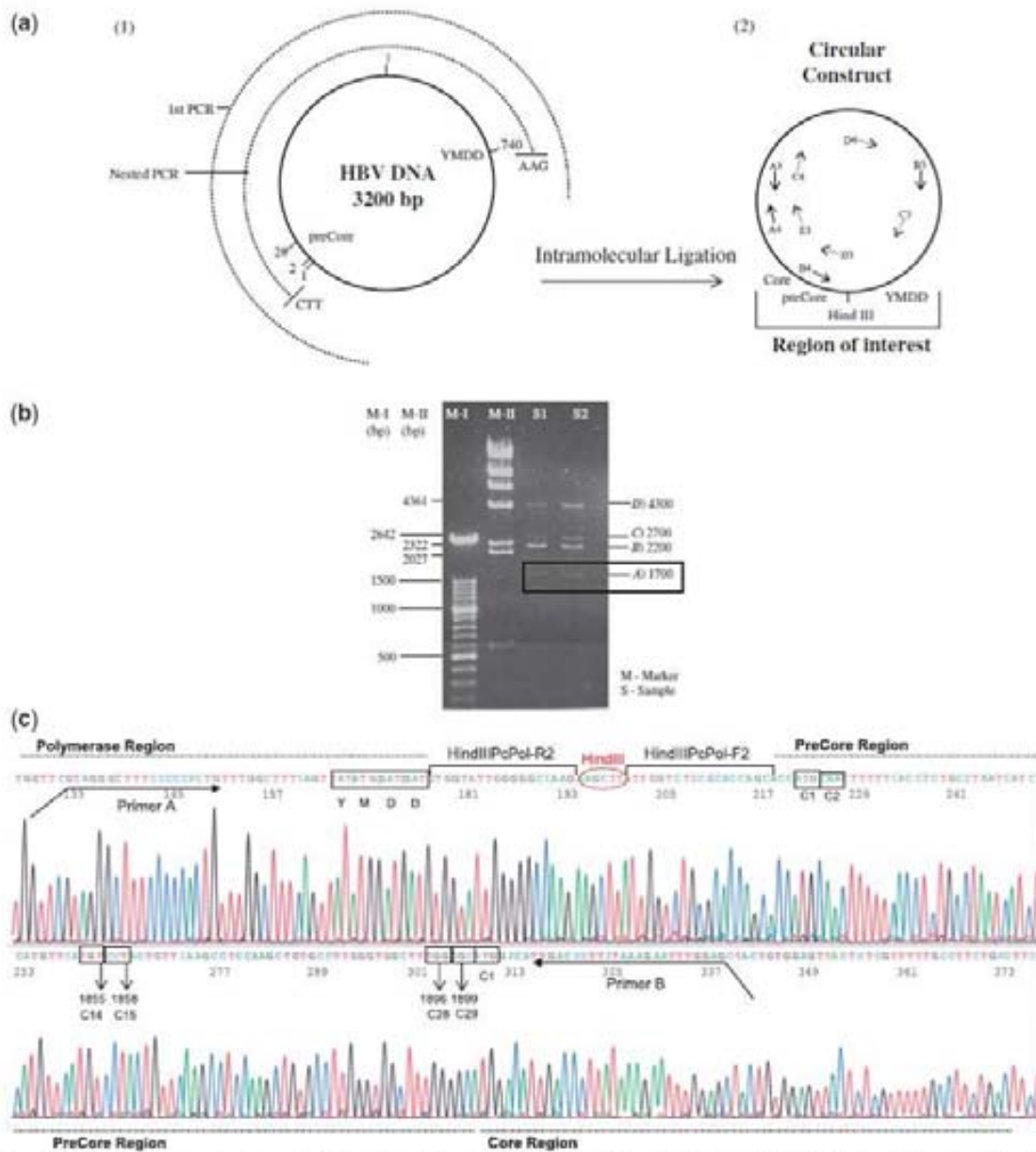
Figure 2. (a) Schematic representation of the intramolecular ligation technique to obtain the preCore and YMDD Pol motif close to each other. (1) The circle in a continuous line represents the complete HBV genome (3200 bp). The semicircles in discontinuous lines represent the first (first PCR, 2700 bp) and the nested (2194 bp for genotype A2 sequences and 2155 bp for genotype D sequences) PCR products. (2) The product obtained after intramolecular ligation and primers used to sequence the circular molecule (Table 2) are indicated in the figure. (b) Agarose gel obtained after intramolecular digestion. Two markers (M-I and M-II) were loaded in the gel; band lengths are shown at the left side of the image. Two samples (S1 and S2) were loaded and four bands were detected; band lengths are indicated at the right side of the image. *A* band with an apparent size of 1700 bp was the circular molecule, with the preCore and YMDD Pol motif close to each other. Band *B* was the nested PCR product without ligation, and band *D* were dimers of nested PCR products without ligation. Band *C* corresponded to the ligated dimers. *B*, *C* and *D* bands were not used, because the main interest was to analyze regions from the same genome. (c) Sequence of the circular construct including the preCore region, YMDD polymerase motif (in box), and HindIII sequence. PreCore (1, 2, 14, 15, 28 and 29) and Core (1) codons analyzed in the text are also enclosed in boxes.

Table 2. Sequences of primers used in the study

Name	Position of the primer and direction in HBV genome (bold characters)	Sequence (5'–3')
F1	1703, Fw	YATAAGAGGACTCTTGGACT
R1	1018, Rv	GCAAANCCCAAAAGACCCA
HindPcPol-F2	1792, Fw	TAC*AAGCTT*ATTGGTCTGCGCACCAGC
HindPcPol-R2	764, Rv	TAC*AAGCTT*CTTGGCCCCCAATACCAC
A3	2489, Rv	CCCMGTAAARTTYCCCACCTTAT
B3	245, Fw	GTCTAGACTCGTGGTGGACTTCTCTC
C3	550, Fw	ATGTWTCCCTCWTGTTGCT
D3	1864, Fw	TCAAGCCTCCAAGCTGTGCC
E3	2267, Fw	GGAGTGTGGATTCGCACT
A4	2270, Fw	GTGTGGATTCGCACTCCTCCAG
B4	1927, Fw	AAATTCTTTATAAGGGTCAATGT
C4	2817, Fw	TCACCATATTCTTGGGAACA
D4	57, Fw	TGCTGGTGGCTCCAGTTC
UDPS-Primer A	703, Fw	GCCTCCCTCGCGCCATCAG*GTAGGGCTTTCCCCCACTG*
UDPS-Primer B	1931, Rv	GCCTTGCCAGCCCGCTCAG*CTCCAAATTCTTTATAWGGGTCRA*
Fw T1855C	1834, Fw	CTAATCATCTCTTGTACATGcCCCRCTGTTCAAGCCTCCAAGCTG
Rv T1855C	1879, Rv	CAGCTTGGAGGCTTGAACAGYGGGgCATGTACAAGAGATGATTAG
C3-Rv	569, Rv	AGCAACAWGAGGGAWACAT
pTriEx-Fw	–	TCTGCGCTCTGCTGAAGCCAGTTA
pTriEx-Rv	–	ACGATCGGAGGACCGAAGGAGCT

The position and direction of the 5' primer in the HBV genome (bold characters in the sequence), and the mutagenized position (bold lower cases) are indicated

interfaces) was used to calculate the lowest free energy of the ε structure. Briefly, each nucleotide sequence was automatically converted to a 'seq' file, with the input format required for the program. A 'bat' file was created for each 'seq' file, enabling execution of the Fold module of the program for each sequence and creating the output 'ct' files, which contained the free energy of the structure. The circular construct analyzed included nucleotides from Positions 1814 to 1912, corresponding to the main sequence of the 5' ε signal (8,11,12). The sequences obtained were used to calculate the epsilon energy of the RNA ensemble. Because the epsilon energy data are not normally distributed, between-group comparisons were done using a permutation test based on the standard *t* score for two groups. The Monte Carlo approach to compute *P*-values was carried out, with $n = 2000$ permutated samples. *P*-values in these comparisons were corrected for multiple testing problems following the false discovery rate procedure. The standard RNA-Structure Fold (5.3) program was used to predict the secondary structures of the 5' (Figure 1a) and 3' (Figure 1b) (Position 1730–1930) ε signals. Both predicted structures showed a pseudo-triloop at the edge of the apical stem, as has been demonstrated by magnetic resonance (MR) analysis (29).

Phenotyping, mutagenesis and cell culture

A more than full-length HBV (1.1 genome unit) (able to produce functional replicative pgRNA) from a patient chronically infected with HBV genotype A was cloned in pTriEx-mod vector (kindly provided by Dr F. Zoulim), as described by Durantel *et al.* (30). This construct was used to simulate HBV variants by site-directed mutagenesis. In the present study, a thymine to cytosine mutation at Position 1855 (T1855C) was generated, using the

Multi-Site-Directed Mutagenesis Kit (Agilent Technologies), following the manufacturer's instructions. This is a silent change that does not induce an amino acid substitution (both code for a cysteine). The forward primer was Fw-T1855C, and the reverse primer Rv-T1855C (Table 2). Because of the 60-nt redundancy in HBV pgRNA, Position 1855 is repeated at the 5'- and 3'-ends. Thus, the mutagenesis process yielded three different constructs that contained the T1855C change at the 5'-end, the 3'-end or simultaneously at both ends. Generation of these mutations was confirmed by direct sequencing of the clones with primers annealing pTriEx-mod-1.1 HBV just before the HBV ligation (promoter-Fw for the 5'-end, and C1D1-Rv for the 3'-end, shown in Table 2).

The three pTriEx-mod-1.1 HBV constructs were transfected in Huh7 human hepatoma cells cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% calf serum. Plasmid transfections were performed using Fugene-HD (Roche Diagnostics GmbH). Cells were seeded in 12-well plates at a density of 10^6 cells/ml and were transfected at 80–90% confluence, following the manufacturer's instructions. The supernatant was used to quantify production of HBsAg (Architecht, Abbot Diagnostics, Sigo, Ireland), HBeAg (Vitros, Ortho-Clinical Diagnostics, High Wycombe, UK) and HBV DNA (COBAS TaqMan HBV V2.0, Roche Diagnostics GmbH). HBV DNA was extracted from the supernatant (QiagenAMP DNA Mini Kit, Qiagen, Hilden, Germany) according to the manufacturer's instructions and used to evaluate HBV DNA production. The HBV viral particle origin of HBV-DNA detected in culture supernatants was confirmed by differential PCR. Briefly, consecutive dilutions (1/1, 1/10, $1/10^2$, $1/10^3$ and $1/10^4$) of extracted DNA were performed, and specific

regions of HBV and pTriEx-mod were amplified, using the paired primers C4 and C3-Rv for HBV, and pTriEx-Fw and pTriEx-Rv for a 592-bp region of the plasmid (Table 2). Two negative controls were performed in the study, one that contained only Fugene, which enabled evaluation of Huh7 replication in the presence of the transfection reagent, and another that contained only HBV DNA without reagent. HBsAg and HBeAg tested negative, but quantification of HBV DNA was positive (2.15×10^2 UI/ml). Differential PCR analysis, using primers for both HBV and pTriEx amplification, were positive only at a dilution of 1/10. HBV PCR of wild-type and mutant clones remained positive up to $1/10^3$ to $1/10^4$ dilutions, indicating production of HBV DNA resulting from HBV transfection of Huh7 cells. Amplification of the pTriEx region in HBV transfection reactions was only detectable at a dilution up to 1/10; in further dilutions, PCR was negative.

RESULTS

HindIII restriction sequence as internal control of UDPS

In the 200 817 (Table 3) sequences analyzed, only 189 nt changes were observed at the HindIII restriction enzyme site. Sixteen HindIII mutated sequences were found at a frequency ranging from 1 to 24 (average, 22) and all carried only one amino acid change. The single nucleotide error rate was 1.6×10^{-4} and the codon error rate was 4.7×10^{-4}. These values were considered the UDPS error rates, and the cut-off for the sequence analysis was set at 0.03% (twice the nucleotide error rate).

Analysis of the main preCore codons

PreCore codons 1, 2 and 28 were analyzed; changes in these codons are the main cause of HBeAg non-expression. The percentage of changes in each sample processed is shown in Table 3. In four of the five HBeAg-negative patients (Patients 1, 3, 5 and 7), HBeAg-negative biochemical status was due to a stop mutation in codon 28 (W28stop, main preCore mutation). In Patient 2, HBeAg-negative status was due to mutations in codon 1 of the preCore region (99.87%). This mutation was also detected at a lower frequency in Patient 6 (9.81%). Patients 2 and 6 were both infected with genotype A2.

The circular construct included the ε structure sequence of the 5′ pgRNA end. The base pairing rate between Position 1896 in codon 15 and Position 1858 in codon 28 was analyzed. Among a total of 90 386 sequences with $TA_{1896}G$ in codon 28, 89 837 (99.39%) showed CCT_{1858} in codon 15 and among all sequences with CCT_{1858} in codon 15, 99.76% showed $TA_{1896}G$. These results indicate that correct base pairing between Positions 1896 and 1858 is an extremely important constraint in the G1896A mutation. The second most frequent nucleotide substitution, G1899A (28.35%), was paired with T_{1855} in 99.3% of cases, and was observed in the presence of G1896A in 99.6% of cases.

Table 3. Percentage of changes in the polymerase, precore and core positions of each sample processed

Sample	Pt ID	HBeAg	HBV Genotype	Total sequences	Amino acid at codon 204 of polymerase (YMDD motif)				preCore Region												Core Region			Free energy (kcal/mol) in consensus sequence
									Codon 1		Codon 2				Codon 28				Codon 1			
					M	Vᵃ	Iᵃ	Other	WT (M)	Otherᵇ	WT (Q)	Stopᵇ	Otherᵇ	WT (W)	Stopᵇ	Otherᵇ	WT (M)	Other	
Baseline 1		N	D	19 421	99.70	0.11	0.09	0.09	99.73	0.27	99.80	0.03	0.17	0.14	99.80	0.06	99.26	0.74	−26.1
Baseline 2		N	A2	29 024	99.63	0.08	0.20	0.10	0.13	99.87	99.82	0.02	0.16	99.64	0.25	0.11	99.53	0.47	−27.6
Baseline 3		N	A2	15 585	99.89	0.02	0.05	0.04	99.88	0.13	99.94	0.02	0.04	0.02	99.96	0.02	99.89	0.11	−25.9
Baseline 4		P	D	34 182	99.33	0.08	0.47	0.12	99.48	0.52	99.71	0.06	0.23	91.59	8.31	0.10	99.49	0.51	−25.2
VBK 4		P	D	25 175	0.25	93.90	5.63	0.23	99.51	0.49	99.73	0.11	0.16	98.56	0.21	0.23	99.59	0.41	−25.2
VBK 5ᶜ		N	A2	20 683	0.08	0.20	99.46	0.27	99.54	0.46	99.65	0.14	0.21	0.10	99.85	0.05	99.09	0.91	−25.5
VBK 6ᶜ		P	A2	25 175	1.06	98.69	0.01	0.24	90.19	9.81	99.75	0.07	0.18	97.61	1.31	1.08	99.21	0.79	−27.1
VBK 7ᶜ		N	D	31 572	0.07	99.69	0.00	0.24	99.49	0.51	99.79	0.09	0.12	0.14	99.70	0.16	99.30	0.70	−26.1

Bold values indicate percentages higher than 1%.
ᵃLMV-resistant variants. Mutations other than V and I were grouped as 'Others'.
ᵇHBeAg-abolishing variants.
ᶜSamples in which polymerase and precore-mutated variants are predominant.

Conserved positions in the preCore region and maintenance of base pairing in the basal and apical stems

A schematic representation of base-pairing between the preCore positions analyzed, corresponding to ϵ, is shown in Figure 3, which includes the percentage of nucleotide and amino acid changes for all sequences. Although a high level of conservation was observed, there was a low percentage of variability ($\geq 0.05\%$, close to the UDPS error rate). Interestingly, the base-paired positions in the basal (1851–1859 and 1895–1903) and apical (1866–1877 and 1881–1894) stems showed similar percentages of changes to maintain base-pairing.

Attending to the upper stem positions, a T to C change in 1884, 1885 and 1893 was observed in >0.1% of sequences (0.15, 0.14 and 0.13%, respectively), with a complementary change of A to G in their base-paired Positions (1875, 1874 and 1867) in very similar percentages (0.13, 0.1 and 0.13%, respectively). Additionally, the low percentage (0.1%) of G to A changes in Position 1891, which represent a new stop codon in the preCore sequence (W26stop), were also compensated by a complementary C to T change in a similar percentage (0.09%) in the base-paired C1869T position. In all the apical stem positions, base-pairing of the structure was maintained.

The hexanucleotide edge of the upper stem (Positions 1877–1882) was also analyzed, with emphasis on the TGT nucleotides (1878–1880), which form a pseudo-triloop structure essential for pgRNA packaging (31). The pseudo-triloop was found to be highly (but not completely) conserved (99.5% of cases). Four variants were detected at a frequency of >50 relative to the master: 403 CGT, 330 TGC, 118 TAT and 66 AGT. Among the three nucleotides, the most highly conserved was G_{1879} (0.06% variability), and this was one of the most highly conserved nucleotides in the complete ϵ signal. Outside the pseudo-triloop, C_{1882} showed 0.15% variability, and the single nucleotide bulge T_{1889}, located in the middle part of the apical stem, had 0.2% variability (Figure 3).

Interestingly, the most frequent changes in the ϵ structure were located in the basal stem (G1896A and G1899A), where the rule of base-pairing maintenance was also observed (see 'Analysis of main preCore codons'). The low percentage of T to C changes (0.3%) in Position 1855 was surprising considering that this change did not induce an amino acid substitution. Its complementary base, Position 1899, showed a higher rate of G to A mutations (28.3%), despite that fact that a significant amino acid change (G29D) was produced. Study of Position T1855 to C by mutagenesis is described in the next section.

Stability of the pgRNA 5′ ε signal: free energy of the stem loop structure

The free energy of the thermodynamic ensemble of the stem loop structure of the pgRNA 5′ ε signal was calculated for all sequences obtained by UDPS analysis (energy values for master sequences shown in Table 3). The distribution obtained, attending to HBV genotype and the presence (mutated) or absence (wild-type) of mutations in preCore codons 1 (pc1) and 28 (pc28) is shown

in Table 4. An analysis was performed between values from the same genotype group; however, comparisons including pcC1 were not taken into account because the first preCore codon (pc1) is not present in the pgRNA 5′-end (it starts at Position 1818). The presence of a mutation in pc28 destabilized the ϵ structure in genotype A2 sequences (higher energy in codon 28 mutations), whereas in genotype D, the opposite was observed, indicating that in this case, pc28 mutations stabilize the ϵ structure (lower energy in pc28 mutations).

The low percentage of T to C changes in Position 1855 mentioned above was even more surprising when stabilization energy was calculated. The rare T1855C (0.30%) variant showed the lowest energy value (−30.1 kcal/mol in genotype A2) and therefore, the highest theoretical thermodynamic stability. This observation, in addition to the finding that T1855C did not induce an amino acid change, prompted us to mimic the T1855C mutation by site-directed mutagenesis. Two clones that are found in natural HBV replication were used to transfect Huh7 cells: the WT clone and another with the 1855 C mutation in both the 3′- and 5′-ends. Two more clones with the mutation in only one of the two pgRNA ends were selected to analyze possible differences in the nucleotide depending on its location in pgRNA. Each transfection was performed at least six times and detailed results of the average expression of HBsAg, HBeAg and HBV DNA are shown in Table 5. In the presence of the double 5′ plus 3′ mutation, HBsAg, HBeAg and HBV-DNA had systematically lower values than T1855 WT, indicating a possible essential nature of the nucleotide. Surprisingly, quantification of HBsAg, HBeAg and HBV DNA yielded higher levels in clones carrying only one mutation in either the 5′- or 3′-end.

The 4-nt bulge and the acceptor site

The ϵ structure bulge contains 4 nt (TTCA) that act as the template for synthesizing the 4 nt primer to start minus DNA strand synthesis. This small sequence shifts from the 5′ bulge to an acceptor site (AS) located in the 3′ direct repeat region 1 (DR1) (Figure 1b). Conservation of the 4-nt bulge and AS in DR1 seemed to be essential for HBV replication; hence, both regions were analyzed. Variants occurring at a frequency of >50 (2-fold the most prevalent variants in the HindIII control sequence: >0.03%) were considered significant. A total of 189 737 sequences were included, and the vast majority (98.53%) showed the same sequence in the bulge and DR1 regions. The remaining sequences ($N = 2783$, 1.47%) contained changes that prevented complete annealing of 4 nt primer in the AS DR1 site, due to a single nucleotide substitution. Less than a third of mutated sequences ($N = 920$, 0.48%) contained the change only in the bulge sequence, while in the other cases ($N = 1863$, 0.98%), the change was only present in the AS of DR1.

The 2783 sequences without correct homology between the 4 nt primer (TTCA) and the AS DR1 site were classified into nine different variants by homology of the bulge sequence (Figure 4): variants 1–4 were defined by mutations within the bulge (Positions 1863–1866) and

Left table — AA changes / Nucleotide Changes / Master Sequence

%MT	WT	N	A	C	T	G	%Tot	Pos	Base
			0	0,17	0	0	0,18	1880	T
			0,06	0	0	0	0,06	1879	G
0,21% P	L	22	0,03	0,21	0	0,01	0,26	1878	T
			0,02	0	0,06	0	0,08	1877	C
0,13% R / 0,10% E	K	21	0,06	0	0,01	0	0,07	1876	G C
			0	0,01	0,02	0,13	0,17	1875	A T
			0	0,01	0,01	0,10	0,13	1874	A T
0,21% P	S	20	0,01	0	0,06	0	0,08	1873	C G
			0	0	0,06	0	0,07	1872	C G
			0,02	0,21	0	0	0,24	1871	T G
	A	19	0,04	0	0,05	0	0,09	1870	C G
			0	0	0,09	0	0,10	1869	C G
			0,07	0	0,03	0	0,10	1868	C G
0,15% R	Q	18	0	0,04	0,05	0,14	0,23	1867	A T
			0	0	0	0,15	0,16	1866	A T
			0	0	0,04	0	0,05	1865	C
0,25% A	V	17	0	0,07	0	0	0,07	1864	T
			0	0,25	0	0	0,26	1863	T
			0,07	0	0,01	0	0,09	1862	G
	T	16	0,01	0,20	0	0	0,22	1861	T
			0	0	0,07	0	0,08	1860	C
			0	0	0	0,07	0,08	1859	A
0,15% S	P	15	0,01	0	71,57	0	71,59	1858	C G
			0,06	0	0,09	0	0,15	1857	C G
			0	0	0,15	0	0,16	1856	C G
0,2 % R	C	14	0,04	0,30	0	0	0,34	1855	T G
			0,06	0	0	0	0,08	1854	G C
			0	0,20	0	0	0,21	1853	T A
69,6% S	T	13	0	0	0,01	0,12	0,15	1852	A T
			0	0	0,10	0	0,11	1851	C G
			0	0,07	68,62	0,04	68,73	1850	A
0,014% R	C	12	0,01	0,29	0	0	0,30	1849	T
			0,07	0	0	0	0,08	1848	G
			0	0,14	0	0	0,14	1847	T
0,23% P	S	11	51,77	0,12	0	0,05	51,94	1846	T
			0,01	0	0,06	0	0,07	1845	C
			0	0,23	0	0	0,24	1844	T
16,9% T	I	10	0,01	0	0,05	0	0,06	1843	C
			0	16,89	0	0	16,90	1842	T
			0	0	0	0,11	0,11	1841	A
3,1% V	I	9	0,02	0	0,05	0	0,07	1840	C
			0	0,09	0	0	0,10	1839	T
			0	0,01	0,02	3,07	3,10	1838	A
0,18% P	L	8	0	0	0,01	0,05	0,07	1837	A
			0,02	0,18	0	0	0,21	1836	T
			0,02	0	0,06	0	0,07	1835	C
0,19% R	C	7	0,04	0	0,08	0	0,12	1834	C
			0,06	0,03	0,03	0	0,12	1833	G
			0,01	0,19	0	0	0,21	1832	T
0,15% P	L	6	0,01	0	0,05	0	0,06	1831	T
			0,02	0,15	0	0	0,18	1830	T
			0,02	0	0,06	0	0,08	1829	C
0,25% P / 0,16% Y / 0,15% R	H	5	0,05	0	0,08	0	0,14	1828	C
			0	0,25	0,02	0,15	0,42	1827	A
			0	0	0,16	0	0,17	1826	C
0,3% L / 0,22% S	F	4	0	0,22	0	0	0,23	1825	T
			0	0,22	0	0	0,23	1824	T
			0,02	0,29	0	0	0,32	1823	T
0,33% P / 0,1% H	L	3	0,08	0,13	0	0,01	0,22	1822	T
			0,10	0,33	0	0,02	0,45	1821	T
			0	0	0,05	0	0,06	1820	C
0,1% R / 0,07 Stop	Q	2	0	0,02	0,01	0,10	0,13	1819	A
			0	0	0,02	0,10	0,13	1818	A
			0	0	0,06	0	0,07	1817	G
16,43% L * / 0,14% I * / 0,14% T *	M	1	0,14	0	0,02	0	0,16	1816	G
			0,01	0,18	0	0,01	0,21	1815	G
			0	16,46	0,03	0,07	16,56	1814	A

Apical Loop

Right table — Master Sequence / Nucleotide Changes / AA changes (Apical Stem and Basal Stem)

Pos	Base	%Tot	A	C	T	G	N	WT	%MT
1881		0,05	0,06	0	0	0			
1882	C	0,15	0	0	0,15	0	23	C	0,17% R
1883	G	0,07	0,02	0	0,05	0			
1884	T	0,21	0,04	0,15	0	0,01	24	L	0,15% P
1885	T	0,15	0	0,14	0	0			
1886	G	0,15	0,09	0	0,05	0			
1887	G	0,12	0,06	0	0,04	0	25	G	
1888	G	0,10	0,09	0	0,01	0			
1889	T	0,20	0	0,19	0	0			
1890		0,09	0,07	0	0,02	0	26	W	0,19% R / 0,17% Stop *
1891		0,11	0,10	0	0	0			
1892		0,10	0,03	0	0,07	0			
1893		0,16	0,01	0,13	0	0,01	27	L	0,13% P
1894		0,05	0	0,04	0	0			

Bulge

Pos	Base	%Tot	A	C	T	G	N	WT	%MT
1895	T	0,08	0	0,06	0	0,01			
1896	G	45,00	44,60	0	0,10	0	28	W	45,2% Stop *
1897	G	0,18	0,08	0	0,09	0			
1898	G	0,10	0,06	0	0,04	0			
1899	C	28,35	28,33	0	0,01	0	29	G	28,3% D
1900		0,09	0,01	0	0,08	0			
1901	A	0,14	0	0	0	0,13	1 Core	M	0,13% V
1902	T	0,13	0,09	0,03	0	0			0,1% K
1903	C	0,13	0,13	0	0	0			0,13% I

Center labels: Apical Loop · Apical Stem · Bulge · Basal Stem · Acceptor Site

Figure 3. Schematic representation of base pairing between the preCore positions analyzed, which correspond to the stem–loop structure. The columns in the center indicate the consensus sequence obtained from the laboratory's entire collection of genotypes A sequences (bold characters). The 'Nucleotide changes' column indicates the percentage of each nucleotide change (%A, %C, %T and %C) and the total changes (%Tot) for all positions, attending to the master sequence. The 'Amino acid (AA) changes' column indicates the codon representing the position (N), the amino acid coded from the master sequence (WT), and the percentage of sequences that code for a different amino acid (%MT).

10   Nucleic Acids Research, 2011

variants 5–9 by mutations within the DR1 AS (Positions 1824–1827). Sequences from variants 5–9, which represented 1232 cases (44 0.65% of total sequences), showed complete 4-nt homology with a sequence located in Positions 1849–1852 (TTCA corresponding to HBV

Table 4. Epsilon energy distribution attending to HBV genotype and the presence or absence of mutations in the preCore (codons 1 and 28)

Genotype	28	1	
		WT	MT
A2	WT	−25.44$_a$	−25.98$_c$
	MT	−24.30$_b$	−23.79$_d$
D	WT	−23.76$_e$	−24.81$_g$
	MT	−24.39$_f$	−25.48$_h$

a versus b: $P = 0.001$; a versus c: $P = 0.006$; a versus d: $P = 0.018$; b versus c: $P = 0.001$; b versus d: $P = NS$; c versus d: $P = 0.001$; e versus f: $P = 0.0015$; e versus g: $P = 0.0210$; e versus h: $P = 0.0015$; f versus g: $P = NS$; f versus h: $P = 0.0210$; g versus h: $P = NS$ a versus e: $P = 0.002$; b versus f: $P = NS$; e versus g: $P = 0.007$; d versus h: $P = 0.013$.

Table 5. Average expression of HBsAg, HBeAg, and HBV DNA in the different transfection experiments

Experiment	HBsAg (IU/ml)	HBeAg (arbitrary units)	HBV DNA (log IU/ml)
Wild-type (T1855)	65.89	5.52	6.5
1855C mutation in 5′ and 3′	55.47	3.58	6.1
1855C mutation in 5′	115.90	6.96	6.3
1855C mutation in 3′	175.37	11.70	6.8

genotype D), which was considered a putative acceptor site (AS3). Variant 3, with 119 sequences (4, 0.06% of the total) showed 3-nt homology with an upstream AS3 sequence. The sequence groups of variants 2, 4, 8 and 9 ($N = 864$ cases; 31, 0.46% of the total) showed 3-nt homology with alternatives to the AS sequences located at Positions 1818–1823 ['ω' or AS2 in Positions 1835–1846 (27)]. Lastly, 568 sequences (20.5, 0.28% of the total) had an AS canonical region. In summary, when sequences of the 4 nt bulge primer and its AS DR1 canonical site were not completely complementary, 44% of cases showed 4-nt homology with the putative AS3 acceptor site (HBV genotype A2 characteristic sequence) while the remaining 56% showed only 3-nt homology, mainly located in previously described alternative acceptor sites (32).

YMDD analysis

The YMDD motif was analyzed in four baseline samples and in four samples containing LMV-resistant mutations, detected by InnoLipa. Percentages of the sequences obtained with different nucleotides are shown in Table 3; the columns indicate the amino acid detected. The wild-type form was represented as methionine (M), and mutants conferring resistance were valine (V) and iso-leucine (I). Amino acids other than M, V or I, detected in a lower percentage, were grouped as 'Others'. As was seen in the baseline samples, the wild-type form of Position 204 was highly predominant (>99%), but in the four samples, a low percentage of the main rtM204V/I mutant forms (0.07–0.2%) and a significant percentage of other variants (0.09–0.12%) were detected. In the four samples corresponding to a time of viral breakthrough (VBK) analyzed, the LMV-resistant variants rtM204V/I were the main populations (98.7–99.7%). However, low
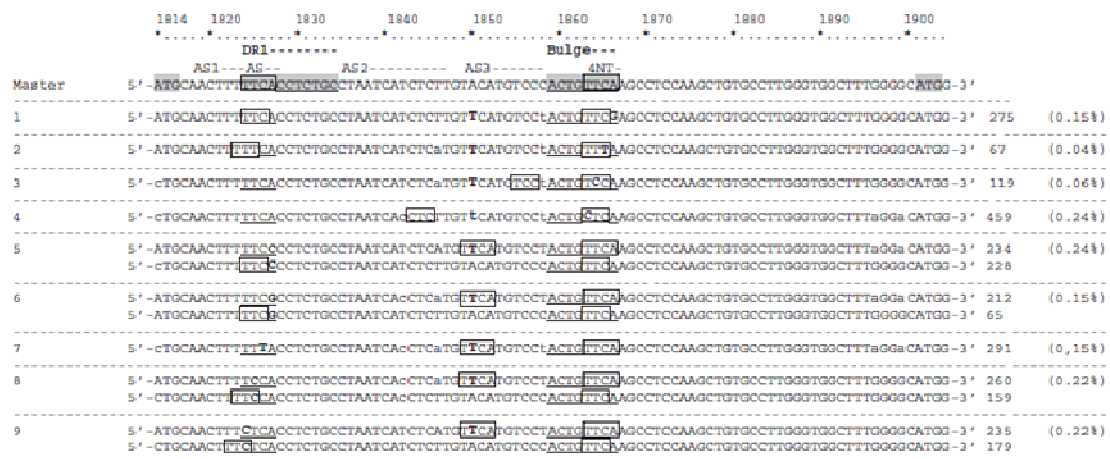
Figure 4. Sequences without correct homology between the 4 nt primer in the bulge (1863–1866) and acceptor site (AS) in direct repeat region 1 (DR1) (1824–1827). The genotype A2 consensus sequence from our laboratory collection used as the pattern is presented in the first line, and variants are added below. The number of sequences of each variant and the percentage of the total represented is indicated at the end of each line. Bold characters indicate mutations from the master, and lower cases indicate a mixture with the nucleotide of the consensus sequence. Boxes indicate homologous sequences and suggest alternative ASs.

Table 6. Percentage of sequences, depending on the presence or absence of mutations in YMDD and preCore positions

YMDD	preCore		Baseline Samples				VBK samples			
			Pt 1 HBe N Gen D	Pt 2 HBe N Gen A2	Pt 3 HBe N Gen A2	Pt 4 HBe P Gen D	Pt 4 HBe P Gen D	Pt 5 HBe N Gen A2	Pt 6 HBe P Gen A2	Pt 7 HBe N[?] Gen D
204	Codon 1-2[a]	Codon 28								
Wt	Wt	Wt	0.20	0.05	0.04	**90.93**	0.25	0.00	0.05	0.00
Wt	Mt	Wt	0.00	**99.41**	0.00	0.55	0.00	0.03	0.00	0.00
Wt	Wt	Mt	**99.21**	0.01	**99.71**	**7.82**	0.00	0.04	**1.01**	0.07
Wt	Mt	Mt	0.30	0.16	0.14	0.03	0.00	0.01	0.00	0.00
Mt	Wt	Wt	0.01	0.00	0.00	0.21	**97.94**	0.13	**88.78**	0.29
Mt	Mt	Wt	0.00	0.29	0.00	0.00	0.59	0.00	**9.86**	0.01
Mt	Wt	Mt	0.29	0.08	0.11	0.46	**1.20**	**99.24**	0.28	**99.03**
Mt	Mt	Mt	0.00	0.00	0.00	0.00	0.02	0.56	0.02	0.60

The negative (N) or positive (P) status of HBe antigen expression and HBV genotype (Gen A2 or D) are indicated in each patient column.
Bold values indicate percentages higher than 1%.
[a]WT codon 1 and 2 are not mutated. MT codon 1 and/or codon 2 are mutated. However, mutations are mainly found in codon 1.

percentages of wild-type sequences were also detected (0.07–1.06%). In three patients with VBK (cases 4, 6 and 7) the rtM204V mutation was predominant, but Patient 4 had the predominant mutation in combination with a considerable percentage of rtM204I (5.6%). The rtM204I mutation alone predominated in Patient 5, but a small percentage (0.2%) of rtM204V was also detected. These results agree with those obtained by Lipa (Table 1). Interestingly, the percentage of 'Other' variants in Position 204 (0.23–0.27%) was higher in VBK samples than in baseline samples, suggesting that these variants might also be involved in LMV resistance.

Simultaneous detection of preCore and polymerase gene mutations in the same HBV genome

The possibility of simultaneous mutations in the YM204DD Pol motif (Position 204) and preCore (codon 1, 2 and 28) genes was analyzed in the eight circular constructs processed. Sequences were grouped according to the presence or absence of mutations in the four regions (preCore codons 1 and 2 were analyzed together), and the percentages found are shown in Table 6. In Patients 5 and 7, who were both HBeAg-negative and had mutations against LMV, the sequences showing simultaneous mutations in the preCore and Pol genes comprised the major populations. Interestingly, sequences carrying mutations in the preCore and Pol genes were detected in significant percentages in all cases, particularly Patient 6, in whom 9.9% of sequences showed simultaneously occurring LMV mutations and preCore codon 1 mutations.

Sequential study of the preCore region and YMDD motif

In Patient 4, the circular construct was obtained from samples taken at baseline and at VBK. Mutations conferring LMV resistance were detected in the baseline sample (0.55%), and were much more prevalent at VBK (99.54%) (Table 3). Although the rt204I mutation was detected in a higher percentage than rt204V at baseline (0.47% versus 0.08%), rt204V was more frequent (93.9%) than rt204I (5.63%) at VBK. As to preCore mutations, the percentage of stop codon 28 significantly

decreased at VBK (0.12%) relative to the baseline sample (8.31%) (Table 3).

Start codon of the core gene

The circular construct included the start codon of the Core gene, and significant percentages (0.11–0.91%) of mutations in this codon were detected in all samples processed (Table 3). It must be kept in mind that these mutations abolished Core protein production; hence HBV genomes carrying these changes would be defective for viral capsid production.

preCore fragment overlapping the HBx gene

The HBx gene spans Position 1374–1838 of the HBV genome, overlapping the preCore gene in the last eight codons (Positions 1814–1838). For this reason, mutations in the HBx stop codon (Positions 1836–1838) and its implications on the preCore ORF were analyzed. Although most sequences (194 423; 96.82%) had the expected TAA stop codon, 5837 sequences (2.9%) carried the TAG amber stop codon, which represents an isoleucine to valine amino acid substitution in the overlapping preCore region (ATC to GTC). Of note, a significant percentage of sequences (324, 0.16%) showed a T1836C substitution, which changes the ochre stop codon TAA to CAA (Gln). After this substitution, the HBx protein would be translated to the next in-frame stop codon TAG at Position 1992, inducing expression of 51 additional amino acids at the COOH-terminal region. In the preCore ORF, the T1836C mutation also induces a change of leucine to proline. The remaining 233 sequences showed other amino acid substitutions, which were not analyzed because they occurred at a frequency below our established cut-off.

DISCUSSION

The circular construct developed in the present study provided information about the HBV RT YMDD Pol and preCore regions, thus enabling analysis of simultaneous mutations in both regions of the same genome. The obligate presence of the 6-bp HindIII sequence, which

linked the two regions, was used as an internal control of fidelity of the UDPS process. The calculated nucleotide error rate (1.6×10^{-4}), with more transitions than transversions (60% versus 40%), was similar to reported error rate values obtained with clonal DNA as the external control (18–20,26,25), thus validating the use of HindIII as the exclusive error rate control. Nonetheless, to be more precise and concise, only variants in a percentage higher than twice this error rate (>0.03%) were taken into consideration.

Analysis of the YMDD motif of the Pol gene confirmed the presence of low but significant percentages (0.07–0.55%) of LMV resistance mutations in baseline samples from all patients studied, thus illustrating the complexity of the HBV quasispecies. Interestingly, significant populations of other YMDD variants (0.09–0.12%) were also detected at baseline. At VBK, the rtM204V/I variants were the main populations detected (99.7–93.9%), but wild-type sequences were also found in small percentages (0.07–1.06%). Once again, the presence of YMDD variants was associated with LMV resistance (14). Variant rtM204V was predominant in three of four cases of VBK, whereas rtM204I was predominant in only one case. In the sequentially studied case, rt204I was detected in a higher percentage than rt204V at baseline (0.47% versus 0.08%), but after VBK, rt204V was more prevalent (93.9%) than rt204I (5.63%). This suggests that the type and relative frequency of mutations at baseline does not predict the mutations selected during treatment (33).

The high percentages of viral populations with preCore and LMV-resistant mutations observed in LMV-resistant anti-HBe cases seem to indicate that mutations in these regions simultaneously occur in the same genome. In the limited available studies directly investigating this simultaneous presence (22,23), there were no correlations of LMV-resistant mutations with HBV genotype or the presence (or not) of the preCore stop codon. Furthermore, *in vitro* study has shown that progeny DNA levels were restored in constructs with preCore and LMV mutations (34,35). Only a small number of clones were analyzed in the previous studies (22,23) because the preCore and YMDD regions are more than 1 kb apart, a fact that makes their study difficult. When the present research was performed, clonal analysis by UDPS was limited to a 250-bp sequence length (~400 bp in its current version). Nonetheless, the methodology reported here enabled ultra-deep clonal study of both types of variants in the same genome. The results obtained with this novel PCR-based technique support the idea that simultaneous mutations in the preCore region and YMDD Pol motif are common in the HBV quasispecies in CHB patients, regardless of the viral genotype (A2 or D) or HBeAg status.

The presence of genomes without the Core antigen start codon suggests the existence of minor viral populations that are defective for viral capsid production, another indication of the complexity of HBV viral quasispecies. The presence of these genomes in HBV infection can be explained by possible collaboration between viruses from the same infection by a trans-complementation mechanism: if wild-type and Core-defective genomes coinfect the same liver cells, Core proteins produced in excess by the wild-type viral genome might be used to encapsidate Core-defective HBV genomes. Possible reduction of HBV core proteins in the cytoplasm pool due to the presence of Core-defective genomes could decrease their expression in the cell membrane, thus modulating host immune pressure and maintaining the infection. Therefore, the presence of defective genomes might establish cooperation with wild-type viruses, thus favoring viral fitness. This mechanism of trans-complementation has been described in *in vitro* studies analyzing defective HBV Core clones (36). Further mutagenic studies are needed to evaluate the possible decrease in host immune response in this situation. Research evaluating possible roles of the minor (0.16%) mutated sequences detected (additional amino acid in the HBx COOH-terminal peptide due to a mutation in the HBx stop codon) would also help to clarify the role of the HBx gene in HCC pathogenesis.

In agreement with previous studies (37), the two HBeAg-positive cases had significantly higher percentages of preCore stop mutations in codon 28 (8.31 and 1.31%), suggesting easy selection of preCore mutations when the host immune response increases during viral infection. Likewise, in HBeAg-negative patients, wild-type codon 28 preCore genomes were detected in lower percentages (0.02–0.14%), explaining cases of natural or LMV-related seroreversion (38,39). An interesting observation in the longitudinally studied patient was the significant decrease in the main preCore mutation at VBK, which coincides with reports suggesting that HBV strains carrying preCore stop codon 28 and no Pol resistant variants are more sensitive to LMV treatment than preCore wild-type sequences (35,38,39). However, additional longitudinal UDPS analyses, as reported in the present study, must be performed to further substantiate this possibility.

Epsilon elements are present twice on pgRNA, one located near the 5′ terminus and the other at the 3′ terminus. Correct base-pairing between Position 1896 (codon 28) and 1858 (codon 15) from the lower stem of the encapsidation signal is one of the main constraints for preCore variability and ϵ thermodynamic stability, as has been reported previously (8) and was highly reinforced by our results. Regarding HBV genotypes, the main preCore mutation (A1896G variant) was clearly predominant in genotype D, while in genotype A2, significant percentages of preCore mutation were located at codon 1. This thermodynamic restriction between HBV genotypes corroborates results from previous studies using clonal or direct sequencing techniques (8,11).

After initial ϵ-Pol binding, the HBV apical stem loop of ϵ has to open up to form a primer synthesis-competent complex, but the highly stable HBV upper stem is a barrier. The free energy to overcome this barrier may come from interaction with RT and by capping the TGT pseudo-triloop with capsid proteins (31). Highly conserved positions and maintenance of correct base pairing in the upper stem were observed in our results. A change of T to C in Positions 1884, 1885 and 1893

was found in 0.15, 0.14 and 0.13%, respectively, but interestingly, the complementary change of A to G in their base-paired Positions (1875, 1874 and 1867) showed very similar percentages (0.13, 0.1 and 0.14%, respectively). Additionally, the low percentage (0.1%) of G to A changes detected in Position 1891, which in fact represent a new stop codon in the preCore sequence, was also compensated with complementary C to T changes in a similar percentage (0.09%) in the 1869 base-paired position. Observation of correct base pairing of the apical and upper stem of ε might reinforce the idea that structural rather than sequence restrictions are involved in conservation of the ε encapsidation signal.

The hexanucleotide sequence located at the top of the apical stem (from 1877 to 1882) was found to be highly conserved among 1200 HBV strains (40). MR analysis (41) has shown that the $_{1878}TGT_{1880}$ sequence folds in a pseudo-triloop (Figure 1), a structure correctly predicted by the RNA folding program used in the present study (27,28). The pseudo-triloop is not needed for RT binding, but is required for pgRNA encapsidation, suggesting its interaction with capsid proteins (29,42). High conservation of T_{1878}, G_{1879} and T_{1880} nucleotides was observed in our study, particularly G_{1879}, which suggests that they have a predominant role in the ε-capsid interaction (43). Structural and functional experiments must be performed to analyze the effect of these changes on the HBV replication process.

The mutation combination of G1899A plus G1896A was frequently detected in our study. Guarnieri *et al.* (36) reported that G1899A does not significantly increase *in vitro* viral replication, despite significant theoretical stabilization of the ε structure. G1899 is found in the Kozac sequence and precedes the start codon of HBcAg. The reason why G1899A is a highly prevalent change, in contrast to its paired stem position, the conserved T1855C, remains unknown. According to our results, the $_{1855}T{:}G_{1899}$ to $_{1855}T{:}A_{1899}$ change was 100 times more frequent than the $_{1855}T{:}G_{1899}$ to $_{1855}C{:}G_{1899}$ change. Furthermore, in contrast to previously reported results (36), the site-directed mutagenesis and the transfection studies showed that presence of the T1855C mutation in both the 5′- and 3′-ends led to a slight, but systematic, reduction in HBsAg, HBeAg and HBV-DNA levels. If T_{1855} were to have an essential role, selection of the C substitution would be limited and explain the low prevalence of this change. The most striking findings of our study were that chimerical (but not natural) constructs with only one 1855C mutation at the 5′- or 3′-end were associated with higher HBsAg, HBeAg and HBV-DNA levels than wild-type clones. Although the 5′ and 3′ structures have differing functions in the singular HBV replication cycle (4,5), enhancement of viral replication activity in chimerical constructs with a single mutation remains unexplained. Specific experiments using primary hepatocyte cells, such as macaque hepatocytes or other *in vitro* cell systems are needed to explore the possible involvement of T_{1855}.

Analysis of nucleotide variability in the 4-nt bulge primer (4 nt primer) used as a template for synthesis of the minus DNA strand and its corresponding AS in the DR1 region confirmed the high conservation of both regions. It has been speculated that the 4 nt primer is more essential than DR1 AS. The existence of alternative acceptor sites (32) is suggested, as well as the idea that 3-nt sequences would be completely functional for minus strand DNA synthesis (36). Our results support these concepts because the 4 nt bulge primer was slightly more conserved than AS DR1 (99.5% versus 98.9%). As was expected, the ε structural motifs were strongly conserved in our study; very few cases (1.47%) had different sequences in the 4 nt primer and its canonical DR1 AS. However, 44% showed complete 4-nt homology with the putative acceptor site, AS3 (Positions 1849–1852), whereas the remaining 56% showed only 3-nt homology with canonical DR1 AS or with sequences located in previously reported alternative acceptor sites (32). Therefore, our results seem to agree with the existence of an alternative to the DR1 acceptor sites for the 4 nt primer (32). Analysis of the theoretical secondary structure of the 3′-end ε signal (Figure 1b) showed that the AS3 sequence has tighter Watson–Crick pairing than canonical DR1 AS. This fact could represent an energy barrier for 4 nt annealing, and explain the low prevalence of mutated bulge variants to enable correct annealing of 4 nt with AS3.

The high cost of the pyrosequencing method used in this research is a limit to the number of patients that can be studied, but the large number of sequences obtained for each patient provides deeper information on the HBV viral quasispecies than has been previously reported. The construct was designed to evaluate specific positions in the preCore and polymerase regions, but not the entire gene. Thus, our conclusions are restricted to the YMDD and preCore regions. Because of the same cost concerns, only LMV-treated patients were selected; other RT positions associated with LMV and other antiviral treatments were not included. The circular construct designed for this study was useful for analyzing two widely separated regions. This method can be applied to other genes if a fragment with the two regions of interest at the ends can be amplified by PCR.

In conclusion, UDPS study confirmed the feasibility of detecting simultaneously occurring mutations in the preCore region and YMDD polymerase motif. The presence of LMV-resistant variants in baseline naïve samples, detection of preCore mutations in HBeAg-positive patients and the presence of WT preCore variants in HBeAg-negative cases were also demonstrated. Low percentages of defective Core genomes and mutated HBx strains were found, illustrating the complexity of the HBV quasispecies. Viral strains present in small percentages can act as reservoirs and be selected in response to external changes (e.g. administration or elimination of drug pressure, or an increased host immune response). Genomes with mutations in the start codon of the Core gene might result in defective particles, whose presence might be allowed due to cooperation between viruses. The thermodynamic stability of the ε signal was confirmed to be the principal restriction for selection of the main preCore mutation, which is responsible for abolishment of HBeAg expression. Furthermore, analysis of ε signal variability revealed the essential role of

structural ϵ motifs and possible involvement of some nucleotides in ϵ signal function. Lastly, we found that correct annealing between the 4 nt primer of the bulge and the AS does not seem to be an absolute condition for the 4 nt primer switch needed for minus DNA strand synthesis, suggesting that alternative *cis*-acting sequences, such AS3 defined in the present report, might contribute to define the normal acceptor site in the 3' of pgRNA.

REFERENCES

1. Nassal,M. (2008) Hepatitis B viruses: reverse transcription a different way. *Virus Res.*, 134, 235–249.
2. Kim,S., Lee,J. and Ryu,W.S. (2009) Four conserved cysteine residues of the hepatitis B virus polymerase are critical for RNA pregenome encapsidation. *J. Virol.*, 83, 8032–8040.
3. Wang,G.H. and Seeger,C. (1993) Novel mechanism for reverse transcription in hepatitis B viruses. *J. Virol.*, 67, 6507–6512.
4. Beck,J. and Nassal,M. (1998) Formation of a functional hepatitis B virus replication initiation complex involves a major structural alteration in the RNA template. *Mol. Cell. Biol.*, 18, 6265–6272.
5. Rieger,A. and Nassal,M. (1996) Specific hepatitis B virus minus-strand DNA synthesis requires only the 5'ε encapsidation signal and the 3'-proximal direct repeat DR1. *J. Virol.*, 70, 585–589.
6. Carman,W.F., Jacyna,M.R., Hadziyannis,S., Karayiannis,P., McGarvey,M.J. and Makris,A.T.H. (1989) Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet*, 9, 588–591.
7. Okamoto,H., Yotsumoto,S., Akahane,Y., Yamanaka,T., Miyazaki,Y., Sugai,Y., Tsuda,F., Tanaka,T., Miyakawa,Y. and Mayumi,M. (1990) Hepatitis B viruses with precore region defects prevail in persistently infected hosts along with seroconversion to the antibody against e antigen. *J. Virol.*, 64, 1298–1303.
8. Rodriguez-Frias,F., Buti,M., Jardi,R., Cotrina,M., Viladomiu,L., Esteban,R. and Guardia,J. (1995) Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. *Hepatology*, 22, 1641–1647.
9. Jardi,R., Rodriguez,F., Buti,M., Costa,X., Valdes,A., Allende,H., Schaper,M., Galimany,R., Esteban,R. and Guardia,J. (2004) Mutations in the basic core promoter region of hepatitis B virus. Relationship with precore variants and HBV genotypes in a Spanish population of HBV carriers. *J. Hepatol.*, 40, 507–514.
10. Pollack,J.R. and Ganem,D. (1994) Site-specific RNA binding by a hepatitis B virus reverse transcriptase initiates two distinct reactions: RNA packaging and DNA synthesis. *J. Virol.*, 68, 5579–5587.
11. Lok,A.S., Akarca,U. and Greene,S. (1994) Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *Proc. Natl Acad. Sci. USA*, 91, 4077–4081.
12. Kidd,A.H. and Kidd-Ljunggren,K. (1996) A revised secondary structure model for the 3ε-end of hepatitis B virus pregenomic RNA. *Nucleic Acids Res.*, 24, 3295–3301.
13. Dienstag,J.L., Perrillo,R.P., Schiff,E.R., Bartholomew,M., Vicary,C. and Rubin,M. (1995) A preliminary trial of lamivudine for chronic hepatitis B infection. *N. Engl. J. Med.*, 333, 1657–1661.
14. Allen,M.I., Deslauriers,M., Andrews,C.W., Tipples,G.A., Walters,K.A., Tyrrell,D.L., Brown,N. and Condreay,L.D. (1998) Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. *Hepatology*, 27, 1670–1677.
15. Lok,A.S.F., Lai,C.L., Leung,N., Yao,G.B., Cui,Z.Y., Schiff,E.R., Dienstag,J.L., Heathcote,E.J., Little,N.R., Griffiths,D.A. *et al.* (2003) Long-term safety of lamivudine treatment in patients with chronic hepatitis B. *Gastroenterology*, 125, 1714–1722.
16. Chien,R.N., Yeh,C.T., Tsai,S.L., Chu,C.M. and Liaw,Y.F. (2003) Determinants for sustained HBeAg response to lamivudine therapy. *Hepatology*, 38, 1267–1273.
17. Lok,A.S., Zoulim,F., Locarnini,S., Bartholomeusz,A., Ghany,M.G., Pawlotsky,J.M., Liaw,Y.F., Mizokami,M. and Kuiken,C. (2007) Antiviral drug-resistant HBV: standardization of nomenclature and assays and recommendations for management. *Hepatology*, 46, 254–265.
18. Zagordi,O., Klein,R., Däumer,M. and Beerenwinkel,N. (2010) Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res.*, 38, 7400–7409.
19. Wang,C., Mitsuya,Y., Gharizadeh,B., Ronaghi,M. and Shafer,R.W. (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res.*, 17, 1195–1201.
20. Solmone,M., Vincenti,D., Prosperi,M.C.F., Bruselles,A., Ippolito,G. and Capobianchi,M.R. (2009) 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.*, 83, 1718–1726.
21. Margeridon-Thermet,S., Shulman,N.S., Ahmed,A., Shahriar,R., Liu,T., Wang,C., Holmes,S.P., Babrzadeh,F., Gharizadeh,B., Hanczaruk,B. *et al.* (2009) 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.*, 199, 1275–1285.
22. Lok,A.S., Hussain,M., Cursano,C., Margotti,M., Gramenzi,A., Grazi,G.L., Jovine,E., Benardi,M. and Andreone,P. (2000) Evolution of hepatitis B virus polymerase gene mutations in hepatitis B e antigen-negative patients receiving lamivudine therapy. *Hepatology*, 32, 1145–1153.
23. Cho,S.W., Hahm,K.B. and Kim,J.H. (2000) Reversion from precore/core promoter mutants to wild-type hepatitis B virus during the course of lamivudine therapy. *Hepatology*, 32, 1163–1169.
24. Rodriguez-Frias,F., Jardi,R., Schaper,M., Gimferrer,M., Elefsiniotis,I., Tabernero,D., Esteban,R. and Buti,M. (2007) Redetection of HBV lamivudine-resistant mutations in a patient under entecavir therapy, who had been treated sequentially with nucleos(t)ide analogues. *J. Med. Virol.*, 79, 1671–1673.
25. Campbell,P.J., Pleasance,E.D., Stephens,P.J., Dicks,E., Rance,R., Goodhead,I., Follows,G.A., Green,A.R., Futreal,P.A. and Stratton,M.R. (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl Acad. Sci. USA*, 105, 13081–13086.
26. Huse,S.M., Huber,J.A., Morrison,H.G., Sogin,M.L. and Welch,D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.*, 8, R143.
27. Reuter,J.S. and Mathews,D.H. (2010) RNA structure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11, 129.
28. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical

modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.

29. Flodell,S., Petersen,M., Girard,F., Zdunek,J., Kidd-Ljunggren,K., Schleucher,J. and Wijmenga,S. (2006) Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res.*, **34**, 4449–4457.

30. Durantel,D., Carrouée-Durantel,S., Werle-Lapostolle,B., Brunelle,M.N., Pichoud,C., Trépo,C. and Zoulim,F. (2004) A new strategy for studying in vitro the drug susceptibility of clinical isolates of human hepatitis B virus. *Hepatology*, **40**, 855–864.

31. Hu,J. and Boyer,M. (2006) Hepatitis B virus reverse transcriptase and ε RNA sequences required for specific interaction *in vitro*. *J. Virol.*, **80**, 2141–2150.

32. Abraham,T.M. and Loeb,D.D. (2007) The topology of hepatitis B virus pregenomic RNA promotes its replication. *J. Virol.*, **81**, 11577–11584.

33. Yuen,M.F., Tanaka,Y., Fong,D.Y.T., Fung,J., Wong,D.K.-H., Yuen,J.C.H., But,D.Y.K., Chan,A.O.O., Wong,B.C.Y., Mizokami,M. *et al.* (2009) Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. *J. Hepatol.*, **50**, 80–88.

34. Chen,R.Y.M., Edwards,R., Shaw,T., Colledge,D., Delaney,W.E., Isom,H., Bowden,S., Desmond,P. and Locarnini,S.A. (2003) Effect of the G1896A precore mutation on drug sensitivity and replication yield of lamivudine-resistant HBV in vitro. *Hepatology*, **37**, 27–35.

35. Tacke,F., Gehrke,C., Luedde,T., Heim,A., Manns,M.P. and Trautwein,C. (2004) Basal core promoter and precore mutations in the hepatitis B virus genome enhance replication efficacy of lamivudine-resistant mutants. *J. Virol.*, **78**, 8524–8535.

36. Guarnieri,M., Kim,K.-hwan, Bang,G., Li,J., Zhou,Y., Tang,X., Wands,J. and Tong,S. (2006) Point mutations upstream of hepatitis B virus core gene affect DNA replication at the step of core protein expression. *J. Virol.*, **80**, 587–595.

37. Volz,T., Lutgehetmann,M., Wachtler,P., Jacob,A., Quaas,A., Murray,J.M., Dandri,M. and Petersen,J. (2007) Impaired intrahepatic hepatitis B virus productivity contributes to low viremia in most HBeAg-negative patients. *Gastroenterology*, **133**, 843–852.

38. Chen,C.H., Lee,C.M., Lu,S.N., Changchien,C., Wang,J.C., Wang,J.H., Hung,C.H. and Hu,T.H. (2006) Comparison of sequence changes of precore and core promoter regions in HBeAg-positive chronic hepatitis B patients with and without HBeAg clearance in lamivudine therapy. *J. Hepatol.*, **44**, 76–82.

39. Kuwahara,R., Kumashiro,R., Murashima,S., Ogata,K., Tanaka,K., Hisamochi,A., Hino,T., Ide,T., Tanaka,E., Koga,Y. *et al.* (2004) Genetic heterogeneity of the precore and the core promoter region of genotype C hepatitis B virus during lamivudine therapy. *J. Med. Virol.*, **72**, 26–34.

40. Flodell,S., Schleucher,J., Cromsigt,J., Ippel,H., Kidd-Ljunggren,K. and Wijmenga,S. (2002) The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.*, **30**, 4803–4811.

41. Girard,F.C., Ottink,O.M., Ampt,K.A.M., Tessari,M. and Wijmenga,S.S. (2007) Thermodynamics and NMR studies on duck, heron and human HBV encapsidation signals. *Nucleic Acids Res.*, **35**, 2800–2811.

42. Ampt,K.A.M., van der Werf,R.M., Nelissen,F.H.T., Tessari,M. and Wijmenga,S.S. (2009) The unstable part of the apical stem of duck hepatitis B virus epsilon shows enhanced base pair opening but not pico- to nanosecond dynamics and is essential for reverse transcriptase binding. *Biochemistry*, **48**, 10499–10508.

43. Petzold,K., Duchardt,E., Flodell,S., Larsson,G., Kidd-Ljunggren,K., Wijmenga,S. and Schleucher,J. (2007) Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. *Nucleic Acids Res.*, **35**, 6854–6861.

Third Study

4 THIRD STUDY: Variability and dynamics of hepatitis B virus quasispecies reflected by ultra-deep pyrosequencing of the main epitopic regions of the HBV core gene

4.1 Hypothesis and Aims

4.1.1 Introduction

The present work was a continuation of the first study, in which we observed high variability of changes in the main epitopic regions between baseline sequences and their consensus genotype. After application of UDPS in the second study, we concluded that this method was a powerful tool to analyze viral quasispecies variability (124). However, due to the amplicon length limitation of 250 bp, study of the Core gene was restricted to the main epitopes, Th50-69 and B74-84 (22-25,27,28,65,66,114). The main epitopic regions of the Core gene are mainly related with immune pressure; however, the fragment also contains the tip of the spike of the Core antigen (codons 74-84), which includes codons involved in electrostatic interactions between the Core and HBsAg (125). Cryo-electron microscopy studies have shown that codons 78 and 79 are within the contact area of Core with envelope proteins (17).

Thus, we designed this study, in which UDPS was used to analyze the quasispecies composition of the Th50-69 and B74-84 regions of the HBV Core gene in the absence of antiviral treatment, to test the natural evolution of the quasispecies, whose variability was previously reported by direct Sanger sequencing (126). We also focused on analysis of one particular case after a treatment-free period and after administration of lamivudine (LVD) antiviral treatment. UDPS analysis of an HBV clonal sequence was included to evaluate the error rate of the technique, because in this type of study, use of an internal control sequence, as was done in the second study, was not possible.

4.1.2 Hypothesis

The variability and conservation of the Core gene positions described in the first study might be detected by UDPS in individual samples. UDPS analysis of the main epitopes might reflect conserved positions that could indicate possible essential roles, such as interactions between Core and surface antigens. In addition, it was hypothesized that minor variants present in the baseline quasispecies might be selected by natural evolution or under antiviral therapy.

4.1.3 Aims

1. Describe baseline variability in the main epitopic regions of the Core gene.

2. Analyze linkage of different mutations in the Core gene.

3. Evaluate the evolution of the main epitopes, Th50-69 and B74-84, of the Core gene in the same patient under different conditions: a treatment-free period (baseline samples) and after LVD breakthrough.

4.2 Summary of the Study

The baseline variability of the main epitopic regions of HBV Core gene was analyzed attending to the percentage of changes detected in the 55 codons studied (from codon 40 to 95 of the Core gene). Four baseline samples were included, 2 genotype A and 2 genotype D. One genotype A patient was HBeAg negative due to mutations in the first codon of the preCore. The two genotype D patients were both HBeAg-negative, due to a main preCore mutation in codon 28.

All patients were diagnosed with active HBV replication and were treated with LVD. All of them presented mutations conferring resistance to treatment after 18 to 24 months. Owing to their similarities in LVD non-response, they were selected for inclusion in the study of baseline variability. A sequential study was performed in Patient 4, and 3 samples were UDPS-analyzed: in addition to the baseline sample, one treatment-free sample and one VBK sample after LVD were included.

For the present study, the UDPS sequence error values were determined by parallel UDPS processing of a cDNA clone in triplicate. Results from the clone resulted in a multiparameter Poisson model, with the distribution of the error rate per site represented in arrays and according to the type of nucleotide change (Detailed description of the filter developed to establish the error rate in the clone is presented in the Appendix).

In addition, to obtain the percentage of amino acid variability in each sample, the total number of amino acid substitutions was divided by the total number of amino acids analyzed. This value, which represented the theoretical variability of each position, was used to estimate the expected variability for each epitopic and non-epitopic (remaining positions) region studied. To obtain the expected variability, the theoretical variability was multiplied by the length of the epitope: 20 for Th50-69, 11 for B74-84, and 25 for the remaining positions). For example, if 3 changes in TH50-69, 3 changes in B74-84 and 2 changes in the remaining sequences were detected, the total variability for each position would be ((3+3+2)/56)=0.14. From this value, the expected variability in TH50-69 would be 0.14x20=2.8; in B74-84, 0.14x11=1.5, and in the remaining non-epitopic positions, 0.14x25=3.5. From these results, we deduced that B74-84 variability (2 changes) is higher than was theoretically expected (1.5), suggesting a predominant evolutive pressure over this region.

A total of 108 403 sequences were obtained and analyzed. Each sample showed specific patterns of amino acid substitutions. Patient 1 (genotype A, HBeAg-negative) presented the lowest rates of variability (Figure 25a), but showed a mutated sequence in the motif located in amino acids 64 to 67. The sequence flanked by these amino acids is commonly defined by $E_{64}LMT_{67}$ and is recognized by T cells (65). The simultaneous E64D and T67N change has been reported to reduce T-cell proliferation *in vitro* (65), and interestingly, the sequence detected in Patient 1 was defined as DVTN. The other genotype A sample (Patient 2, HBeAg-positive) presented higher percentages of variability (Figure 25b), and the two main substitutions at position 41 (mutant: A, 7.51%) and 59 (mutant: I, 4.21%) coincided with the consensus

sequence of Patient 1. The variability detected in Patient 1 (0.1% average amino acid variability), ranging from 0.69% to values under the cut-off (<0.05%), was the lowest in all 4 samples. In this patient, the main epitopic regions contained 67.7% of the changes, a percentage 1.2 times higher than would be expected by the length of these regions, and the changes were equally distributed between the two epitopes. In contrast, Patient 2 had higher variability (0.35% average amino acid variability).

Figure 25 Variability detected in genotype A cases: Patient 1 (a) and Patient 2 (b).

(a)
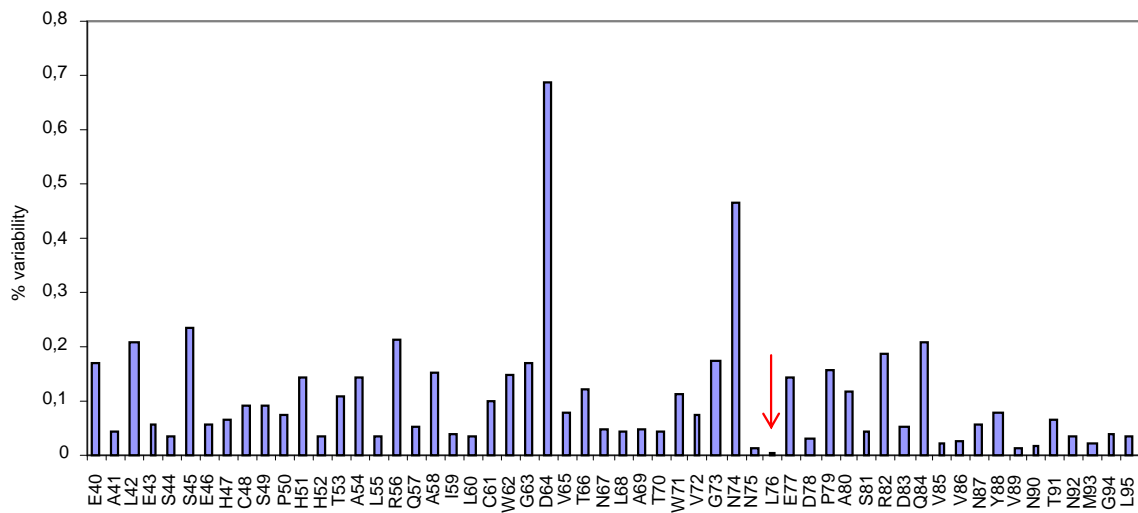
(b)
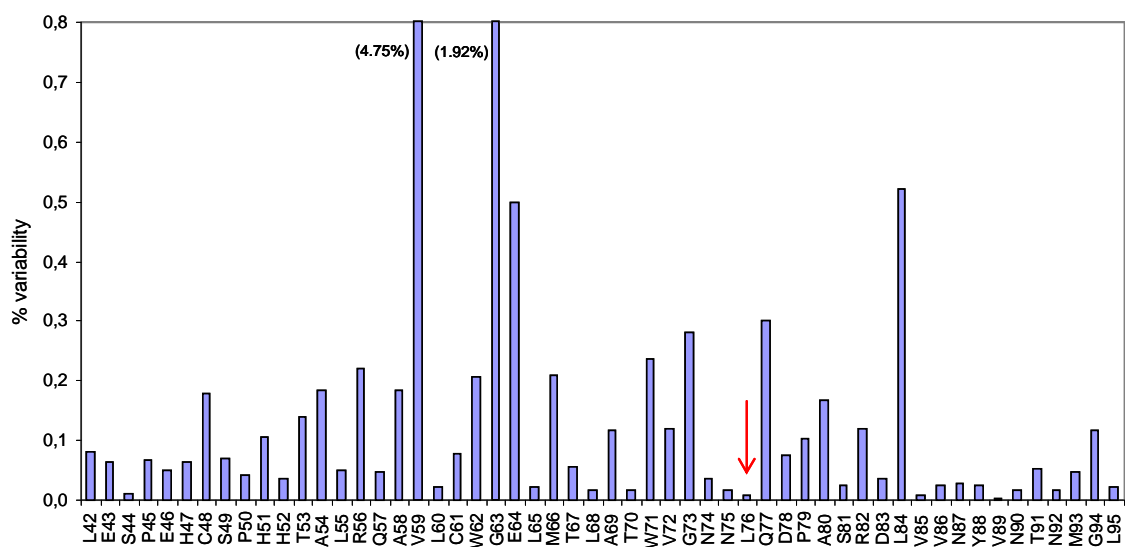
The two genotype D baseline samples (Patients 3 and 4, Figure 26) had the same consensus sequence, except in codons 64, 74, 80, and 93, which were also the most highly variable in Patient 4. In Patient 3, five codons with more than 1% variability were detected: A41 (1.58%), I59 (1.68%), A74 (2.1%), E77 (1.94%), and S87 (1.93%) (3 of them in epitopic regions). The average amino acid variability was 0.26%, and 57.8% of changes were located in the main epitopic regions. Overall, this percentage was not higher than expected; however, changes in B74-84 were 1.6 times higher than the expected random percentage (31.6% vs 19.7%). In Patient 4, variability was higher than 1% in 8 codons: Q40 (36.27%), L55 (18.58%), D64 (29.96%), V74 (13.72%), E77 (3.03%), T80 (8.08%), T92 (18.93%), and V93 (18.89%) (5 of them in epitopic regions). Despite the high total amino acid variability in Patient 4 (2.69%), only 49.8% of changes affected positions located in main epitopic regions, a value lower than was expected in both Th50-69 and B74-84.

Regarding the number of highly variable codons (variability >1%) in genotype A and D cases, it was observed that genotype D contained more variable codons than genotype A (2 codons in genotype A cases and 14 codons in genotype D).

Figure 26 Variability detected in Patient 3 (a) and Patient 4 (b), both of whom were HBeAg-negative and infected by genotype D.

a)
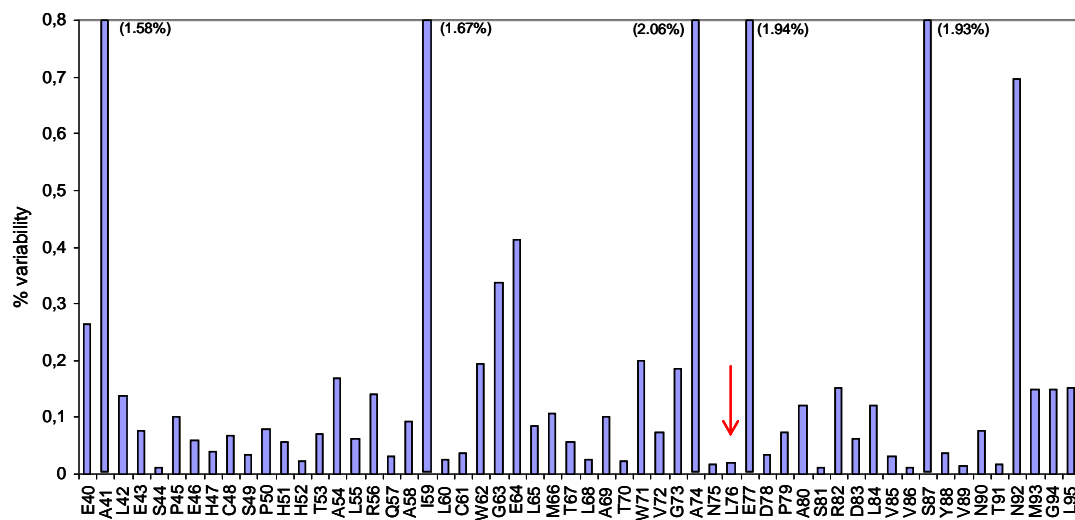
b)
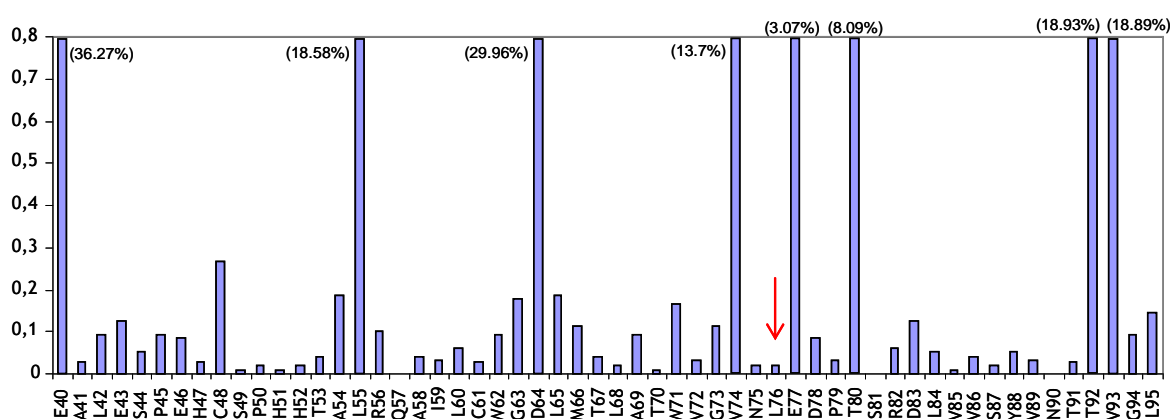
As to the median baseline variability of the 4 samples, 12 codons showed a variability lower than the system error rate (<0.05%) (positions 44, 52, 57, 68, 70, 75, 76, 78, 81, 85, 86, and 89). The most highly conserved was leucine in codon 76, with frequencies clearly below the system error rate (0.003%-0.02%) and a median baseline error of 0.013%. Under the hypothesis that high conservation could indicate an essential role of the conserved amino acid, we decided to analyze one of the most conserved positions by site-directed mutagenesis. Codon 76, coded by leucine, claimed our attention because it has never been defined as essential and it is located in the B74-84 epitope. The mutations induced in codon 76 in Huh7 cell culture were a valine (V) and a proline (P). The presence of P significantly decreased production of HBsAg, HBeAg, and HBV DNA, in comparison with the wild-type form (HBsAg=56.7 [L] vs. 28.2 [P] IU/mL; HBeAg=5.97 [L] vs. 1.19 [P] arbitrary units and HBV DNA =6.3 [L] vs. 5.8 [P] logs IU HBV-DNA/mL). However, the presence of V did not differ from the wild type in HBV DNA (both yielding 6.3 logs of IU HBV-DNA/mL), but it reduced HBsAg levels (HBsAg=56.7 [L] vs. 36 [V] IU/mL) and, surprisingly, increased HBeAg (HBeAg=5.97 [L] vs 24.6 [V] arbitrary units).

Regarding the longitudinally studied patient, 3 samples (baseline, after 36 months without treatment, and after 18 months of LVD treatment at VBK) from Patient 4 were processed. A decrease of Core variability in the treatment-free period and after LVD treatment was observed. Linkage analysis was performed of the 10 most variable positions, which were selected from the highest standard deviation values obtained from the frequency of mutations for each position in the 3 samples, defined as the most variable positions. The haplotype distribution of these 10 most variable positions was established for the 3 samples (Figure 27

and Table 9). The baseline sample presented the highest complexity of haplotype distribution, 16 haplotypes in ≥1% (the master and 15 with one or more variants) reflecting high complexity of its quasispecies (Figure 27 and Table 9a). In the treatment-free sample, which represented a period of natural evolution of the region, a minor baseline variant was selected as the master sequence (variant 12: 1.31% at baseline). Moreover, only two minor variants were detected in the treatment-free sample (Figure 27 and Table 9b), therefore, evidencing lower complexity than the baseline sample. After LVD treatment, the variant selected as master and the two minor variants detected after the treatment-free period were maintained (Figure 27 and Table 9c); thus representing less quasispecies complexity.

Figure 27 Distribution of HBV Core region haplotypic composition in the three sequentially analyzed samples (see % in Table 3).
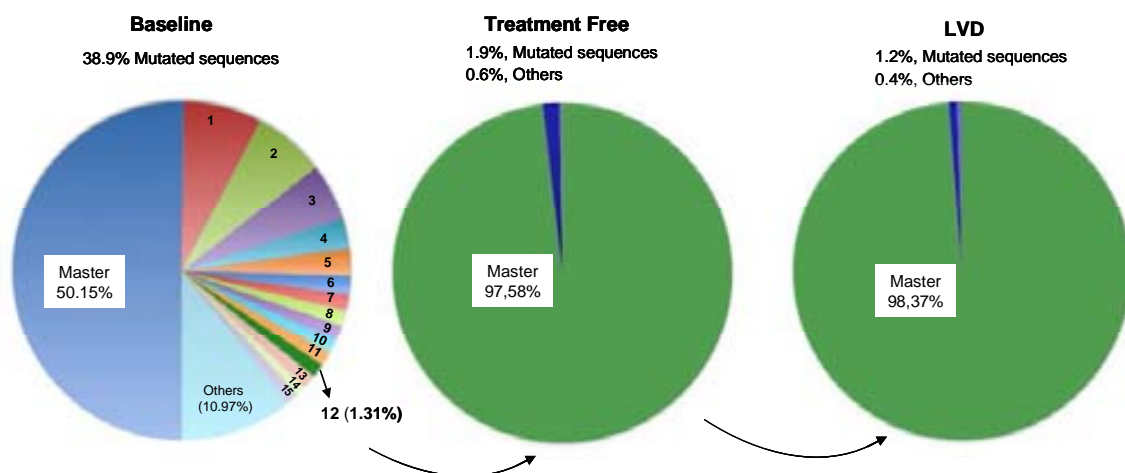
Table 9 Linkage analysis of the most highly variable codons in the sample from Patient 4.

a)

Codon	40	41	55	59	64	74	77	80	87	92	93	%	Variants
Master	E	A	L	I	D	V	E	T	S	T	V	50.2	
	Q	.	.	.	.	.	.	.	.	.	.	7.5	1
	Q	.	.	.	E	G	.	.	.	N	M	6.84	2
	Q	.	I	.	E	.	.	.	.	.	.	5.45	3
	Q	.	I	.	.	.	.	.	.	.	.	2.91	4
	Q	.	I	.	E	.	.	G	.	.	.	2.56	5
	.	.	.	.	.	.	.	.	.	N	M	1.77	6
Baseline	.	.	.	.	E	G	.	.	.	N	M	1.54	7
	Q	.	E	.	.	.	.	.	.	.	.	1.48	8
	Q	.	.	.	E	G	.	.	.	.	.	1.43	9
	Q	.	.	.	.	.	D	.	.	N	M	1.38	10
	.	.	.	.	E	.	.	.	.	.	.	1.36	11
	Q	.	I	.	E	.	.	G	.	H	M	1.31	12
	.	.	I	.	E	.	.	.	.	.	.	1.29	13
	Q	.	I	.	E	.	.	.	.	N	M	1.03	14
	.	.	.	.	.	.	.	G	.	.	.	0.99	15

b)

Codon	40	41	55	59	64	74	77	80	87	92	93	%
Master	Q	A	I	I	E	V	E	G	S	H	M	97.58
Treatment	E	S	L	V	.	N	Q	A	N	N	.	1.57
Free	.	.	.	.	K	.	.	.	.	.	.	0.28

c)

Codon	40	41	55	59	64	74	77	80	87	92	93	%
Master	Q	A	I	I	E	V	E	G	S	H	M	98.37
Lamivudine	E	S	L	V	.	N	Q	A	N	N	.	0.771
treatment	.	.	.	.	K	.	.	.	.	.	.	0.468

In conclusion, this study validates application of UDPS to study the variability of the main epitopes of the HBV Core gene and substantiates a significant richness of the baseline HBV quasispecies. The largest number of variable codons was mainly detected in the Th50-69 and B74-84 regions. A probable relationship with HBV genotype and Core variability was suggested, because higher variability was detected in genotype D samples than in genotype A samples. However, more extensive analyses in a larger number of samples must be performed to confirm this possibility. In addition, important variability was associated with well characterized Th cell motifs, such as $E_{64}LMT_{67}$, suggesting that the host immune system might be the main factor responsible for HBV Core evolution. Furthermore, UDPS showed that codons 78 and 79 were highly conserved, in keeping with their involvement in the interaction between the HBV virion capsid and envelope (65). However, codon 76 was found to be the most highly conserved, even to greater degree than codons 78 and 79. Thus, we postulate possible involvement of codon 76 in interactions with HBsAg (similar to codons 78 and 79) or an involvement in HBeAg conformation.

The dynamism of the HBV quasispecies was tested in one patient in whom strong selection of one of the minor variants present in the baseline quasispecies was seen, coinciding with a decrease in Core variability during a treatment-free period and on lamivudine treatment. This decrease seemed to be the result of a "steady state" situation of the HBV quasispecies after selection of the most highly fit variant after the treatment-free period. UDPS proved to be a useful massive cloning method to deeply analyze viral quasispecies that is mainly limited by the current high cost of this technology. However, because of the advantages of UDPS in the study of viral quasispecies, the cost will probably decrease as application of this technology increases, enabling processing of samples from a large number of patients. At the methodological level, the error rate of the complete system measured by the processing of a clonal sequence (0.05%) was similar to that obtained in the second study (0.03%), measured by an internal control sequence.

Discussion

5 Discussion

5.1 Methods for Studying HBV Variability And Quasispecies Composition

In individual infection, HBV circulates as a quasispecies, whose evolution results in selection of different variants depending on the environmental situation. In chronic hepatitis B patients, HBV infection is currently treated with antiviral therapy, and the presence of these drugs represents an important factor for HBV evolution and selection of viral variants (127). The host immune pressure developed during HBV infection can also change the quasispecies population by inducing selection of variants able to escape the immune response (107).

The evolution of HBV at different time points can be studied at different levels, depending on the percentage of the variant of interest in the viral population. For example, if a variant comprises 20% of the population, conventional Sanger sequencing can detect it. However, linkage analysis of mutants detected by conventional sequencing cannot be performed. For example, the Sanger sequence of a 50:50 mixture of the variants TTT**A**GG**G**CA and TTT**G**GG**A**CA will result in the consensus sequence TTT(**A/G**)GG(**A/G**)CATG, and it will not be clear whether G and A bases are present in the same strain. If the study were limited to a variant accounting for 5% of the population, the InnoLipa method could detect it, but the variant would have to be defined before detection and linkage analysis would not be possible in this case either.

Cloning and sequencing of viral fragments is the classic method for studying low-percentage (minor) variants in viral quasispecies, and it also allows linkage study of the mutations. However, cloning and sequencing is time consuming. For example, to detect variants in percentages of 1%, at least 100 clones must be sequenced and one of them should contain the mutation; however, it is assumed that multiple processing of single sequences is required. Despite this limitation, classic clonal methods have enabled study of the quasispecies composition of several viruses, such as HCV (128).

Since 2005, next-generation sequencing (NGS) technologies have represented an important step ahead in the study of viral quasispecies, among other possible applications, such as complete sequencing of human genome (129,130), identification of bacterial genomes, improvements in cancer research (131) and the study of human microbiomes (132,133). Nonetheless, as with all techniques, NGS has limitations, although it has been applied to study

several viral types. Specifically, the 454 technology has been used to study HIV (134,135), and HCV (136).

In the first study of the present thesis, the aim was to analyze the variability of the HBV Core region by direct sequencing (ie, variants present as more than 20% of the quasispecies), and compare the variability and selection of variants in sequentially obtained samples. To this end, we studied the Core region by conventional Sanger sequencing technology. Of course, this method presents limited sensitivity, but the cost is assumable and therefore, a large number of samples can be included. Using conventional sequencing, we were able to study the HBV Core gene in numerous samples (Results are discussed in section 5.2.2, Core gene variability).

While the first study was underway, our group had the opportunity to use 454 technology (Roche) based on ultra-deep pyrosequencing (UDPS). Among all the NGS techniques, UDPS enabled analysis of the longest DNA fragments (250 nt, when our studies were designed), and was the method that best fitted our aim: to study HBV quasispecies variability at a deeper level. Thanks to the availability of this method, the last two studies of the present thesis were based on application of UDPS. For our requirements, UDPS presents three main limitations: the cost, establishment of the error rate, and the length of the region analyzed (250 bp at the time of the studies). We tried to bypass the first limitation by careful selection of the number and the types of samples included. Regarding establishment of the error rate, we performed several control experiments in each study. Taking advantage of the design of the second study, we were able to include an internal control sequence (a restriction target sequence) within the amplicon analyzed. This was a new strategy that had not been used in previously reported UDPS studies. Conservation of this internal sequence was an absolute requirement for development of the experiment and it was as susceptible to errors as the entire amplicon sequence. Therefore after UDPS, any nucleotide difference observed in this sequence had to be considered a sequencing error. This enabled definition of the UDPS error rate and establishment of the cut-off value, which allowed us to define what percentage of variants could be considered true members of the viral quasispecies.

To define the error rate of amplicons for the third study, we included a clone in the UDPS experiments (131,134) and improved the clone error rate by developing a computational filtering step. The filter was based on random Poisson distribution of errors, adjusted by an array of values attending to the type of nucleotide substitution (A to C, A to T, A to G, C to A, C to T, etc). Interestingly, despite the fact that two different strategies were used in the second and third study, similar percentages were defined for the error rates of the technique, 0.03%

and 0.05%, respectively, which were lower than any previously reported value for UDPS analysis of viral quasispecies.

As was mentioned above, at the time of the study, the length of the amplicons for UPDS analysis was 250 bp. For this reason, the third study was limited to two main epitopic regions, Th50-69 and B74-84, representing a total of 210 bp. This length restriction to 250 bp was especially relevant in the second study because we wanted to analyze two regions located at more than 1-kb distance from each other in the same HBV genome, the preCore (ORF nt 1814-2452) and the YMDD motif of the polymerase (nt positions 736-747). We circumvented this limitation by designing a technique to obtain an HBV construct in which the two regions of interest would be close enough to be studied by UDPS. To obtain this HBV construct, the sequence of a restriction enzyme (Hind III) was added to the primers used for PCR amplification. The primers designed were located near the regions of interest, so that the amplified fragment contained the preCore at one end and the polymerase at the other. Hind III was the restriction enzyme selected because it does not cut inside the HBV genome. After digestion of the amplicon with Hind III and intramolecular ligation, both ends of the amplicon came together in a new circular construct, separated by a short 42-bp fragment containing the primer sequences (18-bp each primer) and the 6-pb Hind III sequence. Thus, in the new circular HBV construct, the preCore and the YMDD motif of the polymerase of the same viral genome were in close proximity. This molecule was the template for UDPS. The Hind III sequence (AAGCTT) was the internal control sequence to establish the error rate of the technique. To obtain the HBV circular construct, Hind III ligation of the ends was a unique, absolute requirement. In addition, ligation could not occur if the Hind III sequence contained nucleotide errors, because ligation of protuberant ends requires complementarity of the bases. For this reason, all errors detected in this 6-bp sequence were considered errors sources from the UDPS process and were eliminated.

Of particular interest, the technique designed for our second study can be applied to research investigating simultaneous mutations in distanced regions of any genome. In addition, because the circular construct included an internal sequence, the technique allows establishment of the UDPS error rate. The only requirement for building the PCR construct is that the two regions of interest and the region in between must be amplified by PCR; currently, the maximum DNA length amplified by PCR is 5 kb (137).

The error rate from the HBV clone of the third study was set at 0.05%, similar to the 0.03% obtained with the internal control sequence used in the circular preCore-Pol construct of the

second study; therefore, the error rate of UDPS sequencing of a 250-bp amplicon was established at <0.05%, regardless of the experimental approach used. With the use of UDPS technology we were able to study the variability of the HBV quasispecies in the second and third study. The results obtained in both studies clearly show that UDPS is a powerful tool for analysis of the HBV quasispecies composition.

5.2 HBV Variability and Quasispecies Composition

5.2.1 Quasispecies distribution of the preCore region

The circular construct designed in the second study enabled analysis of the distanced preCore and YMDD motif of the polymerase regions. We focused our interest on YMDD because it is the main catalytic motif of the polymerase, and the region where the main polymerase mutations resistant to LVD are clustered. LVD has been massively used as antiviral therapy, but over time, patients present emergence of high rates of resistance mutations (38). We assumed that if there was some restriction for the simultaneous presence of preCore and polymerase mutations, this limitation would be confirmed by analysis of the YMDD polymerase and preCore of the same viral genome. Nonetheless, the remainder of the polymerase was not analyzed, and it may contain other interesting regions of interest for further studies.

The high percentages of viral populations with preCore and YMDD mutations observed in LVD-resistant anti-HBe patients might indicate that mutations in these regions occur simultaneously in the same genome. However, there has been little direct study of this simultaneous presence (104,105). Lok et al. (104) examined a small number of clones from HBeAg-negative patients treated with LVD for one year and found no correlation between detection of LVD-resistant mutants and HBV genotype or presence (or not) of the preCore stop codon. Another study that analyzed the entire HBV genome (105) showed no clear relationship between preCore promoter mutations and selection of LVD-resistant polymerase mutations. It must be kept in mind that only a small number of clones were analyzed in these studies, mainly because a long fragment (>1 kb) must be cloned and sequenced, a difficult task. Cloning of preCore-YMDD fragments from HBeAg-positive patients would be even more difficult because any preCore mutants present would be in small populations and a large number of cloned sequences would be required to detect them.

The novel PCR-based technique designed in the second study supports the idea that simultaneous mutations in the preCore region and YMDD Pol motif are common in the HBV quasispecies in CHB patients, regardless of the viral genotype (A or D) or HBeAg status. This hypothesis was supported by the simultaneous presence of preCore and YMDD mutations in the two HBeAg-positive cases after LVD VBK, although they were detected in low proportions (1.2% in patient 4 and 0.28% in patient 6). These variants would be extremely difficult to detect by classic clonal and Sanger sequencing, where more than 200 clones longer than 1 kb would be needed. In agreement with previous studies (44), the two HBeAg-positive cases analyzed in the second study showed highly significant percentages of preCore HBeAg-negative mutations (8.9% and 11.2%), which suggests possible selection of preCore mutations when the host immune response increases during viral infection. Likewise, in HBeAg-negative patients, wild-type preCore genomes were detected in lower, but significant percentages (0.04%-0.29%). These variants could represent a reservoir in the HBeAg-positive cases prone to be re-selected (seroreversion) without treatment or during LVD therapy, if HBeAg-positive variants were more resistant to antiviral therapy than HBeAg-negative ones. In fact, this seroreversion phenomenon has been frequently observed (106,107). In addition, *in vitro* studies have shown that progeny DNA levels are restored in constructs with preCore and LVD mutations (138,139). In relation to the hypothesis that HBeAg-negative variants are more sensitive to antiviral therapy than preCore wild-type (HBeAg positive) sequences (106,138,139), an HBeAg-positive patient from the second study (patient 4) was longitudinally followed up, and the proportion of HBeAg-negative variants strongly decreased after LVD from 8.9% to 1.22%. This observation might support the hypothesis and would also explain the common observation of reversion to wild type in codon 28 during LVD treatment (106,140). The presence of these minor genomes in the HBV quasispecies of chronic hepatitis B patients reflects the complexity of the HBV population during infection.

HBV pgRNA has redundant ends, and the "ε" structural elements (coded by the preCore sequence) are present twice in pgRNA, one near the 5' terminus and the other at the 3' terminus. The key event for HBV genome replication is binding of the viral polymerase to the ε stem loop of the encapsidation signal located at the 5' end of pgRNA (12,48,55). The thermodynamic stability of this signal is reported to be essential for its function; in this sense, correct base-pairing between positions 1896 (codon 28) and 1858 (codon 15) at the lower stem of the "ε" encapsidation signal is one of the main constraints for preCore variability and ε thermodynamic stability, as has been reported (12). The results obtained in the second study highly reinforced the importance of base pairing between nucleotides constituting the

encapsidation signal. We observed highly conserved positions and maintenance of correct base-pairing in the upper stem. A change of T to C in positions 1884, 1885, and 1893 was found in 0.15%, 0.14%, and 0.13% respectively, and interestingly, the complementary change of A to G in their base-paired positions (1875, 1874, and 1867) showed very similar percentages (0.13%, 0.1%, and 0.13%, respectively). Additionally, the low percentage (0.1%) of G to A changes detected in position 1891 (which represents a new stop codon in the preCore sequence) was also compensated by complementary A to G changes in a similar percentage (0.09%) in the 1866 base-paired position. Observation of correct base-pairing of the apical and upper stem of ε might reinforce the idea that structural rather than sequence constrictions are involved in conservation of the ε encapsidation signal. Regarding HBV genotype, the main preCore mutation (A1896G variant) was clearly predominant in genotype D, while in genotype A, significant percentages of preCore mutation were located at codon 1 (observed in patients 2 and 6). This thermodynamic restriction between HBV genotypes corroborates results from previous studies using clonal or direct sequencing techniques (12,55).

It has been suggested that HBV polymerase recognizes the ε structure through a direct interaction with eukaryotic translation initiation factor (eIF4E) (141). After initial ε-Pol binding, the apical stem loop of ε should be opened to form a primer synthesis-competent complex, but the highly stable HBV upper stem represents a barrier for this process. The free energy to overcome this barrier may come from interaction with RT and by capping the TGT apical pseudo-triloop with capsid proteins (Figure 10) (142). The hexanucleotide sequence located at the top of the apical stem (from 1877 to 1882) was found to be highly conserved among 1200 HBV strains (54). Magnetic Resonance analysis (143) has shown that the $_{1878}TGT_{1880}$ sequence folds in a pseudo-triloop, a structure correctly predicted by the RNA folding program used in the second study (144,145). The pseudo-triloop is not needed for RT binding, but is required for pgRNA encapsidation, suggesting its possible interaction with capsid proteins (54,146). High conservation of T_{1878}, G_{1879}, and T_{1880} nucleotides was observed in our study, particularly G_{1879} (extremely low variability, at the limit of the UDPS error rate), which suggests that they have a predominant role in the ε-capsid interaction (147).

In duck HBV, no specific nucleotide sequence seems to be necessary for RT binding, and base-pairing stabilizing the apical stem abrogates RT binding. In contrast, thermodynamic studies have shown that the HBV apical stem is more stable than duck HBV (146). It has been postulated that ε-RT binding occurs by fitting the kinked shape of the apical stem into the palm domain of RT, involving U_{1889} (146). However, in the second study, a small but significant

percentage of changes were observed in position 1889 (0.2%). Interestingly, this change maintains the kinked shape of the apical stem and the thermodynamic stability of the epsilon structure. Dynamic Magnetic Resonance experiments should be performed to test C_{1889} functionality.

Guarnieri et al (148) reported that G1899A does not significantly increase *in vitro* viral replication, despite a significant theoretical stabilization of the ε structure. In our second study, the G1899A plus G1896A mutation combination was frequently detected. It has been postulated that the G1899A change might be associated with its juxtaposition to the Core initiation codon in the Kozac sequence, resulting in an increase in HBcAg expression (148). Interestingly, in our results, T1855 was highly conserved, whereas its paired position, 1899 was highly variable. The reason why G1899A is a highly prevalent change, in contrast to the conserved T1855C, remains unknown. This finding is even more surprising taking into account that a T to C change in 1855 did not result in any amino acid substitution and thermodynamically stabilized the ε, while the change in its paired 1899 position yielded an amino acid substitution (Gly to Asp). In addition, the change of $_{1855}T:G_{1899}$ to $_{1855}T:A_{1899}$ was 100 times more frequent than the $_{1855}T:G_{1899}$ to $_{1855}C:G_{1899}$ change. Site-directed mutagenesis in position 1855 and the transfection experiments performed in the second study showed that the T1855C mutation in both the 5' and 3' ends led to a slight, but systematic, reduction in HBsAg, HBeAg, and HBV-DNA levels, supporting a possible essential nature of this nucleotide. Nonetheless, the single study analyzing the T1855C change reported to date detected a replication rate similar to that of the wild-type (148). The most striking results of our second study were that chimerical (but not natural) constructs that only contained one 1855C mutation at the 5' or 3' end and evaluated in cell culture were associated with higher HBsAg, HBeAg, and HBV-DNA levels than wild-type clones. Although the 5' and 3' structures were found to have differing functions in the singular HBV replication cycle, enhancement of viral replication activity in chimerical constructs with a single mutation remains unexplained.

In the second study, analysis of nucleotide variability in the 4-nucleotide bulge primer (4NT primer) used as the template for synthesis of the minus DNA strand and its corresponding acceptor site (AS) in the DR1 region confirmed a high conservation of both regions. It has been speculated that the 4NT primer is more essential than DR1 AS. The existence of alternative acceptor sites has been suggested (149), as well as the idea that 3-nucleotide sequences would be completely functional for minus strand DNA synthesis (148). Our results support these concepts because the 4NT bulge primer was slightly more conserved than AS DR1 (99.5% vs

98.9%). Abraham et al. (149) reported that sequences surrounding the normal AS would play crucial roles in accurately placing the nascent minus DNA strand. As was expected, the ε structural motifs were strongly conserved in our study; very few cases (1.47%) had different sequences in the 4NT primer and its canonical DR1 AS. However, 44% showed complete 4-nucleotide homology with the putative acceptor site, AS3 (position 1849-1852), whereas the remaining 56% showed only 3-nucleotide homology with canonical DR1 AS or with sequences located in previously reported alternative acceptor sites . Therefore, our results seem to agree with the existence of an alternative to the DR1 acceptor sites for the 4NT primer (149), mainly located between positions 1849 and 1852. Analysis of the theoretical secondary structure of the 3' end ε signal showed that the AS3 sequence has tighter Watson Crick pairing than canonical DR1 AS. This fact could represent an energy barrier for 4NT annealing, and explain the low prevalence of mutated bulge variants to enable correct annealing of 4NT with AS3.

5.2.2 Core gene variability

Study of the Core gene is of interest because (in contrast to the polymerase and surface) it shows little overlapping with other HBV genes; this makes it the most useful HBV region for analyzing the quasispecies structure (115). The variability of this region is also interesting (variants present in percentages higher than 20%), because of its relation with the immune response. As was described in the Introduction, the nucleocapsid is the most immunogenic of all the HBV antigens (20) and because of this and the above characteristics; we focused on the variability of this region. In the first study, we analyzed the Core gene by conventional sequencing technologies, and we compared the variability detected in sequential samples.

An increase in the number of AA substitutions in HBcAg during chronic HBV disease has been described, mainly clustering in the CD4+, CD8+, and B cell immunodominant epitopes (66,119). Although these variants are associated with persistence of chronic infection and sensitivity to IFN therapy, their real significance remains controversial (22-24,66,107,119,150), and their distribution does not show specific patterns (66,151). Our first study investigated the natural occurrence and selection of HBV Core variants in treatment-free CHB patients and those under treatment (IFN, LVD, or ADV) as the external evolutionary factor. The effect of IFN has been described previously (119,152), but when our study was performed, there were no reports investigating Core variants in patients receiving nucleos(t)ide analogs. We examined HBV Core substitutions while active viral replication was occurring. Hence, possible associations of Core nucleotide changes with response to antiviral therapy were not determined. The study of Core

variability focused on 3 epitopic regions, immunodominant Th50-69 and B74-84, and minor Th28-47, as well as the region from AA1-11, described as conserved (122). The Th18-27 epitope (153,154), which is also located in the 760-bp preCore/Core region analyzed, was not investigated because of the small number of changes detected (only 7 cases, data not shown).

5.2.2.1 Differing host immune response between chronic hepatitis B patients and inactive carriers

One part of the first study included analysis of baseline and treatment-free samples from CHB patients, and two sequential samples from inactive carriers. The baseline samples corresponded to the first available sample from each patient following diagnosis of infection. In a comparison with the corresponding master sequences of each genotype, several changes were detected. A large number of changes were found in the baseline samples, suggesting an intrinsic variability in samples from the same genotype. This variability was supposed to be putative changes occurring in the natural evolution of HBV genomes. The changes were particularly found in the main epitopic regions, Th50-69 and B74-84, coinciding with their role in immune system stimulation (22,24-26,114). It is important to mention that although the AA1-11 region has been described as a conserved region (Chain), some positions showed remarkable variability in the present study, particularly codons 3 and 4. The fact that region AA1-11 partially overlaps the Th CD4+ epitope 1-25 (26,114) suggested that there may be some immunological pressure against this epitope, which we mainly observed after ADV treatment (Discussion in section 5.2.2.2, Effect of antiviral treatments in Core gene). Regarding HBeAg status, it was observed that HBeAg-negative patients presented higher variability in Th50-69 than HBeAg-positive ones, indicating a possible relationship with HBeAg loss and Core variability (detailed discussion in section 5.2.2.3, Core gene variability and possible relationship with the HBeAg expression). We also found that patients with a larger number of changes in the B74-84 epitope had higher HBV DNA levels. This result coincides with the findings of other authors (155,156), but it contrasts with Sendi's study in which AA substitutions were found more often in patients with low HBV DNA replication and even in inactive carriers (154). Based on our results, we postulated that the presence of AA changes enables HBV to evade immune clearance, thereby lengthening the life of infected hepatocytes and increasing the viral population. This mechanism could result in a higher probability of infecting additional hepatocytes and ultimately, raising HBV DNA in serum.

The natural evolution of the HBV Core gene was evaluated in a group of samples from treatment-free (TF) CHB patients. Higher conservation in the AA1-11 region was observed in all TF samples. In addition, the highest variability was located in the main epitopic regions, supporting host immune pressure on AA regions 50-69 and 74-84, defined as the main epitopes.

In the first study, we also established a comparison between CHB patients and inactive carriers (ICs). In the Mediterranean area, ICs are characterized by low rates of viral replication; therefore, amplification of the Core gene was limited to the viral loads. Although 20 ICs were initially selected for the first study, only four could ultimately be included. The low number of IC patients included limited further analysis and definite interpretations. However, the differences observed between the TF samples from CHB patients and the IC results were used to suggest potential alternative mechanisms. The AA substitutions detected in epitopic regions in both IC and TF samples indicated persistent immune activity and suggested viral escape mechanisms (65,119), but the host response seemed to differ between these groups: the percentage of changes in AA1-11, Th28-47 and B74-84 in ICs was higher than in the TF samples. This larger number of changes in epitopes in ICs has been described in another study (154). This observation led us to speculate that possible alternative mechanisms of variability in the Core gene might develop in IC patients, as compared to untreated CHB patients.

Despite the small number of HBsAg-positive inactive carriers included in the first study, strong selection of the E40D/Q change was found in most of the patients, at levels similar to those observed at baseline in a recent study (154), suggesting differential immunological activity in the development of inactive HBsAg-positive carrier status and CHB status: immune activity specifically addressed to minor Th epitopes in cases of inactive HBsAg-positive carriers and activity addressed to changes in immunodominant epitopes in CHB. However, this observation must be viewed with caution because of the small number of inactive HBsAg-positive carriers studied.

5.2.2.2 Effect of antiviral treatments in Core gene

Regarding the effect of antiviral therapy on Core gene variability, we observed a tendency to conserve the AA1-11 region under treatment, although the variability was slightly higher than in treatment- free or baseline samples. The highest percentage of changes in Th50-69 and B74-84 was observed under IFN treatment, in keeping with its immune stimulating activity

(119,157). Core gene variability under IFN has been described (119). In our first study, however, a patient-by-patient analysis of a treatment-free sample and IFN sample was performed and the effect of IFN on the main immunodominant epitopes was confirmed. One important aspect to mention from our study was that most IFN-treated patients were anti-HBeAg-positive, and seroconversion of HBeAg has been described to increase Core variability (107). However, we did not find differences between HBeAg-positive and -negative patients.

Attending to the effect of NUCs, we expected no effect or a low effect on Core gene variability, as no immune-stimulating activity has been described. Thus, we expected to detect variability similar to that observed in the treatment-free samples. Surprisingly, the minor Th28-47 epitope was higher after NUC treatment than was seen in the untreated samples, suggesting that this minor epitope might be stimulated under the effect of NUCs. To explain this possible immunological stimulation, it should be remembered that NUCs inhibit reverse transcription of pregenomic HBV RNA, but not messenger RNA expression. Therefore, viral protein synthesis remains active and this could result in an excess of viral antigens such as HBcAg, which do not form viral particles. This intracellular accumulation of HBcAg could stimulate proteasome production of viral protein fragments, increasing antigen expression on the surface of infected hepatocytes (the greater the expression of viral epitopes, the greater the probability that HLA will bind to them). This sequence of events could potentially represent another type of immune system stimulation, as an alternative to IFN-induced HLA stimulation. We suggest that changes in Th28-47 of the Core region are selected during NUC treatment and could confer an escape mechanism for immune pressure. However, the reason why this theoretical phenomenon was more evident in the Th28-47 epitope than in the main epitopes remains unexplained.

After IFN therapy, the most frequent AA substitutions were located in the immunodominant regions, whereas after ADV, they were similarly distributed between the immunodominant regions and minor Th28-47 region. In addition, specific changes in the minor Th28-47 epitope (codons 41, 45, and 46) were seen more often after ADV therapy than in the treatment-free period. Both these findings suggest differential immune stimulation during ADV, more specifically associated with this minor epitope. Despite the fact that the substitution patterns detected were nonspecific, some changes were especially interesting. The E64D substitution in the Th50-69 region was notable, being observed in all the periods analyzed except after ADV. This change significantly reduces T cell proliferation *in vitro* when linked to T67N, altering the $E_{64}LMT_{67}$ sequence, a common motif in T cell epitopes (65); thus, it is selected by the effect of

immune pressure. In our patients, the double E64D-N67T change was observed in 46% of patients with an E64D substitution.

The main variants observed in the B74-84 region over the entire study were N74A/D and E77Q, which were similarly distributed between the baseline and sequential groups, indicating that their selection was not the result of specific evolutionary pressures. Both codons are highly polymorphic (122), and their variants are often detected in CHB and IC cases (22,26,114,119,152,158). However, changes in codon 77 have been associated with the HLA-B*4001 allele (159) and may be involved in viral escape mechanisms. Large studies are needed to determine the effect of these variants on immunologic reactivity against HBV. Interestingly, the selection of changes in codon 84 (mainly L84Q) was associated with IFN therapy. Although highly polymorphic positions, such as 74, 77, and 84 (122) have been observed in IFN therapy (119,152), there are no studies investigating their status after other therapies or in the absence of treatment.

5.2.2.3 Core gene variability and possible relationship with HBeAg expression (main preCore mutation)

The samples analyzed in the first study by conventional sequencing technologies showed a large number of changes in the Th50-69 region at baseline in HBeAg-negative patients. This may indicate an enhanced immune response during HBeAg loss. It is known that HBeAg can contribute to the outcome and pathogenesis of HBV infection, inducing tolerance in Core-specific T cells and reducing their potential to kill HBV-infected cells (10). In addition, an increased evolutionary rate has been described during the change from HBeAg-positive to HBeAg-negative status (107). However, conventional sequencing of the Core gene has not shown a specific a relationship with HBV genotype (160).

Surprisingly, after LVD therapy, numbers of AA substitutions in the immunodominant Th50-69 and B74-84 regions were higher in wild-type cases than in preCore mutant cases. This might be due to preferential immune selection of preCore wild-type sequences because they have a higher number of Core substitutions than sequences with preCore variants. In the sequentially studied patient from the second study, we also observed a decrease in the viral population containing preCore mutations. These findings from both studies could explain the reported cases of reselection of wild-type sequences after LVD therapy (161). We postulate that preCore mutants might be more sensitive to LVD treatment possibly associated with the fact

that wild type preCore genomes accumulate more changes in the Core gene under LVD and this might confer an alternative immune escape (as deduced in the first study).

5.2.2.4 Quasispecies variability of the main epitopic regions

In the third study, baseline variability of the main epitopic regions of the Core gene was also examined by UDPS, and indirectly, we were able to evaluate the effect of host immune pressure on the composition of the HBV quasispecies in 4 different CHB patients. An important richness in variability of the quasispecies composition in the main HBV Core epitopic regions was observed (Th50-69 and B74-84), and a possible relationship between quasispecies variability and HBV genotype.

Four baseline samples were analyzed in the third study; two genotype A and two genotype D cases. Due to the high cost of UDPS technology and the need to include a clone in the analysis, we were only able to examine four baseline samples and two additional ones from an LVD-treated patient. In the baseline study, the most variable codons (median baseline variability ≥1%) were 40, 41, 55, 59, 64, 74, 77, 80, 92, and 93, all of which have been previously described in an analysis of acute exacerbations in HBeAg-negative patients by classic clonal methods (162). Furthermore, some of these highly variable codons (codon 64, 74, and 77) were also detected as common changes in untreated chronic hepatitis B patients included in the first study. The baseline variability of the 4 samples suggested differences between genotypes A and D. Genotype A samples had fewer codons with variability >1%, and consequently, lower average variability than genotype D samples. Furthermore the two genotype D cases presented some highly variable codons located outside the main epitopes, indicating that the immune response against the minor epitopes, Th28-47 and Th 82-101, might be higher than in genotype A cases. The UDPS results on quasispecies variability of the main epitopes indicates possible differences against the host immune response according to HBV genotype.

HBV quasispecies evolution was evaluated in a single longitudinal case by analysis of a baseline sample and 2 additional samples, one taken after a period without treatment and one taken after LVD treatment. The highest variability, which was detected in the baseline sample, dropped during the treatment-free period, suggesting a reduction in HBV quasispecies composition due to immune pressure during the period without treatment, and coinciding with an exacerbation (162). This dynamism in the composition of the viral quasispecies was

also observed by linkage analysis, which detected a dramatic reduction in the number of viral strains and their percentage within the total variants in the treatment-free and LVD samples (less than 2% of the total sequences in both samples). Interestingly, one of the minor baseline viral strains (variant 12, 1.31% at baseline) was strongly selected and became the master sequence after the treatment-free and LVD periods. Hence, this variant can be labeled as an escape mutant whose selection would be related to better fitness, regardless of its initial frequency.

Based on these results, we postulated that the HBV quasispecies had achieved a kind of "steady state" after the treatment-free period that did not change with LVD treatment, suggesting that immune pressure might have decreased during treatment. The significant differences in average variability in the two periods suggest that the quasispecies equilibrium is dynamic. The structure of the HBV quasispecies in the 3 samples represents a complex reservoir of different minor variants resulting from natural HBV evolution and likely affected by antiviral treatment (63).

5.2.3 Peculiarities of the HBV quasispecies

5.2.3.1 The ELMT motif

The ELMT motif delimited in codons 64 to 67 of the Core gene showed considerable variability in samples analyzed in the first and third study. ELMT is commonly recognized by T-cells and the double mutation E64D and T67N, has been described to reduce T-cell proliferation *in vitro* (65).

The E64D substitution was commonly detected in samples analyzed by Sanger sequencing in the first study, except in samples under ADV treatment. The E64D and T67N double mutation was observed in 46% of patients with E64D, indicating a possible advantage of this double mutation under treatment or in treatment-free periods.

This motif also claimed our attention in the study of quasispecies composition of the main Core epitopes (third study) because the ELMT region was included in the UDPS fragment analyzed. Patient 1 of this study presented a sequence in codon 64 to 67 with complete changes from ELMT to DVTN. In addition, this patient presented the lowest rates of variability in the UDPS region analyzed. This leads us to speculate that mutations in this motif might confer an immune escape mechanism (65), and therefore, additional changes in the

quasispecies need not be selected. DVTN might suffice for immune escape and lead to the temporary "steady state" situation, until further changes in immune system activation against the new major variant occur and produce a new increase in quasispecies complexity.

Indeed, in the case report from the third study, the ELMT motif also presented peculiarities, specifically in position 64. The escape variant detected in the baseline sample in a low percentage (1.31%) and later selected as the major variant in the treatment-free and lamivudine samples, had the E64D mutation. This could indicate that the mutation (in combination with other Core gene mutations) was selected as an immune escape variant.

The importance of the ELMT motif was made patent in the work analyzing the Core gene (first and third studies) and we speculated that it might be involved in immune escape; in particular, position 64 might have an important role in evading the host immune response.

5.2.3.2 Codon 76 in the Core gene

The HBV Core sequence is involved in capsid conformation by interacting with the surface antigens in viral particle assembly. However, the interaction between HBsAg and HBcAg in virions is still unclear (18,163,164). The Core region analyzed by UDPS in the third study is part of the assembly domain (amino acids 1-149); thus, the changes in the region analyzed could potentially modify the shell conformation. Electrostatic interactions between Core and HBsAg take place at the tip of the spike of the Core antigen (codons 74-84) coinciding with the location of the main Th epitope and representing a constriction factor for the variability of this epitope (16,125). Cryo-electron microscopy studies have shown that codons 78 and 79 are within the contact area of Core with envelope proteins (17). This essential role might explain the high level of conservation of these positions (especially codon 78, 0.05%) observed in the third study. However, surprisingly, the most conserved codon of the main Core gene epitopes on UDPS analysis of baseline and sequential samples was leucine at position 76. Interestingly, leucine 76 is also located at the tip of the spike, next to the major immunodominant region (MIR) and is a part of the B74-84 main epitope. This position has never been described as essential, in contrast to the nearby positions 78 and 79 (17).

To our knowledge, few studies (163,164) have analyzed the effect of single Core amino acid mutations on HBV replication *in vitro*. Only one such study conducted by Ponsel and Bruss (164) evaluated the leucine 76 (L76) position (among others) by inducing a change to alanine; no significant reduction in nucleocapsid or virion production was observed. In the present

study, we induced changes in the hydrophobic characteristics of position 76 and determined the effect on HBsAg, HBeAg, and HBV DNA production. In contrast to the results of Ponsel and Bruss, we found a significant reduction in HBsAg production with both the L76V and L76P changes, suggesting possible involvement of L76 of HBcAg in the HBsAg interaction. A significant decrease in HBV replication in the presence of P76 was detected, leading us to speculate that the hydrophobic characteristics of position 76, conferred by the presence of L or V, are needed for HBV replication. We observed a surprising increase in HBeAg production in the presence of the L76V mutation and no HBV DNA increase. This was a notable finding, particularly because some authors have reported a correlation between HBeAg and HBV-DNA levels (165). Nonetheless, this correlation was not found by other authors (166) and currently remains controversial. We suggest that this unexpected HBeAg increase may indicate alternative pathways between HBeAg and HBV replication, as has been indicated previously (167). Based on our *in vitro* results, we postulate that the amino acid changes induced in the Core sequence are not as important as the structure adopted by the capsid and HBeAg.

5.2.3.3 Core defective genomes

The construct analyzed in the second study included UDPS analysis of the first codon of the Core gene. Significant percentages (clearly higher than the error rate) of genomes with mutations in the ATG start codon of the Core gene were detected in all samples. The presence of genomes without the Core antigen start codon suggests the existence of minor viral populations that are defective for viral capsid production, and therefore defective for producing viral progeny. The presence of these defective genomes in HBV infection may be explained by possible collaboration between viruses from the same infection by a trans-complementation mechanism: If wild-type and Core-defective genomes coinfect the same liver cells, Core proteins produced by the wild-type viral genome might be used to encapsidate Core-defective HBV genomes. It must be taken into account that HBsAg is produced in excess, whereas HBcAg is detected in limited proportions (8). Therefore, the relative "consumption" of HBcAg by encapsidation of these defective genomes could represent a possible reduction of the HBV Core protein pool in the cytoplasm and a decrease in Core antigens expression in the cell membrane, thus modulating host immune pressure and maintaining the infection.

Therefore, the presence of defective genomes might establish cooperation with wild-type viruses and favor viral fitness. This mechanism of trans-complementation has been described in *in vitro* studies analyzing defective HBV Core clones (148).

5.2.3.4 X gene deletions

The construct analyzed in the second study included analysis of the last 8 codons of the X gene. We detected low percentages of sequences with mutations in the stop codon of the X gene, which might cause an elongation of HbxAg by 51 amino acids.

The HBx gene plays a crucial role in HCC pathogenesis by interacting with cellular oncogenes, although the carcinogenic mechanism currently remains unclear (39). HCC occurs in only a small percentage of HBV-infected patients by an unknown event that triggers carcinogenic cascades, such as COOH-terminal truncation of HBx during HBV integration into the host genome (168). Ma et al. (169) proposed that the HBx gene contains two functional domains: an oncogenic domain (in the N terminal through the middle peptide) and a proapoptotic domain (in the COOH-terminal peptide). In HBV-infected hepatocytes, there is a balance between these two functions, but when the proapoptotic domain is deleted, the balance is broken and the oncogenic function becomes dominant, leading to HCC. According to this hypothesis, the low percentage (0.17%) of sequences with an additional amino acid in the HBx COOH-terminal peptide due to the mutation in the HBx stop codon might be associated with the risk of developing HCC during HBV infection. A possible role of these minor HBx mutated variants in modulating viral replication can also be postulated.

Conclusions

6 Conclusions

From the present thesis, and attending to the aims defined in the three studies, the following is concluded:

1. Ultra-deep pyrosequencing is a useful tool to study the composition of the HBV viral quasispecies. However, a control to establish the error rate of the technique is needed (an internal control sequence or a clone).

2. The simultaneous presence of mutations in the YMDD motif of the polymerase and preCore region of the same HBV viral genome is not restricted, regardless of HBeAg expression or the HBV genotype.

3. The variability of the preCore region is highly restricted, mainly because it contains the epsilon encapsidation signal sequence. This restriction in variability is more limited to the structural conformation of this signal than the sequence itself.

4. Selection of changes in the Core gene is mainly due to the immunomodulator effect of interferon administration; however, the host immune response also induces evolutive pressure. The variability induced by these two factors is mainly clustered in the main Th50-69 and B74-84 epitopes.

5. Despite the fact that NUCs have no immunomodulator effect, administration of Lamivudine or adefovir induces changes in the minor Th28-47 epitope.

6. Inactive HBV carriers (ICs) present alternative mechanisms of variability in the Core gene relative to that seen in untreated chronic hepatitis B (CHB) patients. Immune pressure in the CHB group is mainly directed against the main epitopes Th50-69 and B74-84, whereas in the IC group, the immune system may act against minor epitopes, such as Th28-47.

7. HBV variants within preCore mutants seems to be more sensitive to Lamivudine.

8. Baseline variants in the Core epitopes, such as mutants of the $Th_{64}ELMT_{67}$ motif, which are able to evade the host immune system, are selected to maintain HBV infection in the absence of antiviral treatment.

9. Ultra-deep sequencing technology enables study of highly conserved positions in the HBV quasispecies to investigate possible essential roles. However, highly sensitive sequencing methods have shown that the virus can select variability in these potentially essential positions or motifs.

   a. Correct annealing between the 4 NT primer of the bulge and the acceptor site does not seem to be an absolute requirement for synthesis of the negative DNA strand.

   b. The presences of defective mutants (Core mutants in the initial codon or X antigen elongation) in a low percentage in the quasispecies strongly suggest cooperation between members of the HBV quasispecies.

10. Ultra-deep analysis of preCore/Core regions of HBV has revealed the high complexity of the viral quasispecies.

Appendix

7 Appendix

7.1 Complete manuscript of the Third Study

1 Variability and Dynamics of Hepatitis B Virus Quasispecies Reflected By Ultra-Deep

2 Pyrosequencing of The Main Epitopic Regions of The HBV Core Gene

3 Maria Homs, Maria Buti, David Tabernero, Josep Quer, Rosendo Jardi, Israel Ortega, Álex

4 Sanchez, Melanie Schaper, Noelia Corral, Rafael Esteban, Francisco Rodriguez-Frias.

5

6 Running title: Ultra-Deep analysis of HBV-Core main epitopes

7

8 Institutions: Maria Homs, Maria Buti, David Tabernero, Josep Quer, Melanie Schaper, Rafael

9 Esteban and Francisco Rodrigue-Frias, Centro de Investigación Biomédica en Red de

10 Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto Carlos III, Barcelona 08036, Spain.

11 Maria Homs, David Tabernero, Rosend Jardi, Melanie Schaper, Noelia Corral and Francisco

12 Rodríguez-Frias, Department of Biochemistry, Hospital Vall d'Hebron, Universitat Autònoma de

13 Barcelona, Barcelona 08035, Spain.

14 Maria Buti and Rafael Esteban, Department of Hepatology, Hospital Vall d'Hebron, Universitat

15 Autònoma de Barcelona, Barcelona 08035, Spain.

16 Israel Ortega and Alex Sanchez, Statistics and Bioinformatics Unit, Research Institut, Hospital

17 Vall d'Hebron, Barcelona 08035, Spain.

18

30 **Corresponding author**: Francisco Rodriguez-Frias, PhD. Proteins, Hepatitis and Molecular

31 Genetics Unit, Department of Biochemistry, Hospital Vall d'Hebron. Passeig Vall d'Hebron, 119-

32 129, Barcelona, 08035, Spain. frarodri@vhebron.net

33

34 **Telephone** +34 932746991 and **Fax** +34 932746831

35

2

36 **ABSTRACT**

37 In this study, the variability of the main immunodominant motifs of hepatitis B virus (HBV) Core

38 gene are evaluated by ultra-deep pyrosequencing (UDPS). Four samples corresponding to 4

39 naive patients were sequenced to assess baseline variability. Two additional samples were

40 selected to sequentially analyze a lamivudine-treated patient. After computer filtering of UDPS

41 data, 108 403 baseline sequences were analyzed. Positions with highest variability rates were

42 mainly located in the main Core epitopes, showed some relationship with HBV genotype, and

43 were particularly associated with the T-helper motif, $E_{64}LMT_{67}$. Thus, immune system pressure

44 is suggested to be the main cause of HBV Core evolution. Conservation of essential codons 78

45 and 79 for Core-HBsAg interaction was confirmed. However, codon 76 was also highly

46 conserved, perhaps indicating a role in HBsAg interactions, similar to codons 78 and 79, or

47 even in HBeAg conformation. The dynamism of the HBV quasispecies tested in one patient

48 showed strong selection of one minor baseline variant, coinciding with a decrease in Core

49 variability during a treatment-free and lamivudine-treated period. The drop in variability seemed

50 to result from a "steady state" situation of the HBV quasispecies after selection of the variant

51 with greatest fitness. UDPS analysis is an optimal tool for determining the composition of viral

52 quasispecies.

53

54 **KEYWORDS**

56 **INTRODUCTION**

57 Hepatitis B virus (HBV) infection is a global health problem. Around 350 million people are

58 chronically affected with this pathogen, which confers a higher risk of developing liver disease,

59 liver cirrhosis, and hepatocellular carcinoma. The course of HBV infection is closely related to

60 the host immune response and genetic factors [1], and disease progression is related to

61 mutations in the HBV Core gene [2-4].

62 HBV Core gene codes for two partially collinear proteins, the hepatitis B "e" antigen (HBeAg)

63 and hepatitis B Core antigen (HBcAg). These proteins, together with the surface antigen

64 (HBsAg) are important targets for antiviral immunity, but HBcAg seems to be the most

65 immunogenic [5]. Several epitopes have been identified in the HBV Core gene. Among them, two

66 regions play a particularly important immunodominant role: the sequence from amino acid 50 to

67 69, which immunostimulates CD4+ T-helper lymphocytes (Th50-69) [6] and the sequence from

68 amino acid 74 to 84, which stimulates B lymphocytes (B74-84) [7,8].

69 During chronic HBV infection, a large number of amino acid substitutions are seen in the Core

70 gene, mainly clustered in epitopic regions. These amino acid changes have been associated

71 with viral persistence because of their impact on the host immune response and the natural

72 course of HBV infection [7,9-14]. The largest number of Core gene changes is associated with IFN

73 therapy [10-12]. The effect of nucleoside/nucleotide analogues has been little investigated,

74 although some variability in a minor epitope (Th28-47) was recently reported [11]. In another

75 recent study, entecavir and adefovir were associated with an enhanced immune response [15].

76 Selection of amino acid changes in the Core gene might result in evasion of HBV from the host

77 immune system, thereby lengthening the life of infected hepatocytes. For this reason, Core

78 gene baseline variability in chronic hepatitis B patients might be crucial for understanding the

79 evolution of the viral quasispecies in response to host immune pressure.

80 Next-generation sequencing technologies enable deep assessment of gene variability and are

81 especially useful to study the dynamics of viral quasispecies [16-20]. Core gene variability can be

82 studied with this technology, specifically, the 454 FLX platform, which analyzes fragments of

83 250 to 400 bp length. Although this length does not permit complete analysis of the gene, ultra-

84 deep analysis of the main immunodominant regions of Core protein (Th50-69 and B74-84) is

4

85   possible. The aim of this study was to analyze the variability of these main HBV Core epitopes

86   in chronic hepatitis B patients by ultra-deep pyrosequencing (UDPS).

87

88   **METHODS**

89   **Patients and samples**

90   Four chronic hepatitis B patients with complete clinical documentation were selected for the

91   study; baseline characteristics are indicated in Table 1. All patients were diagnosed with active

92   HBV replication and treated with lamivudine (LVD) 100 mg/day (Zeffix, Glaxo Wellcome UK).

93   After 18 to 24 months, they all presented mutations conferring resistance to treatment. Owing to

94   their similarities in LVD non-response, they were selected for inclusion in the study. To evaluate

95   baseline variability, a sample taken at the time of the diagnosis (hence, antiviral treatment-

96   naïve) was selected for each patient. HBV DNA of all samples was retested by real time PCR

97   TaqMan (Roche) technology, and all presented values higher than $5 \log_{10}$ IU/mL.

98   Patient 4 was sequentially analyzed and 3 samples were selected for UDPS analysis. HBV-

99   DNA had been quantified using the branched-DNA (bDNA) technology available at that time,

100   but the samples selected were retested with TaqMan technology for this study. At the time of

101   the diagnosis (baseline sample), high HBV-DNA levels were detected by bDNA (retesting with

102   TaqMan, $7 \log_{10}$ IU/mL). However, HBV-DNA spontaneously dropped below the limit of

103   detection of bDNA technology and the patient remained untreated, according to the guidelines

104   at that time. Some other samples obtained during this period without treatment were also

105   retested and showed 4 to $5 \log_{10}$ IU/mL. After 36 months (treatment-free sample), bDNA

106   significantly increased ($>8 \log_{10}$ IU/mL on TaqMan retesting) and LVD was started. After an

107   initial suboptimal response (HBV DNA decrease to 4 logs), viral breakthrough ($7 \log_{10}$ HBV-

108   DNA) was observed after 18 months, and the rtL180M and rtM204V HBV polymerase variants

109   were selected (LVD non-response sample). Ultimately, adefovir was added to LVD.

110

111   **Epitopic region amplification and UDPS amplicon preparation**

112   All the samples included in this study had HBV viral loads higher than 6 logs IU/mL. HBV-DNA

113   was extracted from serum by QIAamp microspin columns (QIAamp DNA Mini Kit, QIAGEN,

114   Hilden, Germany), according to the manufacturer's instructions. To obtain optimal amplification

5

115 of HBV DNA, the process was optimized with two PCRs. To minimize the error rate of the PCR

116 process (false nucleotide substitutions), high fidelity polymerase (Pfu Ultra-II, Stratagene, La

117 Jolla, USA) was used. At the time the study was designed, the maximal amplicon length that

118 could be analyzed by the FLX platform was 250 nucleotides; PCR primers were selected for

119 amplification of a specific 210-bp HBV fragment, which included main epitopic regions (Th50-69

120 and B74-84). The first PCR primers were sense (position 1662-1681); 5'-

121 $^C/_T$ATAAGAGGACTCTTGGACT-3' and anti-sense primers (position 2912-2931); 5'-

122 TGTTCCCA$^A/_G$GAATA$^A/_T$GGTGA-3'. The nested primers included the recognition site for UDPS,

123 in italics. The sequence of the sense primer (position 1997-2016) was 5'-

124 GCCTCCCTCGCGCCA*TCAG*ACCGCCTCAGCTCT$^C/_T$TATCG-3', and the anti-sense primer

125 (position 2178-2206) was 5'-GCCTTGCCAGCCCGC*TCAG*CCACA$^A/_G$AGTTGCCTGA$^A/_G$CTT-3'.

126 PCR products were isolated from 0.9% agarose gel, quantified using Quan-iT Picogreen sDNA

127 reagent (Invitrogen). Prior to the sequencing reaction, each amplicon was pooled to obtain a

128 concentration of 4×10^6 molecules of the HBV region. This working solution was enriched with

129 the capture beads needed for sequencing. After optimal enrichment, clonal amplification in

130 beads was performed in forward and reverse directions (emPCR kits II and III, 454 Life

131 Sciences). UDPS was carried out in the Genome Sequencer FLX (454 Life Sciences). The HBV

132 region analyzed covered amino acids 40 to 95 of the Core gene.

133

134 **Bioinformatics filter**

135 A total of 133 813 sequences were obtained. Reads were obtained with forward and reverse

136 sequences and were aligned according to the primer sequence (designed by our group). Initial

137 raw data filtering was performed as previously reported [20,21]. Briefly, reads with insertions and/or

138 deletions, reads in which the primer did not match the reference primer by at least 75%, those

139 that did not reach a fixed length, and those containing ambiguous calls were excluded. The

140 remaining sequences were then filtered again with an R commander program, based on

141 previous studies [16-18,20,22].

142 UDPS sequence error values were determined by UDPS processing of a cDNA clone in

143 triplicate. This strategy resulted in a multiparameter Poisson model [19] with the distribution of the

144 error rate per site (μ_φ) of this clone represented in arrays, according to the type of nucleotide

6

145 change (i and j indicate the nucleotide at each site, and "r" refers to the homopolymeric [r=1] or

146 non-homopolymeric [r=2] region) (Table 2).

147 Minor variants from our amplicons were defined as the number of substitutions of an expected

148 nucleotide (nt_i^s) to an observed variant (nt_j^s) in the number of reads (N) covering a site (s),

149 represented by k_j^s at each position. We calculated the probability at which that these minor

150 variants would occur if there were sequencing errors with the following equation:

$$p = 1 - \sum_{l=0}^{k_j^s - 1} \frac{e^{-\lambda} \cdot \lambda^l}{l!}$$

151

152 in which $\lambda = \mu_{ij} \times N$, and N indicates the total number of sequences. After applying the equation,

153 variants with a p value of <0.05 were included in the analysis.

154 The empirical distribution of mismatch errors determined by UDPS analysis of an HBV DNA

155 clone from the same region yielded an average of 0.006%; however in 8 positions, errors were

156 higher than 0.02% but lower than 0.05%. Therefore, the sensitivity of UDPS to detect mutations

157 was primarily limited by the highest mismatch error rate in the HBV DNA clone of 0.05%, which

158 is similar to the value recently obtained in UDPS amplicons including an internal sequence as a

159 control [20]. The measurements and biological conclusions in this study are based only on

160 mutations present at this frequency.

161

162 **Data analysis statistical tests**

163 To obtain the percentage of amino acid variability in each sample, the total number of amino

164 acid substitutions was divided by the total number of amino acids analyzed. This value gave a

165 theoretical variability for each position, and was used to estimate the expected variability for the

166 studied regions (the theoretical variability was multiplied by the length of the epitope: 20 for

167 Th50-69, 11 for the B 74-84, and 25 for the remaining positions).

168 Fisher's exact test was used to evaluate a possible relationship between the most variable

169 codons (variability ≥1%) and their position in the epitopic region or other regions. The Wilcoxon

170 signed-rank test was used to compare the evolution of the codons in the sequential analysis.

171

172 **Phenotyping, Mutagenesis, and Cell Culture**

7

128

173   Cloning of a more than full-length HBV genome [20] in pTriEx-mod vector was performed as

174   described by Durantel et al. [23]. The influence on HBV viral replication of the most conserved

175   position observed in our study was analyzed by site-directed mutagenesis (Agilent

176   Technologies, Stratagene, USA) following the manufacturer's instructions. The most conserved

177   position observed after UDPS analysis, codon 76, (see *Results*) was selected for mutagenesis.

178   The wild-type clone had an L in codon 76, which was changed to V or P to test the effect of

179   maintaining or deeply altering the physical-chemical properties of Core codon 76. The

180   introduction of mutations was confirmed by direct sequencing, as previously described [20].

181   Huh7 human hepatoma cells were cultured in Dulbecco's modified Eagle medium (DMEM)

182   supplemented with 10% calf serum. Transfection of plasmids was performed as previously

183   described [20] using Fugene-HD (Roche, Germany). The supernatant was used to quantify

184   HBsAg (Architecht, Abbot), HBeAg (Vitros, Johnson & Johnson), and HBV DNA

185   (CobasTaqman, Roche) production. The results were statistically analyzed with the Student *t*

186   test. DNA was extracted from the supernatant (QiagenAMP DNA Mini Kit, Qiagen, Germany)

187   following the manufacturer's instructions and used to evaluate HBV DNA production. As has

188   been previously described [20], to confirm that HBV-DNA detected after transfection was the

189   result of HBV replication and not due to contamination from the HBV genome in the pTriEx-mod

190   vector of the transfection experiments, the supernatant extractions were used in 1/10, $1/10^2$,

191   $1/10^3$, and $1/10^4$ dilutions and PCR amplification of HBV-DNA and pTriEx-mod was performed.

192

193   **RESULTS**

194   **Baseline variability of main epitopic regions of HBV Core gene**

195   The amplicon analyzed was limited to codons 40 to 95, which include the main Th50-69 and

196   B74-84 epitopes. A total of 122 814 sequences corresponding to 4 baseline samples were

197   analyzed, and among them, 108 403 were validated by bioinformatics and Poisson filtering. A

198   total of 61 499 were from genotype A samples and the remaining from genotype D. Variability

199   was analyzed attending to the percentage of changes in all codons of the amplicon, and the

200   results obtained for each position are shown in Table 3.

201   The two genotype A samples (patients 1 and 2) showed differences between their master

202   sequences in ten codons (41, 45, 59, 64, 65, 66, 67, 77, 84 and 87, Table 3), seven of which

8

203 were located in Th50-69 or B74-84. Of particular note, the master sequence of the motif

204 delimited by codons 64 to 67, commonly defined by $E_{64}LMT_{67}$ and recognized by T-cells [13],

205 differed in patient 1. The sequence found, $D_{64}VTN_{67}$, was completely different from the master

206 sequence of genotype A and D. The variability detected in patient 1 (0.1% average amino acid

207 variability), ranging from 0.69% to values under the cut-off (<0.05%), was the lowest in all 4

208 samples. In this patient, the main epitopic regions contained 67.7% of the changes, a

209 percentage 1.2 times higher than would be expected by the length of these regions, and the

210 changes were equally distributed between the two epitopes. In contrast, patient 2 had higher

211 variability (0.35% average amino acid variability), particularly in codons 41 (7.54%), 59 (4.75%),

212 and 63 (1.93%). Only 53.1% of these changes were located in epitopic regions, a rate similar to

213 the expected random percentage, but in the Th50-69 epitope the substitutions were 1.3 times

214 higher than would be expected. Interestingly, two of the main substitutions detected in patient 2,

215 S41 (A, 7.51%) and V59 (I, 4.21%), coincided with the consensus sequence of patient 1. The

216 third main variant position was G63 (V, 1.62%).

217 The two genotype D baseline samples (patients 3 and 4) had the same consensus sequence,

218 except in codons 64, 74, 80, and 93, which were also the most highly variable in patient 4. In

219 patient 3, five codons with more than 1% variability were detected: A41 (1.58%), I59 (1.68%),

220 A74 (2.1%), E77 (1.94%), and S87 (1.93%) (3 of them in epitopic regions). The average amino

221 acid variability was 0.26%, and 57.8% of changes were located in the main epitopic regions.

222 Overall, this percentage was not higher than expected; however, changes in B74-84 were 1.6

223 times higher than the expected random percentage (31.6% vs 19.7%). In patient 4, variability

224 was higher than 1% in 8 codons: Q40 (36.27%), L55 (18.58%), D64 (29.96%), V74 (13.72%),

225 E77 (3.03%), T80 (8.08%), T92 (18.93%), and V93 (18.89%) (5 of them in epitopic regions).

226 Linkage analysis showed that some of the main variants seen in this patient (S at codon 41, V

227 at 59, N at 74, E at 77, and N at codon 87) were located in the same viral strain (1.5% of

228 quasispecies). This observation seems to indicate possible selection by the effect of immune

229 pressure on the Core gene. Surprisingly, despite the high total amino acid variability detected in

230 patient 4 (2.69%), only 49.8% of changes affected positions located in main epitopic regions, a

231 value lower than was expected in both Th50-69 and B74-84.

232   Median baseline frequencies were compared between the epitopic and other positions. Only

233   genotype A samples showed high variability approaching significance (p=0.07, sample 1 and

234   p=0.05, sample 2) in epitopic positions. Regarding the median baseline variability (Table 3), 6 of

235   the highest values were located in positions within the main epitopic regions (codons 55, 59, 64,

236   74, 77, and 80), and 4 positions outside the main epitopes (codons 40, 41, 92, and 93)

237   accumulated high percentages of changes. The variability in positions 40, 92, and 93 was due

238   to changes in patient 4, whereas the variability of codon 41 (median 2.3% of changes) was due

239   to changes in patients 2 and 3. Interestingly, positions 64 and 66, corresponding to the

240   $E_{64}LMT_{67}$ motif of Th50-69, showed significant variability in all 4 samples (0.69%, 0.5%, 0.41%,

241   and 29.96% in position 64 and 0.12%, 0.21%, 0.11%, and 0.11% in position 66).

242   Attending to the conserved positions, 12 codons showed variability lower than the system error

243   rate (<0.05%) (positions 44, 52, 57, 68, 70, 75, 76, 78, 81, 85, 86, and 89). The most highly

244   conserved was leucine in codon 76, with frequencies clearly below the system error rate

245   (0.003%-0.02%) and a median baseline error of 0.013% (Table 3). Based on this finding, codon

246   76 was analyzed by site-directed mutagenesis analysis.

247

248   **Case report**

249   Patient 4 was selected for sequential analysis, and 3 samples were processed (Figure 1): a

250   baseline sample (also included in the *Baseline Study*), a sample following a treatment-free

251   period of 36 months, and a sample following 18 months of LVD non-response. After application

252   of the bioinformatic filter, 34 320 sequences from the treatment-free sample and 43 257

253   sequences from the LVD sample were obtained. The average amino acid variability of the

254   baseline sample was higher than that of the treatment-free one (2.69% vs 0.34%, p=0.001) and

255   the average amino acid variability of the treatment-free sample was higher than that of the LVD

256   treatment sample (0.34% vs 0.18%, p=0.001). The main variants detected by sequential

257   analysis are represented in Figure 2. Of note, 6 of these 11 variable positions (codons 55, 57,

258   64, 74, 77, 80) were located in main Core epitopic regions.

259   Linkage analysis was performed to analyze whether the most frequent amino acid substitutions

260   were simultaneously present in the same viral sequence (Figure 3a). At baseline, 50.15% of

261   sequences in the positions studied were wild type, and 38.88% were mutated sequences in the

10

262   most variable positions (15 different mutated variants, Figure 3a); 10.97% showed mutations in

263   other positions. The mutated sequences in highly variable positions detected at baseline were

264   found to be decreased in the treatment-free period (1.85%, Figure 3b) and after LVD

265   breakthrough (1.24%, Figure 3c).

266   Attending to these variable positions, the most common strain at baseline (7.5%) had only one

267   mutated codon (E40Q), followed by a strain (6.84%) with 5 mutated codons (E40Q, D64E,

268   V74G, T92N, and V93M) and another strain (5.45%) with 3 mutated codons (E40Q, L55I, and

269   D64E). Surprisingly, the baseline strain that had been selected as master after the treatment-

270   free period and maintained during LVD was a low-frequency baseline mutant strain (1.31%) with

271   the following substitutions: E40Q, L55I, D64E, T80G, T92H, and V93M (variant 12, Figure 3a).

272   The time period with an absence of therapy (between baseline and treatment-free sampling)

273   represented the complete time of HBV infection, and the HBV Core quasispecies showed a

274   tendency to decreased variability. This was reflected by a drop in the percentage of

275   accumulated variability from baseline (38.88%) to the treatment-free sample (1.85%) and

276   coincided with the change in the master sequence between the two samples, which could

277   indicate a possible alternative immune escape mechanism. No significant differences were

278   observed between samples from the treatment-free and LVD periods, which showed similar

279   composition and percentages of mutated variant strains (Figure 3).

280

281   **Evaluation of the conserved position, leucine 76, by site-directed mutagenesis**

282   As is described above, leucine (L) at codon 76 was the most highly conserved position in all the

283   samples analyzed (Table 3). Although there were other conserved positions (Q57, T70, D78,

284   and S81), codon 76 focused our interest because it was a leucine (one amino acid coded by 6

285   different codons), because of its location in the Core gene, and because it has never been

286   described as essential. L76 was even more conserved than D76 (0.05%) or P79 (0.09%), both

287   of which are reported to be involved in the Core-HBsAg interaction [24]. Mutagenic studies of L76

288   were performed to evaluate a possible essential role of this amino acid. The experiments

289   included a change to valine (V), whose hydrophobicity is similar to that of L, and a change to

290   proline (P), whose physical-chemical properties differ from those of L.

11

291 After transfection, HBsAg, HBeAg, and HBV DNA were quantified in cell culture supernatants.

292 The presence of P in position 76 significantly decreased production of HBsAg, HBeAg, and

293 HBV-DNA in comparison with the wild type (L76) (both, $p<0.001$). However, when V was in

294 position 76, a reduction was observed in HBsAg levels ($p<0.001$), but not in HBV DNA.

295 Surprisingly, the V substitution resulted in four-fold increase in HBeAg production in comparison

296 with L76 ($p<0.001$).

297

298 **DISCUSSION**

299 Several epitopic regions have been described in HBV Core gene. Core region variability under

300 antiviral treatment has been extensively studied, particularly in relation to IFN therapy [10,12].

301 However, differing patterns of amino acid substitutions under the effect of nucleotide/nucleoside

302 analogues and during periods without treatment have been recently described [11]. Some studies

303 have used the classical clonal method to analyze the evolution and composition of the entire

304 HBV Core gene [25,26], including a recent study by Tang et al, which clonally analyzed the

305 complete HBV genome in relation to antiviral therapy [27]. However, in all these reports, only a

306 small number of clones were processed (30 in Tang's study) and minor populations could not

307 be studied. Thus, high-resolution clonal studies are needed to deeply analyze the composition

308 of the HBV Core gene quasispecies. To this end, the recently developed UDPS technology

309 provides an opportunity to bypass the restrictions of the classical clonal method to determine

310 the composition of viral quasispecies [16,17,19,20].

311 When this study was designed, UDPS technology based on the 454 Life Science Platform (GS-

312 FLX, Roche Applied Science) only allowed analysis of around 200 nucleotide sequences.

313 Because of this limitation, the present work was focused on analysis of the main

314 immunodominant N-terminal Core motifs, Th50-69 and B74-84 [6,7]. The high economic cost and

315 complex computing analysis of UDPS strongly restrict the number of samples to be processed.

316 For this reason, we were only able to study six selected samples: four samples to evaluate the

317 baseline variability and two more samples to sequentially analyze one patient.

318 To achieve the main aim of UDPS analysis, a parallel analysis was needed to define the cut-off

319 for analyzing the UDPS data. In our previous report of UDPS analysis of the HBV quasispecies

320 [20], inclusion of an internal control sequence within the analyzed amplicon enabled establishment

12

321 of a cut-off percentage to define the limit of viral variability, which was set at 0.03%. In the

322 present study, a Poisson computational model was applied based on a previous study that used

323 a single parameter to define the probability of system error (μ) [19]. We approached the model by

324 using two 4x4 arrays of error values taking into account the type of nucleotide change and

325 whether the region sequence was homopolymeric or non-homopolymeric (μ_{p}). By processing an

326 HBV-DNA clone, the percentage established to differentiate variability from UDPS error was set

327 at 0.05%, the highest mismatch error rate of the clone and a value similar to the previously

328 established one [20].

329 In the study of baseline HBV quasispecies, the most variable codons (median baseline

330 variability ≥1%) were 40, 41, 55, 59, 64, 74, 77, 80, 92, and 93, all of which have been

331 previously described in a study using clonal methodology to analyze acute exacerbations in

332 HBeAg-negative patients [26]. Furthermore, some of these highly variable codons (codon 64, 74,

333 and 77) were previously described as common changes in untreated chronic hepatitis B [11]. The

334 baseline variability of the 4 samples suggested differences between genotypes A and D.

335 Genotype A sequences showed fewer codons with variability ≥1% and, consequently, lower

336 overall median variability than genotype D. The statistical analysis of the median baseline

337 frequencies showed that genotype A samples had higher variability in epitopic positions,

338 approaching significance ($p=0.07$, sample 1 and $p=0.05$, sample 2). In contrast, in genotype D

339 cases, no significant differences in variability were observed between the epitopic and other

340 positions. These findings could be related to the high frequencies of mutations in codon 40, 87,

341 92, and 93 in genotype D samples (patients 3 and 4), which might be involved in the minor

342 epitopic regions Th28-47 and Th82-101 [8,11]. These differences in HBV Core gene variability

343 seem to indicate an influence of genotype on immune activation.

344 Patient 1 showed the lowest variability, but interestingly, the consensus sequence of codons 64

345 to 67 was defined by the sequence $D_{64}VTN_{67}$, which is completely different from the well

346 characterized $E_{64}LMT_{67}$ motif. This motif is commonly recognized by T-cells, and the

347 simultaneous E64D and T67N change has been reported to reduce T-cell proliferation *in vitro* [13].

348 Hence, we suggest that the changes observed in patient 1 might be the result of an alternative

349 mechanism to escape from immune pressure over the $E_{64}LMT_{67}$ motif, which is located in the

350 immunodominant Th50-69 region. In fact, the finding that positions 64 and 66 of this motif

13

351   showed significant variability in all 4 samples suggests that the motif could be a central target

352   for immune pressure in HBV infection [11]. The low variability observed in sample 1 (0.1% of the

353   total amino acid variability) could reflect attainment of a kind of "steady state" in the viral

354   quasispecies, resulting from strong selection of an escape variant with the mutated motif

355    $D_{64}VTN_{67}$, similar that observed in the longitudinal study of patient 4.

356   Evolution of the HBV quasispecies was evaluated in a case report, which involved a baseline

357   sample and 2 additional samples, one taken after a period without treatment and one taken

358   after LVD treatment. The highest variability, which was detected in the baseline sample,

359   dropped during the treatment-free period, suggesting a reduction in the composition of HBV

360   quasispecies as a consequence of immune pressure during the period without treatment, and

361   coinciding with an exacerbation [25]. This dynamism in the composition of the viral quasispecies

362   was also observed by linkage analysis, which detected a dramatic reduction in the number of

363   viral strains and their percentage among the total variants in the treatment-free and LVD

364   samples (less than 2% of the total sequences in both samples). One of the low baseline viral

365   strains (variant 12, 1.31%, Figure 3) was strongly selected until it became the master of the

366   treatment-free and LVD samples. Hence, this variant can be labeled as an escape mutant,

367   whose selection could be related to better fitness, regardless of its initial frequency.

368   Based on these results, we postulated that the HBV quasispecies achieved a "steady state"

369   after the treatment-free period that did not change with LVD treatment, because immune

370   pressure could also be decreased during treatment. The significant differences in average

371   variability in the two periods suggest that the equilibrium is dynamic. The structure of the HBV

372   quasispecies in the 3 samples represents a complex reservoir of different minor variants,

373   resulting from natural HBV evolution and likely affected by antiviral treatment [28].

374   Several codons in our region were found to be highly conserved (positions 57, 70, 76, 78 and

375   81). Position 76, located at the tip of the spike was of especial interest, being next to the major

376   immunodominant region (MIR) and part of the B 74-84 main epitope. Indeed, this position has

377   never been described as essential, in contrast to the nearby positions 78 and 79 [24]. The HBV

378   Core sequence is involved in the process of capsid conformation by interacting with the surface

379   antigens in viral particle assembly. However, the interaction between HBsAg and HBcAg in

380   virions is still unclear [29-31]. The Core region analyzed in the present study is part of the assembly

14

381 domain (amino acids 1-149); thus, the changes in the region analyzed might potentially modify

382 the shell conformation.

383 Electrostatic interactions between Core and HBsAg take place at the tip of the spike of the Core

384 antigen (codons 74-84) [32,33]. Cryo-electron microscopy studies have shown that codon 78 and

385 79 are within the contact area of Core with envelope proteins. [24] This essential role might

386 explain the high level of conservation of these positions (especially codon 78, 0.05%) observed

387 in our UDPS analysis. However, surprisingly, the most conserved codon of the HBV Core region

388 in baseline and sequential samples was leucine at position 76.

389 To our knowledge, few studies [30,31] have analyzed the effect of single Core amino acid

390 mutations on HBV replication *in vitro*. Only one such study conducted by Ponsel and Bruss [31]

391 evaluated the leucine 76 position (among others) by inducing a change to alanine; no significant

392 reduction in nucleocapsid or virion production was observed. In the present study, we induced

393 changes in the hydrophobic characteristics of position 76 and determined their effect on HBsAg,

394 HBeAg, and HBV DNA production. In contrast to the results of Ponsel and Bruss, we found a

395 significant reduction in HBsAg production with both the V and P changes, suggesting possible

396 involvement of L76 of HBcAg in the HBsAg interaction. A significant decrease in HBV replication

397 in the presence of P76 was detected, leading us to speculate that the hydrophobic

398 characteristics of position 76, conferred by the presence of L or V, are needed for HBV

399 replication. The increase in HBeAg production in the presence of the L76V mutation and the

400 absence of increasing HBV DNA we found are surprising, particularly because some authors

401 have reported a correlation between HBeAg and HBV-DNA levels [34]. Nonetheless, this

402 correlation was not found by other authors[35] and currently remains controversial. We suggest

403 that this unexpected HBeAg increase may indicate alternative pathways between HBeAg and

404 HBV replication, as has been indicated previously [36]. Based on our *in vitro* results, we postulate

405 that the amino acid changes induced in the Core sequence are not as important as the structure

406 adopted by the capsid and HBeAg.

407 Our study is mainly limited by the high cost of the UDPS process. A larger number of samples,

408 additional sequential studies and duplicate UDPS experiments would have given more

409 information about sites in HBV Core potentially involved in quasispecies evolution, and related

410 to HBV chronic infection and to treatment non-response. In addition, at the time of the study, the

15

411 available UDPS methodology only allowed analysis of sequences up to 250 bp. For this reason,

412 we limited the study to the widely described main epitopic regions included in the HBV Core,

413 previously studied by conventional methodologies [6,7,11,25].

414 In conclusion, this study validates application of UDPS methodology to study the variability of

415 the main epitopes of the HBV Core gene and substantiates the significant richness of the HBV

416 baseline quasispecies. Although other studies have analyzed the complete Core region, this is

417 the first study analyzing the HBV Core gene by UDPS technology. The largest number of

418 variable codons was mainly detected in the Th50-69 and B74-84 regions, and a probable

419 relationship with HBV genotype and Core variability was suggested. However, more extended

420 analyses with a larger number of samples must be performed to confirm this possibility. In

421 addition, important variability was associated with well characterized Th-cell motifs, such as

422 $E_{64}LMT_{67}$, suggesting that the host immune system might be the main factor responsible for

423 HBV Core evolution. Furthermore, UDPS showed that codons 78 and 79 were highly

424 conserved, in keeping with their involvement in the interaction between the HBV virion capsid

425 and envelope. However, codon 76 was found to be the most highly conserved, to an even

426 greater degree than codons 78 and 79. Thus, we postulate possible involvement of codon 76 in

427 interactions with HBsAg or in HBeAg conformation. The dynamism of the HBV quasispecies

428 was tested in one patient in whom detection of a strong selection of one of the minor variants

429 present in the baseline quasispecies was seen, coinciding with a decrease in Core variability

430 during a treatment-free period and on lamivudine treatment. This decrease seemed to be the

431 result of a "steady state" situation of the HBV quasispecies after selection of the most highly fit

432 variant. We found that UDPS is a useful massive clonal method to deeply analyze viral

433 quasispecies that is mainly limited by the current high cost of this technology. However,

434 because of the advantages of UDPS in the study of viral quasispecies, cost will probably

435 decrease as application of this technology increases, enabling processing of samples from a

436 large number of patients.

437

438 **ACKNOWLEDGMENTS**

16

440   References

441   1.   Thursz, MR, Kwiatkowski, D, Allsopp, CE, Greenwood, BM, Thomas, HC, Hill, AV .
442   Association between an MHC class II allele and clearance of hepatitis B virus in the
443   Gambia. N Engl J Med 1995; 332: 1065-1069 [PMID: 7898524
444   DOI:10.1056/NEJM199504203321604].

445   2.   Ehata, T, Omata, M, Yokosuka, O, Hosoda, K, Ohto, M . Variations in codons 84-101 in
446   the core nucleotide sequence correlate with hepatocellular injury in chronic hepatitis B
447   virus infection. J Clin Invest 1992; 89: 332-338 [PMID: 1729279 DOI:10.1172/JCI115581].

448   3.   Preikschat, P, Gunther, S, Reinhold, S, Will, H, Budde, K, Neumayer, HH, Kruger, DH,
449   Meisel, H . Complex HBV populations with mutations in core promoter, C gene, and pre-S
450   region are associated with development of cirrhosis in long-term renal transplant
451   recipients. Hepatology 2002; 35: 466-477 [PMID: 11826424
452   DOI:10.1053/jhep.2002.30698].

453   4.   Sung, FY, Jung, CM, Wu, CF, Lin, CL, Liu, CJ, Liaw, YF, Tsai, KS, Yu, MW . Hepatitis B
454   virus core variants modify natural course of viral infection and hepatocellular carcinoma
455   progression. Gastroenterology 2009; 137: 1687-1697 [PMID: 19664630
456   DOI:10.1053/j.gastro.2009.07.063].

457   5.   Vanlandschoot, P, Cao, T, Leroux-Roels, G . The nucleocapsid of the hepatitis B virus: a
458   remarkable immunogenic structure. Antiviral Res 2003; 60: 67-74 [PMID: 14638400.

459   6.   Ferrari, C, Bertoletti, A, Penna, A, Cavalli, A, Valli, A, Missale, G, Pilli, M, Fowler, P,
460   Giuberti, T, Chisari, FV . Identification of immunodominant T cell epitopes of the hepatitis B
461   virus nucleocapsid antigen. J Clin Invest 1991; 88: 214-222 [PMID: 1711541
462   DOI:10.1172/JCI115280].

463   7.   Carman, WF, Boner, W, Fattovich, G, Colman, K, Dornan, ES, Thursz, M, Hadziyannis, S
464   . Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive
465   disease and in T helper cell epitopes during clinical remission. J Infect Dis 1997; 175:
466   1093-1100 [PMID: 9129071.

467   8.   Milich, DR, McLachlan, A . The nucleocapsid of hepatitis B virus is both a T-cell-
468   independent and a T-cell-dependent antigen. Science 1986; 234: 1398-1401 [PMID:
469   3491425.

470   9.   Bozkaya, H, Ayola, B, Lok, AS . High rate of mutations in the hepatitis B core gene during
471   the immune clearance phase of chronic hepatitis B virus infection. Hepatology 1996; 24:
472   32-37 [PMID: 8707278 DOI:10.1002/hep.510240107].

473   10.   Radecke, K, Protzer, U, Trippler, M, Meyer Zum Buschenfelde, KH, Gerken, G .
474   Selection of hepatitis B virus variants with aminoacid substitutions inside the core antigen
475   during interferon-alpha therapy. J Med Virol 2000; 62: 479-486 [PMID: 11074477.

476   11.   Homs, M, Jardi, R, Buti, M, Schaper, M, Tabernero, D, Fernandez-Fernandez, P, Quer,
477   J, Esteban, R, Rodriguez-Frias, F . HBV core region variability: effect of antiviral
478   treatments on main epitopic regions. Antivir Ther 2011; 16: 37-49 [PMID: 21311107
479   DOI:10.3851/IMP1701].

480   12.   Naoumov, NV, Thomas, MG, Mason, AL, Chokshi, S, Bodicky, CJ, Farzaneh, F,
481   Williams, R, Perrillo, RP . Genomic variations in the hepatitis B core gene: a possible factor
482   influencing response to interferon alfa treatment. Gastroenterology 1995; 108: 505-514
483   [PMID: 7835593.

17

138

484   13.   Torre, F, Cramp, M, Owsianka, A, Dornan, E, Marsden, H, Carman, W, Williams, R,
485   Naoumov, NV . Direct evidence that naturally occurring mutations within hepatitis B core
486   epitope alter CD4+ T-cell reactivity. *J Med Virol* 2004; **72**: 370-376 [PMID: 14748060
487   DOI:10.1002/jmv.20016].

488   14.   Alexopoulou, A, Karayiannis, P, Hadziyannis, SJ, Aiba, N, Thomas, HC . Emergence and
489   selection of HBV variants in an anti-HBe positive patient persistently infected with quasi-
490   species. *J Hepatol* 1997; **26**: 748-753 [PMID: 9126785.

491   15.   Jiang, Y, Li, W, Yu, L, Liu, J, Xin, G, Yan, H, Sun, P, Zhang, H, Xu, D, Niu, J . Enhancing
492   the antihepatitis B virus immune response by adefovir dipivoxil and entecavir therapies.
493   *Cell Mol Immunol* 2011; **8**: 75-82 [PMID: 20921939 DOI:10.1038/cmi.2010.37].

494   16.   Margeridon-Thermet, S, Shulman, NS, Ahmed, A, Shahriar, R, Liu, T, Wang, C, Holmes,
495   SP, Babrzadeh, F, Gharizadeh, B, Hanczaruk, B, Simen, BB, Egholm, M, Shafer, RW .
496   Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and
497   nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients.
498   *J Infect Dis* 2009; **199**: 1275-1285 [PMID: 19301976 DOI:10.1086/597808].

499   17.   Solmone, M, Vincenti, D, Prosperi, MC, Bruselles, A, Ippolito, G, Capobianchi, MR . Use
500   of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of
501   hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in
502   reverse transcriptase and hepatitis B S antigen. *J Virol* 2009; **83**: 1718-1726 [PMID:
503   19073746 DOI:10.1128/JVI.02011-08].

504   18.   Simen, BB, Simons, JF, Hullsiek, KH, Novak, RM, Macarthur, RD, Baxter, JD, Huang, C,
505   Lubeski, C, Turenchalk, GS, Braverman, MS, Desany, B, Rothberg, JM, Egholm, M, Kozal,
506   MJ, Terry Beirn Community Programs for Clinical Research on AIDS . Low-abundance
507   drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive
508   patients significantly impact treatment outcomes. *J Infect Dis* 2009; **199**: 693-701 [PMID:
509   19210162 DOI:10.1086/596736].

510   19.   Wang, C, Mitsuya, Y, Gharizadeh, B, Ronaghi, M, Shafer, RW . Characterization of
511   mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance.
512   *Genome Res* 2007; **17**: 1195-1201 [PMID: 17600086 DOI:10.1101/gr.6468307].

513   20.   Homs, M, Buti, M, Quer, J, Jardi, R, Schaper, M, Tabernero, D, Ortega, I, Sanchez, A,
514   Esteban, R, Rodriguez-Frias, F . Ultra-deep pyrosequencing analysis of the hepatitis B
515   virus preCore region and main catalytic motif of the viral polymerase in the same viral
516   genome. *Nucleic Acids Res* 2011 [PMID: 21742757 DOI:10.1093/nar/gkr451].

517   21.   Campbell, PJ, Pleasance, ED, Stephens, PJ, Dicks, E, Rance, R, Goodhead, I, Follows,
518   GA, Green, AR, Futreal, PA, Stratton, MR . Subclonal phylogenetic structures in cancer
519   revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 2008; **105**: 13081-13086
520   [PMID: 18723673 DOI:10.1073/pnas.0801523105].

521   22.   Huse, SM, Huber, JA, Morrison, HG, Sogin, ML, Welch, DM . Accuracy and quality of
522   massively parallel DNA pyrosequencing. *Genome Biol* 2007; **8**: R143 [PMID: 17659080
523   DOI:10.1186/gb-2007-8-7-r143].

524   23.   Durantel, D, Carrouee-Durantel, S, Werle-Lapostolle, B, Brunelle, MN, Pichoud, C,
525   Trepo, C, Zoulim, F . A new strategy for studying in vitro the drug susceptibility of clinical
526   isolates of human hepatitis B virus. *Hepatology* 2004; **40**: 855-864 [PMID: 15382118
527   DOI:10.1002/hep.20388].

528   24.   Seitz, S, Urban, S, Antoni, C, Bottcher, B . Cryo-electron microscopy of hepatitis B
529   virions reveals variability in envelope capsid interactions. *EMBO J* 2007; **26**: 4160-4167
530   [PMID: 17762862 DOI:10.1038/sj.emboj.7601841].

18

484   13.   Torre, F, Cramp, M, Owsianka, A, Dornan, E, Marsden, H, Carman, W, Williams, R,
485   Naoumov, NV . Direct evidence that naturally occurring mutations within hepatitis B core
486   epitope alter CD4+ T-cell reactivity. *J Med Virol* 2004; **72**: 370-376 [PMID: 14748060
487   DOI:10.1002/jmv.20016].

488   14.   Alexopoulou, A, Karayiannis, P, Hadziyannis, SJ, Aiba, N, Thomas, HC . Emergence and
489   selection of HBV variants in an anti-HBe positive patient persistently infected with quasi-
490   species. *J Hepatol* 1997; **26**: 748-753 [PMID: 9126785.

491   15.   Jiang, Y, Li, W, Yu, L, Liu, J, Xin, G, Yan, H, Sun, P, Zhang, H, Xu, D, Niu, J . Enhancing
492   the antihepatitis B virus immune response by adefovir dipivoxil and entecavir therapies.
493   *Cell Mol Immunol* 2011; **8**: 75-82 [PMID: 20921939 DOI:10.1038/cmi.2010.37].

494   16.   Margeridon-Thermet, S, Shulman, NS, Ahmed, A, Shahriar, R, Liu, T, Wang, C, Holmes,
495   SP, Babrzadeh, F, Gharizadeh, B, Hanczaruk, B, Simen, BB, Egholm, M, Shafer, RW .
496   Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and
497   nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients.
498   *J Infect Dis* 2009; **199**: 1275-1285 [PMID: 19301976 DOI:10.1086/597808].

499   17.   Solmone, M, Vincenti, D, Prosperi, MC, Bruselles, A, Ippolito, G, Capobianchi, MR . Use
500   of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of
501   hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in
502   reverse transcriptase and hepatitis B S antigen. *J Virol* 2009; **83**: 1718-1726 [PMID:
503   19073746 DOI:10.1128/JVI.02011-08].

504   18.   Simen, BB, Simons, JF, Hullsiek, KH, Novak, RM, Macarthur, RD, Baxter, JD, Huang, C,
505   Lubeski, C, Turenchalk, GS, Braverman, MS, Desany, B, Rothberg, JM, Egholm, M, Kozal,
506   MJ, Terry Beirn Community Programs for Clinical Research on AIDS . Low-abundance
507   drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive
508   patients significantly impact treatment outcomes. *J Infect Dis* 2009; **199**: 693-701 [PMID:
509   19210162 DOI:10.1086/596736].

510   19.   Wang, C, Mitsuya, Y, Gharizadeh, B, Ronaghi, M, Shafer, RW . Characterization of
511   mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance.
512   *Genome Res* 2007; **17**: 1195-1201 [PMID: 17600086 DOI:10.1101/gr.6468307].

513   20.   Homs, M, Buti, M, Quer, J, Jardi, R, Schaper, M, Tabernero, D, Ortega, I, Sanchez, A,
514   Esteban, R, Rodriguez-Frias, F . Ultra-deep pyrosequencing analysis of the hepatitis B
515   virus preCore region and main catalytic motif of the viral polymerase in the same viral
516   genome. *Nucleic Acids Res* 2011 [PMID: 21742757 DOI:10.1093/nar/gkr451].

517   21.   Campbell, PJ, Pleasance, ED, Stephens, PJ, Dicks, E, Rance, R, Goodhead, I, Follows,
518   GA, Green, AR, Futreal, PA, Stratton, MR . Subclonal phylogenetic structures in cancer
519   revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 2008; **105**: 13081-13086
520   [PMID: 18723673 DOI:10.1073/pnas.0801523105].

521   22.   Huse, SM, Huber, JA, Morrison, HG, Sogin, ML, Welch, DM . Accuracy and quality of
522   massively parallel DNA pyrosequencing. *Genome Biol* 2007; **8**: R143 [PMID: 17659080
523   DOI:10.1186/gb-2007-8-7-r143].

524   23.   Durantel, D, Carrouee-Durantel, S, Werle-Lapostolle, B, Brunelle, MN, Pichoud, C,
525   Trepo, C, Zoulim, F . A new strategy for studying in vitro the drug susceptibility of clinical
526   isolates of human hepatitis B virus. *Hepatology* 2004; **40**: 855-864 [PMID: 15382118
527   DOI:10.1002/hep.20388].

528   24.   Seitz, S, Urban, S, Antoni, C, Bottcher, B . Cryo-electron microscopy of hepatitis B
529   virions reveals variability in envelope capsid interactions. *EMBO J* 2007; **26**: 4160-4167
530   [PMID: 17762862 DOI:10.1038/sj.emboj.7601841].

18

571 **FIGURE LEGEND**

572 **Figure 1.** Variability of all codons analyzed from the 3 sequentially studied samples,

573 corresponding to Patient 4. Percentages higher than 2% are indicated at the top of the bars.

574

575 **Figure 2.** Evolution of the most variable codons of the Core region analyzed in the sequentially

576 studied patient. The variable position and its corresponding master amino acid are shown on

577 the x-axis. The main mutated amino acids are indicated at the top of the bars.

578

579 **Figure 3.** Quasispecies composition of the HBV Core region in the 3 sequentially analyzed

580 samples. **(a)** Baseline variants (from 1-15) in percentages ≥1%, **(b)** Treatment-free variability,

581 and **(c)** LVD sample variability of the most frequent amino acid substitutions defined in Figure 2.

582 Linkage analysis was also performed attending these most variable codons.

20

583 **TABLES**

584

585 **Table 1.** Baseline characteristics of the four patients included in the study

Patient	Sex	Age	ALT (IU/mL)	HBV DNA (log$_{10}$ IU/mL)	Genotype	HBeAg status
1	Male	46	167	7.4	A	- *
2	Male	39	95	7.6	A	+
3	Male	31	392	8.3	D	-
4	Female	55	117	7.5	D	-

586

587 Age at the time of sample collection. * Wild type in main preCore mutation

588

589

21

142

590 **Table 2.** Arrays of μ_{ijr} mismatch frequencies in homopolymeric (r=1) and non-homopolymeric

591 (r=2) regions generated from average misincorporation errors during PCR amplification and

592 pyrosequencing of 3 replicates from an HBV DNA clone

593

Homopolymeric	j=A	j=C	j=T	j=G
i=A	1	2.34×10^{-6}	2.04×10^{-5}	3.49×10^{-6}
i=C	4.98×10^{-6}	1	3.56×10^{-5}	1.95×10^{-6}
i=T	3.99×10^{-6}	2.91×10^{-6}	1	2.04×10^{-6}
i=G	1.50×10^{-5}	3.92×10^{-5}	4.40×10^{-6}	1
Non-homopolymeric	j=A	j=C	j=T	j=G
i=A	1	2.93×10^{-6}	1.47×10^{-5}	7.53×10^{-6}
i=C	4.83×10^{-6}	1	1.01×10^{-5}	3.20×10^{-6}
i=T	1.21×10^{-6}	7.39×10^{-6}	1	1.77×10^{-5}
i=G	4.23×10^{-5}	1.41×10^{-6}	2.91×10^{-6}	1

594

22

595 **Table 3** Baseline variability of all the codons analyzed. Subindexed numbers indicate the
596 patient in whom the master amino acid (AA) was detected.

Codon	Patient N seqs Master AA	Pt 1 33245 % var	Pt 2 28254 % var	Pt 3 27156 % var	Pt 4 19748 % var	Median 597 baseline variability
40	E	0.17	0.00	0.27	36.27	9.18
41	A_{134}/S_2	0.04	**7.54**	**1.58**	0.03	**2.30**
42	L	0.20	0.08	0.14	0.09	0.13
43	E	0.06	0.07	0.08	0.12	0.08
44	S	0.03	0.01	0.01	0.05	0.03
45	S_1/P_{234}	0.24	0.07	0.10	0.09	0.12
46	E	0.06	0.05	0.06	0.08	0.06
47	H	0.06	0.06	0.04	0.03	0.05
48	C	0.09	0.18	0.07	0.27	0.15
49	S	0.09	0.07	0.03	0.01	0.05
50	P	0.07	0.04	0.08	0.02	0.05
51	H	0.14	0.11	0.06	0.01	0.08
52	H	0.03	0.04	0.02	0.02	0.03
53	T	0.11	0.14	0.07	0.04	0.09
54	A	0.14	0.18	0.17	0.19	0.17
55	L	0.03	0.05	0.06	**18.58**	**4.68**
56	R	0.21	0.22	0.14	0.10	0.17
57	Q	0.05	0.05	0.03	0.00	0.03
58	A	0.15	0.18	0.09	0.04	0.12
59	I_{134}/V_2	0.04	**4.75**	**1.68**	0.03	**1.62**
60	L	0.03	0.02	0.03	0.06	0.04
61	C	0.10	0.08	0.04	0.03	0.06
62	W	0.15	0.21	0.19	0.09	0.16
63	G	0.17	**1.93**	0.34	0.18	0.65
64	D_{14}/E_{23}	0.69	0.50	0.41	**29.96**	**7.89**
65	V_1/L_{234}	0.08	0.02	0.09	0.19	0.09
66	T_1/M_{234}	0.12	0.21	0.11	0.11	0.14
67	N_1/T_{234}	0.05	0.06	0.06	0.04	0.05
68	L	0.04	0.02	0.03	0.02	0.03
69	A	0.05	0.12	0.10	0.09	0.09
70	T	0.05	0.02	0.02	0.01	0.02
71	W	0.11	0.24	0.20	0.17	0.18
72	V	0.07	0.12	0.07	0.03	0.07
73	G	0.17	0.28	0.19	0.11	0.19
74	$N_{12}/A_3/V_4$	0.46	0.04	**2.06**	**13.72**	**4.07**
75	N	0.01	0.02	0.02	0.02	0.02
76	L	0.00	0.01	0.02	0.02	0.01
77	G_{134}/Q_2	0.14	0.30	**1.94**	**3.03**	**1.35**
78	D	0.03	0.08	0.03	0.08	0.05
79	P	0.16	0.10	0.07	0.03	0.09
80	A_{123}/T_4	0.11	0.17	0.12	**8.09**	**2.12**
81	S	0.04	0.03	0.01	0.00	0.02
82	R	0.19	0.12	0.15	0.06	0.13
83	D	0.05	0.04	0.06	0.12	0.07
84	Q_1/L_{234}	0.21	0.52	0.12	0.05	0.23
85	V	0.02	0.01	0.03	0.01	0.02
86	V	0.02	0.03	0.01	0.04	0.03
87	N_1/S_{234}	0.06	0.03	**1.93**	0.02	0.51
88	Y	0.08	0.03	0.04	0.05	0.05
89	V	0.01	0.00	0.02	0.03	0.02
90	N	0.02	0.02	0.08	0.00	0.03
91	T	0.07	0.05	0.02	0.03	0.04
92	N	0.03	0.02	0.70	**18.93**	**4.92**
93	M_{123}/V_4	0.02	0.05	0.15	**19.89**	**4.78**
94	G	0.04	0.12	0.15	0.09	0.10
95	L	0.03	0.02	0.15	0.14	0.09

598

23

144

Figure 1

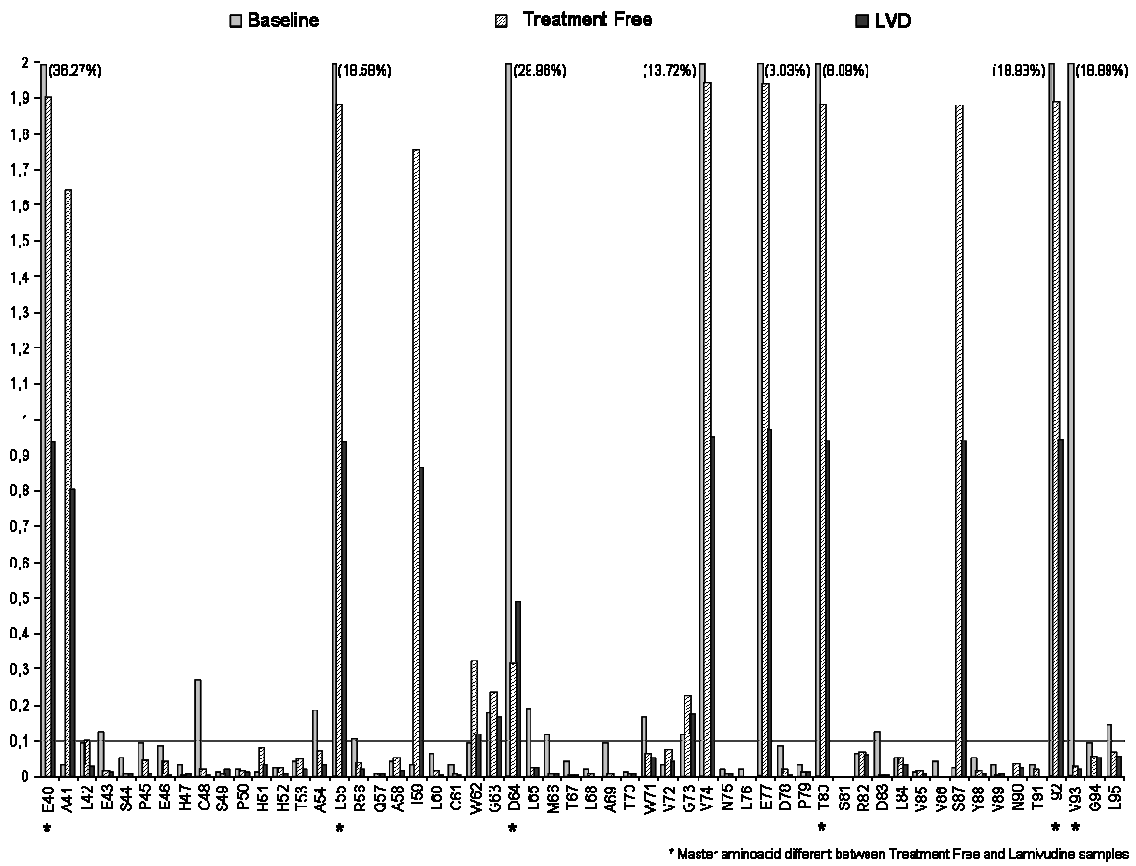* Master aminoacid differert between Treatment Free and Lamivudine samples
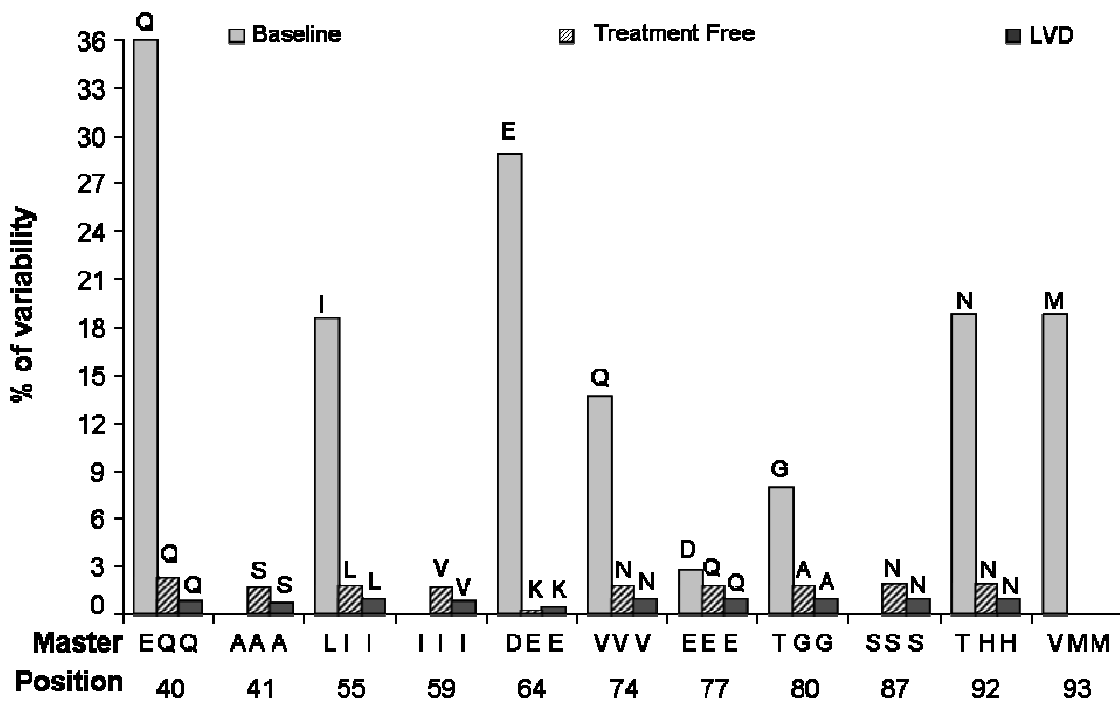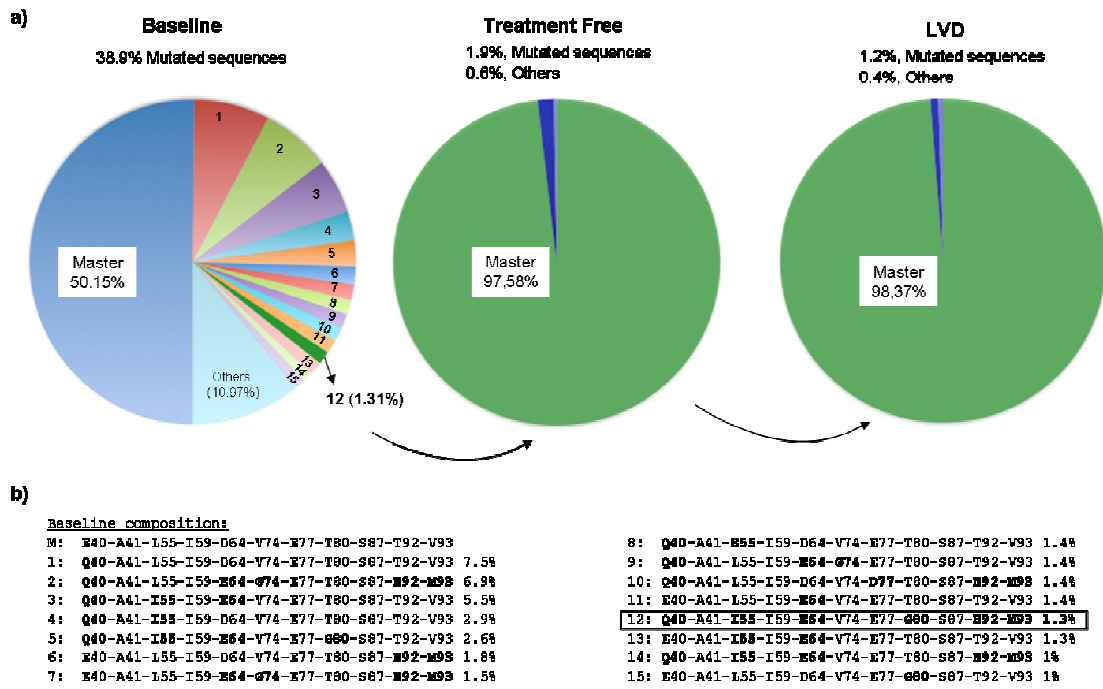
Figure 2

Figure 3

Bibliography

8   Bibliography

(1) Alter HJ, Blumberg BS. Further studies on a "new" human isoprecipitin system (Australia antigen). Blood 1966 Mar;27(3):297-309.

(2) 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 Oct 25;281(5733):646-650.

(3) Hepatitis B vaccine. Lancet 1980 Dec 6;2(8206):1229-1230.

(4) Dejean A, Lugassy C, Zafrani S, Tiollais P, Brechot C. Detection of hepatitis B virus DNA in pancreas, kidney and skin of two human carriers of the virus. J Gen Virol 1984 Mar;65 (Pt 3)(Pt 3):651-655.

(5) Zhang Y, Ma X, Fang L, Lin S, Wu Z, Gu J. The existence and significance of hepatitis B virus DNA in glomerulonephritis.
. Nephrology 1996;2(2):119-125.

(6) WHO. Fact sheet No 204. 2008; Available at: http://www.who.int/mediacentre/factsheets/fs204/en/. Accessed November/22, 2011.

(7) Beck J, Nassal M. Hepatitis B virus replication. World J Gastroenterol 2007 Jan 7;13(1):48-64.

(8) Patient R, Hourioux C, Roingeard P. Morphogenesis of hepatitis B virus and its subviral envelope particles. Cell Microbiol 2009 Nov;11(11):1561-1570.

(9) Ganem D, Prince AM. Hepatitis B virus infection--natural history and clinical consequences. N Engl J Med 2004 Mar 11;350(11):1118-1129.

(10) Yang CY, Kuo TH, Ting LP. Human hepatitis B viral e antigen interacts with cellular interleukin-1 receptor accessory protein and triggers interleukin-1 response. J Biol Chem 2006 Nov 10;281(45):34525-34536.

(11) Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarvey MJ, Makris A, et al. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet 1989 Sep 9;2(8663):588-591.

(12) Rodriguez-Frias F, Buti M, Jardi R, Cotrina M, Viladomiu L, Esteban R, et al. Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. Hepatology 1995 Dec;22(6):1641-1647.

(13) Okamoto H, Yotsumoto S, Akahane Y, Yamanaka T, Miyazaki Y, Sugai 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 Mar;64(3):1298-1303.

(14) Okamoto H, Tsuda F, Akahane Y, Sugai Y, Yoshiba M, Moriyama K, et al. Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. J Virol 1994 Dec;68(12):8102-8110.

(15) Protzer U, Trippler M, Ohl J, Knolle P, Duchmann R, Meyer zum Buschenfelde KH, et al. Rare pre-core stop-codon mutant nt. 1897 predominates over wide-spread mutant nt. 1896 in an unusual course of chronic hepatitis B. J Viral Hepat 1996 May;3(3):155-162.

(16) Nassal M, Rieger A, Steinau O. Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking. J Mol Biol 1992 Jun 20;225(4):1013-1025.

(17) Seitz S, Urban S, Antoni C, Bottcher B. Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. EMBO J 2007 Sep 19;26(18):4160-4167.

(18) Wynne SA, Crowther RA, Leslie AG. The crystal structure of the human hepatitis B virus capsid. Mol Cell 1999 Jun;3(6):771-780.

(19) Diepolder HM, Ries G, Jung MC, Schlicht HJ, Gerlach JT, Gr ner N, et al. Differential antigen-processing pathways of the hepatitis B virus e and core proteins. Gastroenterology 1999 Mar;116(3):650-657.

(20) Vanlandschoot P, Cao T, Leroux-Roels G. The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure. Antiviral Res 2003 Oct;60(2):67-74.

(21) Marinos G, Smith HM, Naoumov NV, Williams R. Quantitative assessment of serum IgM anti-HBc in the natural course and during interferon treatment of chronic hepatitis B virus infection. Hepatology 1994 Feb;19(2):303-311.

(22) Carman WF, Boner W, Fattovich G, Colman K, Dornan ES, Thursz M, et al. Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive disease and in T helper cell epitopes during clinical remission. J Infect Dis 1997 May;175(5):1093-1100.

(23) Milich DR, McLachlan A. The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen. Science 1986 Dec 12;234(4782):1398-1401.

(24) Ferrari C, Bertoletti A, Penna A, Cavalli A, Valli A, Missale G, et al. Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. J Clin Invest 1991 Jul;88(1):214-222.

(25) Salfeld J, Pfaff E, Noah M, Schaller H. Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. J Virol 1989 Feb;63(2):798-808.

(26) Jung MC, Spengler U, Schraut W, Hoffmann R, Zachoval R, Eisenburg J, et al. Hepatitis B virus antigen-specific T-cell activation in patients with acute and chronic hepatitis B. J Hepatol 1991 Nov;13(3):310-317.

(27) Penna A, Del Prete G, Cavalli A, Bertoletti A, D'Elios MM, Sorrentino R, et al. Predominant T-helper 1 cytokine profile of hepatitis B virus nucleocapsid-specific T cells in acute self-limited hepatitis B. Hepatology 1997 Apr;25(4):1022-1027.

(28) Cao T, Desombere I, Vanlandschoot P, Sallberg M, Leroux-Roels G. Characterization of HLA DR13-restricted CD4(+) T cell epitopes of hepatitis B core antigen associated with self-limited, acute hepatitis B. J Gen Virol 2002 Dec;83(Pt 12):3023-3033.

(29) Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY. Molecular evolution of hepatitis B virus over 25 years. J Virol 2006 Nov;80(21):10307-10314.

(30) 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 Feb;64(2):613-620.

(31) Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, et al. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J Gen Virol 2000 Jan;81(Pt 1):67-74.

(32) Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J 1989 Dec 1;8(12):3867-3874.

(33) Das K, Xiong X, Yang H, Westland CE, Gibbs CS, Sarafianos SG, 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 May;75(10):4771-4779.

(34) Rodriguez-Frias F, Jardi R, Schaper M, Buti M, Ferrer-Costa C, Tabernero D, et al. Adefovir for chronic hepatitis B treatment: identification of virological markers linked to therapy response. Antivir Ther 2008;13(8):991-999.

(35) Steitz TA. DNA polymerases: structural diversity and common mechanisms. J Biol Chem 1999 Jun 18;274(25):17395-17398.

(36) Doublie S, Sawaya MR, Ellenberger T. An open and closed case for all polymerases. Structure 1999 Feb 15;7(2):R31-5.

(37) Dienstag JL, Perrillo RP, Schiff ER, Bartholomew M, Vicary C, Rubin M. A preliminary trial of Lamivudine for chronic hepatitis B infection. N Engl J Med 1995 Dec 21;333(25):1657-1661.

(38) Allen MI, Deslauriers M, Andrews CW, Tipples GA, Walters KA, Tyrrell DL, et al. Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. Lamivudine Clinical Investigation Group. Hepatology 1998 Jun;27(6):1670-1677.

(39) Tang H, Oishi N, Kaneko S, Murakami S. Molecular functions and biological roles of hepatitis B virus x protein. Cancer Sci 2006 Oct;97(10):977-983.

(40) Kumar V, Jayasuryan N, Kumar R. A truncated mutant (residues 58-140) of the hepatitis B virus X protein retains transactivation function. Proc Natl Acad Sci U S A 1996 May 28;93(11):5647-5652.

(41) Bouchard MJ, Wang L, Schneider RJ. Activation of focal adhesion kinase by hepatitis B virus HBx protein: multiple functions in viral replication. J Virol 2006 May;80(9):4406-4414.

(42) Stoeckl L, Funk A, Kopitzki A, Brandenburg B, Oess S, Will H, et al. Identification of a structural motif crucial for infectivity of hepatitis B viruses. Proc Natl Acad Sci U S A 2006 Apr 25;103(17):6730-6734.

(43) Yeh CT, Wong SW, Fung YK, Ou JH. Cell cycle regulation of nuclear localization of hepatitis B virus core protein. Proc Natl Acad Sci U S A 1993 Jul 15;90(14):6459-6463.

(44) Volz T, Lutgehetmann M, Wachtler P, Jacob A, Quaas A, Murray JM, et al. Impaired intrahepatic hepatitis B virus productivity contributes to low viremia in most HBeAg-negative patients. Gastroenterology 2007 Sep;133(3):843-852.

(45) Neumann AU. Hepatitis B viral kinetics: a dynamic puzzle still to be resolved. Hepatology 2005 Aug;42(2):249-254.

(46) Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Brechot C, et al. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. Gut 2005 Aug;54(8):1162-1168.

(47) Nassal M. Hepatitis B viruses: reverse transcription a different way. Virus Res 2008 Jun;134(1-2):235-249.

(48) Kidd AH, Kidd-Ljunggren K. A revised secondary structure model for the 3'-end of hepatitis B virus pregenomic RNA. Nucleic Acids Res 1996 Sep 1;24(17):3295-3301.

(49) Noguchi C, Ishino H, Tsuge M, Fujimoto Y, Imamura M, Takahashi S, et al. G to A hypermutation of hepatitis B virus. Hepatology 2005 Mar;41(3):626-633.

(50) Moolla N, Kew M, Arbuthnot P. Regulatory elements of hepatitis B virus transcription. J Viral Hepat 2002 Sep;9(5):323-331.

(51) Chen IH, Huang CJ, Ting LP. Overlapping initiator and TATA box functions in the basal core promoter of hepatitis B virus. J Virol 1995 Jun;69(6):3647-3657.

(52) Pollack JR, Ganem D. Site-specific RNA binding by a hepatitis B virus reverse transcriptase initiates two distinct reactions: RNA packaging and DNA synthesis. J Virol 1994 Sep;68(9):5579-5587.

(53) Knaus T, Nassal M. The encapsidation signal on the hepatitis B virus RNA pregenome forms a stem-loop structure that is critical for its function. Nucleic Acids Res 1993 Aug 25;21(17):3967-3975.

(54) Flodell S, Schleucher J, Cromsigt J, Ippel H, Kidd-Ljunggren K, Wijmenga S. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. Nucleic Acids Res 2002 Nov 1;30(21):4803-4811.

(55) Lok AS, Akarca U, Greene S. Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. Proc Natl Acad Sci U S A 1994 Apr 26;91(9):4077-4081.

(56) Wang GH, Seeger C. Novel mechanism for reverse transcription in hepatitis B viruses. J Virol 1993 Nov;67(11):6507-6512.

(57) Orito E, Mizokami M, Ina Y, Moriyama EN, Kameshima N, Yamamoto M, et al. Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. Proc Natl Acad Sci U S A 1989 Sep;86(18):7059-7062.

(58) Dandri M, Murray JM, Lutgehetmann M, Volz T, Lohse AW, Petersen J. Virion half-life in chronic hepatitis B infection is strongly correlated with levels of viremia. Hepatology 2008 Oct;48(4):1079-1086.

(59) Carman W, Thomas H, Domingo E. Viral genetic variation: hepatitis B virus as a clinical example. Lancet 1993 Feb 6;341(8841):349-353.

(60) Reuman EC, Margeridon-Thermet S, Caudill HB, Liu T, Borroto-Esoda K, Svarovskaia ES, et al. A classification model for G-to-A hypermutation in hepatitis B virus ultra-deep pyrosequencing reads. Bioinformatics 2010 Dec 1;26(23):2929-2932.

(61) Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002 Aug 8;418(6898):646-650.

(62) Cullen BR. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. J Virol 2006 Feb;80(3):1067-1076.

(63) Domingo E, Gomez J. Quasispecies and its impact on viral hepatitis. Virus Res 2007 Aug;127(2):131-150.

(64) Hsu HY, Chang MH, Ni YH, Chen HL. Survey of hepatitis B surface variant infection in children 15 years after a nationwide vaccination programme in Taiwan. Gut 2004 Oct;53(10):1499-1503.

(65) Torre F, Cramp M, Owsianka A, Dornan E, Marsden H, Carman W, et al. Direct evidence that naturally occurring mutations within hepatitis B core epitope alter CD4+ T-cell reactivity. J Med Virol 2004 Mar;72(3):370-376.

(66) 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 Jul;24(1):32-37.

(67) Yu H, Yuan Q, Ge SX, Wang HY, Zhang YL, Chen QR, et al. Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". PLoS One 2010 Feb 19;5(2):e9297.

(68) Rodriguez-Frias F, Jardi R, Buti M, Schaper M, Hermosilla E, Valdes A, et al. Hepatitis B virus genotypes and G1896A precore mutation in 486 Spanish patients with acute and chronic HBV infection. J Viral Hepat 2006 May;13(5):343-350.

(69) Sumi H, Yokosuka O, Seki N, Arai M, Imazeki F, Kurihara T, et al. Influence of hepatitis B virus genotypes on the progression of chronic type B liver disease. Hepatology 2003 Jan;37(1):19-26.

(70) Buster EH, Flink HJ, Simsek H, Heathcote EJ, Sharmila S, Kitis GE, et al. Early HBeAg loss during peginterferon alpha-2b therapy predicts HBsAg loss: results of a long-term follow-up study in chronic hepatitis B patients. Am J Gastroenterol 2009 Oct;104(10):2449-2457.

(71) Wiersma ST, McMahon B, Pawlotsky JM, Thio CL, Thursz M, Lim SG, et al. Treatment of chronic hepatitis B virus infection in resource-constrained settings: expert panel consensus. Liver Int 2011 Jul;31(6):755-761.

(72) Hadziyannis SJ. Natural history of chronic hepatitis B in Euro-Mediterranean and African countries. J Hepatol 2011 Jul;55(1):183-191.

(73) Bertoletti A, Gehring AJ. The immune response during hepatitis B virus infection. J Gen Virol 2006 Jun;87(Pt 6):1439-1449.

(74) Kakimi K, Guidotti LG, Koezuka Y, Chisari FV. Natural killer T cell activation inhibits hepatitis B virus replication in vivo. J Exp Med 2000 Oct 2;192(7):921-930.

(75) Singh R, Kaul R, Kaul A, Khan K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. World J Gastroenterol 2007 Mar 28;13(12):1770-1787.

(76) Fong TL, Di Bisceglie AM, Biswas R, Waggoner JG, Wilson L, Claggett J, et al. High levels of viral replication during acute hepatitis B infection predict progression to chronicity. J Med Virol 1994 Jun;43(2):155-158.

(77) Thimme R, Wieland S, Steiger C, Ghrayeb J, Reimann KA, Purcell RH, et al. CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection. J Virol 2003 Jan;77(1):68-76.

(78) Wieland S, Thimme R, Purcell RH, Chisari FV. Genomic analysis of the host response to hepatitis B virus infection. Proc Natl Acad Sci U S A 2004 Apr 27;101(17):6669-6674.

(79) Guidotti LG, Rochford R, Chung J, Shapiro M, Purcell R, Chisari FV. Viral clearance without destruction of infected cells during acute HBV infection. Science 1999 Apr 30;284(5415):825-829.

(80) Moretta L, Bottino C, Pende D, Vitale M, Mingari MC, Moretta A. Human natural killer cells: Molecular mechanisms controlling NK cell activation and tumor cell lysis. Immunol Lett 2005 Aug 15;100(1):7-13.

(81) Koziel MJ. The immunopathogenesis of HBV infection. Antivir Ther 1998;3(Suppl 3):13-24.

(82) Nakamura I, Nupp JT, Cowlen M, Hall WC, Tennant BC, Casey JL, et al. Pathogenesis of experimental neonatal woodchuck hepatitis virus infection: chronicity as an outcome of infection is associated with a

diminished acute hepatitis that is temporally deficient for the expression of interferon gamma and tumor necrosis factor-alpha messenger RNAs. Hepatology 2001 Feb;33(2):439-447.

(83) Webster GJ, Reignat S, Maini MK, Whalley SA, Ogg GS, King A, et al. Incubation phase of acute hepatitis B in man: dynamic of cellular immune mechanisms. Hepatology 2000 Nov;32(5):1117-1124.

(84) Guidotti LG, Chisari FV. Noncytolytic control of viral infections by the innate and adaptive immune response. Annu Rev Immunol 2001;19:65-91.

(85) Kalams SA, Walker BD. The critical need for CD4 help in maintaining effective cytotoxic T lymphocyte responses. J Exp Med 1998 Dec 21;188(12):2199-2204.

(86) Ciurea A, Hunziker L, Martinic MM, Oxenius A, Hengartner H, Zinkernagel RM. CD4+ T-cell-epitope escape mutant virus selected in vivo. Nat Med 2001 Jul;7(7):795-800.

(87) Iloeje UH, Yang HI, Su J, Jen CL, You SL, Chen CJ, et al. Predicting cirrhosis risk based on the level of circulating hepatitis B viral load. Gastroenterology 2006 Mar;130(3):678-686.

(88) European Association For The Study Of The Liver. EASL Clinical Practice Guidelines: management of chronic hepatitis B. J Hepatol 2009 Feb;50(2):227-242.

(89) European Association For The Study Of The Liver. EASL Clinical Practice Guidelines: management of chronic hepatitis B virus infection. J Hepatol 2012.

(90) Chang TT, Lai CL, Chien RN, Guan R, Lim SG, Lee CM, et al. Four years of lamivudine treatment in Chinese patients with chronic hepatitis B. J Gastroenterol Hepatol 2004 Nov;19(11):1276-1282.

(91) Hadziyannis SJ, Tassopoulos NC, Heathcote EJ, Chang TT, Kitis G, Rizzetto M, et al. Long-term therapy with adefovir dipivoxil for HBeAg-negative chronic hepatitis B for up to 5 years. Gastroenterology 2006 Dec;131(6):1743-1751.

(92) Liaw YF, Gane E, Leung N, Zeuzem S, Wang Y, Lai CL, et al. 2-Year GLOBE trial results: telbivudine Is superior to Lamivudine in patients with chronic hepatitis B. Gastroenterology 2009 Feb;136(2):486-495.

(93) Tenney DJ, Rose RE, Baldick CJ, Pokornowski KA, Eggers BJ, Fang J, et al. Long-term monitoring shows hepatitis B virus resistance to entecavir in nucleoside-naive patients is rare through 5 years of therapy. Hepatology 2009 May;49(5):1503-1514.

(94) Snow-Lampart A, Chappell B, Curtis M, Zhu Y, Myrick F, Schawalder J, et al. No resistance to tenofovir disoproxil fumarate detected after up to 144 weeks of therapy in patients monoinfected with chronic hepatitis B virus. Hepatology 2011 Mar;53(3):763-773.

(95) Bartholomeusz A, Locarnini SA. Antiviral drug resistance: clinical consequences and molecular aspects. Semin Liver Dis 2006 May;26(2):162-170.

(96) Tipples GA, Ma MM, Fischer KP, Bain VG, Kneteman NM, Tyrrell DL. Mutation in HBV RNA-dependent DNA polymerase confers resistance to Lamivudine in vivo. Hepatology 1996 Sep;24(3):714-717.

(97) Ono-Nita SK, Kato N, Shiratori Y, Masaki T, Lan KH, Carrilho FJ, et al. YMDD motif in hepatitis B virus DNA polymerase influences on replication and Lamivudine resistance: A study by in vitro full-length viral DNA transfection. Hepatology 1999 Mar;29(3):939-945.

(98) Lai CL, Leung N, Teo EK, Tong M, Wong F, Hann HW, et al. A 1-year trial of telbivudine, Lamivudine, and the combination in patients with hepatitis B e antigen-positive chronic hepatitis B. Gastroenterology 2005 Aug;129(2):528-536.

(99) Tenney DJ, Levine SM, Rose RE, Walsh AW, Weinheimer SP, Discotto L, et al. Clinical emergence of entecavir-resistant hepatitis B virus requires additional substitutions in virus already resistant to Lamivudine. Antimicrob Agents Chemother 2004 Sep;48(9):3498-3507.

(100) Angus P, Vaughan R, Xiong S, Yang H, Delaney W, Gibbs C, et al. Resistance to adefovir dipivoxil therapy associated with the selection of a novel mutation in the HBV polymerase. Gastroenterology 2003 Aug;125(2):292-297.

(101) Yeon JE, Yoo W, Hong SP, Chang YJ, Yu SK, Kim JH, et al. Resistance to adefovir dipivoxil in Lamivudine resistant chronic hepatitis B patients treated with adefovir dipivoxil. Gut 2006 Oct;55(10):1488-1495.

(102) Reijnders JG, Deterding K, Petersen J, Zoulim F, Santantonio T, Buti M, et al. Antiviral effect of entecavir in chronic hepatitis B: influence of prior exposure to nucleos(t)ide analogues. J Hepatol 2010 Apr;52(4):493-500.

(103) Levine S, Hernandez D, Yamanaka G, Zhang S, Rose R, Weinheimer S, et al. Efficacies of entecavir against Lamivudine-resistant hepatitis B virus replication and recombinant polymerases in vitro. Antimicrob Agents Chemother 2002 Aug;46(8):2525-2532.

(104) Lok AS, Hussain M, Cursano C, Margotti M, Gramenzi A, Grazi GL, et al. Evolution of hepatitis B virus polymerase gene mutations in hepatitis B e antigen-negative patients receiving Lamivudine therapy. Hepatology 2000 Nov;32(5):1145-1153.

(105) Cho SW, Hahm KB, Kim JH. Reversion from precore/core promoter mutants to wild-type hepatitis B virus during the course of Lamivudine therapy. Hepatology 2000 Nov;32(5):1163-1169.

(106) Kuwahara R, Kumashiro R, Murashima S, Ogata K, Tanaka K, Hisamochi A, et al. Genetic heterogeneity of the precore and the core promoter region of genotype C hepatitis B virus during Lamivudine therapy. J Med Virol 2004 Jan;72(1):26-34.

(107) Lim SG, Cheng Y, Guindon S, Seet BL, Lee LY, Hu P, et al. Viral quasi-species evolution during hepatitis Be antigen seroconversion. Gastroenterology 2007 Sep;133(3):951-958.

(108) Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975 May 25;94(3):441-448.

(109) Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A 1977 Feb;74(2):560-564.

(110) Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005 Sep 15;437(7057):376-380.

(111) Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. Science 1998 Jul 17;281(5375):363, 365.

(112) Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008 Nov 6;456(7218):53-59.

(113) Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A 2003 Jul 22;100(15):8817-8822.

(114) Jung MC, Diepolder HM, Spengler U, Wierenga EA, Zachoval R, Hoffmann RM, et al. Activation of a heterogeneous hepatitis B (HB) core and e antigen-specific CD4+ T-cell population during seroconversion to anti-HBe and anti-HBs in hepatitis B virus infection. J Virol 1995 Jun;69(6):3358-3368.

(115) van de Klundert MA, Cremer J, Kootstra NA, Boot HJ, Zaaijer HL. Comparison of the hepatitis B virus core, surface and polymerase gene substitution rates in chronically infected patients. J Viral Hepat 2012 Feb;19(2):e34-40.

(116) Ehata T, Omata M, Yokosuka O, Hosoda K, Ohto M. Variations in codons 84-101 in the core nucleotide sequence correlate with hepatocellular injury in chronic hepatitis B virus infection. J Clin Invest 1992 Jan;89(1):332-338.

(117) Preikschat P, Gunther S, Reinhold S, Will H, Budde K, Neumayer HH, et al. Complex HBV populations with mutations in core promoter, C gene, and pre-S region are associated with development of cirrhosis in long-term renal transplant recipients. Hepatology 2002 Feb;35(2):466-477.

(118) Sung FY, Jung CM, Wu CF, Lin CL, Liu CJ, Liaw YF, et al. Hepatitis B virus core variants modify natural course of viral infection and hepatocellular carcinoma progression. Gastroenterology 2009 Nov;137(5):1687-1697.

(119) Radecke K, Protzer U, Trippler M, Meyer Zum Buschenfelde KH, Gerken G. Selection of hepatitis B virus variants with aminoacid substitutions inside the core antigen during interferon-alpha therapy. J Med Virol 2000 Dec;62(4):479-486.

(120) Chen RY, Bowden S, Desmond PV, Dean J, Locarnini SA. Effects of interferon alpha therapy on the catalytic domains of the polymerase gene and basal core promoter, precore and core regions of hepatitis B virus. J Gastroenterol Hepatol 2003 Jun;18(6):630-637.

(121) Boni C, Penna A, Bertoletti A, Lamonaca V, Rapti I, Missale G, et al. Transient restoration of anti-viral T cell responses induced by Lamivudine therapy in chronic hepatitis B. J Hepatol 2003 Oct;39(4):595-605.

(122) Chain BM, Myers R. Variability and conservation in hepatitis B virus core protein. BMC Microbiol 2005 May 27;5:33.

(123) Chien RN, Yeh CT, Tsai SL, Chu CM, Liaw YF. Determinants for sustained HBeAg response to Lamivudine therapy. Hepatology 2003 Nov;38(5):1267-1273.

(124) Homs M, Buti M, Quer J, Jardi R, Schaper M, Tabernero D, 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 Jul 8.

(125) Gerelsaikhan T, Tavis JE, Bruss V. Hepatitis B virus nucleocapsid envelopment does not occur without genomic DNA synthesis. J Virol 1996 Jul;70(7):4269-4274.

(126) Homs M, Jardi R, Buti M, Schaper M, Tabernero D, Fernandez-Fernandez P, et al. HBV core region variability: effect of antiviral treatments on main epitopic regions. Antivir Ther 2011;16(1):37-49.

(127) Tang YZ, Liu L, Pan MM, Wang YM, Deng GH. Evolutionary pattern of full hepatitis B virus genome during sequential nucleos(t)ide analog therapy. Antiviral Res 2011 Jun;90(3):116-125.

(128) Martell M, Esteban JI, Quer J, Genesca J, Weiner A, Esteban R, et al. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J Virol 1992 May;66(5):3225-3229.

(129) Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, et al. Multiplex amplification of large sets of human exons. Nat Methods 2007 Nov;4(11):931-936.

(130) Hormozdiari F, Alkan C, Eichler EE, Sahinalp SC. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. Genome Res 2009 Jul;19(7):1270-1278.

(131) Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci U S A 2008 Sep 2;105(35):13081-13086.

(132) Siqueira JF,Jr, Fouad AF, Rocas IN. Pyrosequencing as a tool for better understanding of human microbiomes. J Oral Microbiol 2012;4:10.3402/jom.v4i0.10743. Epub 2012 Jan 23.

(133) Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A 2010 Apr 20;107(16):7503-7508.

(134) Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res 2007 Aug;17(8):1195-1201.

(135) Liang B, Luo M, Scott-Herridge J, Semeniuk C, Mendoza M, Capina R, et al. A comparison of parallel pyrosequencing and sanger clone-based sequencing and its impact on the characterization of the genetic diversity of HIV-1. PLoS One 2011;6(10):e26745.

(136) Wang GP, Sherrill-Mix SA, Chang KM, Quince C, Bushman FD. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J Virol 2010 Jun;84(12):6218-6228.

(137) Cheng S, Fockler C, Barnes WM, Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Natl Acad Sci U S A 1994 Jun 7;91(12):5695-5699.

(138) Chen RY, Edwards R, Shaw T, Colledge D, Delaney WE,4th, Isom H, et al. Effect of the G1896A precore mutation on drug sensitivity and replication yield of Lamivudine-resistant HBV in vitro. Hepatology 2003 Jan;37(1):27-35.

(139) Tacke F, Gehrke C, Luedde T, Heim A, Manns MP, Trautwein C. Basal core promoter and precore mutations in the hepatitis B virus genome enhance replication efficacy of Lamivudine-resistant mutants. J Virol 2004 Aug;78(16):8524-8535.

(140) Ciancio A, Smedile A, Rizzetto M, Lagget M, Gerin J, Korba B. Identification of HBV DNA sequences that are predictive of response to Lamivudine therapy. Hepatology 2004 Jan;39(1):64-73.

(141) Kim S, Wang H, Ryu WS. Incorporation of eukaryotic translation initiation factor eIF4E into viral nucleocapsids via interaction with hepatitis B virus polymerase. J Virol 2010 Jan;84(1):52-58.

(142) Hu J, Boyer M. Hepatitis B virus reverse transcriptase and epsilon RNA sequences required for specific interaction in vitro. J Virol 2006 Mar;80(5):2141-2150.

(143) Girard FC, Ottink OM, Ampt KA, Tessari M, Wijmenga SS. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. Nucleic Acids Res 2007;35(8):2800-2811.

(144) Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 2010 Mar 15;11:129.

(145) Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A 2004 May 11;101(19):7287-7292.

(146) Ampt KA, van der Werf RM, Nelissen FH, Tessari M, Wijmenga SS. The unstable part of the apical stem of duck hepatitis B virus epsilon shows enhanced base pair opening but not pico- to nanosecond dynamics and is essential for reverse transcriptase binding. Biochemistry 2009 Nov 10;48(44):10499-10508.

(147) Petzold K, Duchardt E, Flodell S, Larsson G, Kidd-Ljunggren K, Wijmenga S, et al. Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. Nucleic Acids Res 2007;35(20):6854-6861.

(148) Guarnieri M, Kim KH, Bang G, Li J, Zhou Y, Tang X, et al. Point mutations upstream of hepatitis B virus core gene affect DNA replication at the step of core protein expression. J Virol 2006 Jan;80(2):587-595.

(149) Abraham TM, Loeb DD. The topology of hepatitis B virus pregenomic RNA promotes its replication. J Virol 2007 Nov;81(21):11577-11584.

(150) Naoumov NV, Thomas MG, Mason AL, Chokshi S, Bodicky CJ, Farzaneh F, et al. Genomic variations in the hepatitis B core gene: a possible factor influencing response to interferon alfa treatment. Gastroenterology 1995 Feb;108(2):505-514.

(151) Pumpens P, Grens E, Nassal M. Molecular epidemiology and immunology of hepatitis B virus infection - an update. Intervirology 2002;45(4-6):218-232.

(152) Fattovich G, Giustina G, Christensen E, Pantalena M, Zagni I, Realdi G, et al. Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type B. The European Concerted Action on Viral Hepatitis (Eurohep). Gut 2000 Mar;46(3):420-426.

(153) Bertoletti A, Ferrari C, Fiaccadori F, Penna A, Margolskee R, Schlicht HJ, et al. HLA class I-restricted human cytotoxic T cells recognize endogenously synthesized hepatitis B virus nucleocapsid antigen. Proc Natl Acad Sci U S A 1991 Dec 1;88(23):10445-10449.

(154) Sendi H, Mehrab-Mohseni M, Shahraz S, Norder H, Alavian SM, Noorinayer B, et al. CTL escape mutations of core protein are more frequent in strains of HBeAg negative patients with low levels of HBV DNA. J Clin Virol 2009 Nov;46(3):259-264.

(155) Tanaka H, Ueda H, Hamagami H, Yukawa S, Ichinose M, Miyano M, et al. Mutations in hepatitis B virus core regions correlate with hepatocellular injury in Chinese patients with chronic hepatitis B. World J Gastroenterol 2005 Aug 14;11(30):4693-4696.

(156) Mohamadkhani A, Jazii FR, Poustchi H, Nouraein O, Abbasi S, Sotoudeh M, et al. The role of mutations in core protein of hepatitis B virus in liver fibrosis. Virol J 2009 Nov 26;6:209.

(157) Perrillo R. Benefits and risks of interferon therapy for hepatitis B. Hepatology 2009 May;49(5 Suppl):S103-11.

(158) Alexopoulou A, Karayiannis P, Hadziyannis SJ, Aiba N, Thomas HC. Emergence and selection of HBV variants in an anti-HBe positive patient persistently infected with quasi-species. J Hepatol 1997 Apr;26(4):748-753.

(159) Abbott WG, Tsai P, Leung E, Trevarton A, Ofanoa M, Hornell J, et al. Associations between HLA class I alleles and escape mutations in the hepatitis B virus core gene in New Zealand-resident Tongans. J Virol 2010 Jan;84(1):621-629.

(160) Harrison A, Lemey P, Hurles M, Moyes C, Horn S, Pryor J, et al. Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. Viruses 2011 Feb;3(2):83-101.

(161) Suzuki F, Suzuki Y, Tsubota A, Akuta N, Someya T, Kobayashi M, et al. Mutations of polymerase, precore and core promoter gene in hepatitis B virus during 5-year Lamivudine therapy. J Hepatol 2002 Dec;37(6):824-830.

(162) Alexopoulou A, Baltayiannis G, Eroglu C, Nastos T, Dourakis SP, Archimandritis AJ, et al. Core mutations in patients with acute episodes of chronic HBV infection are associated with the emergence of new immune recognition sites and the development of high IgM anti-HBc index values. J Med Virol 2009 Jan;81(1):34-41.

(163) Pairan A, Bruss V. Functional surfaces of the hepatitis B virus capsid. J Virol 2009 Nov;83(22):11616-11623.

(164) Ponsel D, Bruss V. Mapping of amino acid side chains on the surface of hepatitis B virus capsids required for envelopment and virion formation. J Virol 2003 Jan;77(1):416-422.

(165) Deguchi M, Yamashita N, Kagita M, Asari S, Iwatani Y, Tsuchida T, et al. Quantitation of hepatitis B surface antigen by an automated chemiluminescent microparticle immunoassay. J Virol Methods 2004 Feb;115(2):217-222.

(166) Wiegand J, Wedemeyer H, Finger A, Heidrich B, Rosenau J, Michel G, et al. A decline in hepatitis B virus surface antigen (hbsag) predicts clearance, but does not correlate with quantitative hbeag or HBV DNA levels. Antivir Ther 2008;13(4):547-554.

(167) Thompson AJ, Nguyen T, Iser D, Ayres A, Jackson K, Littlejohn M, et al. Serum hepatitis B surface antigen and hepatitis B e antigen titers: disease phase influences correlation with viral load and intrahepatic hepatitis B virus markers. Hepatology 2010 Jun;51(6):1933-1944.

(168) Tu H, Bonura C, Giannini C, Mouly H, Soussan P, Kew M, et al. Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. Cancer Res 2001 Nov 1;61(21):7803-7810.

(169) Ma NF, Lau SH, Hu L, Xie D, Wu J, Yang J, et al. COOH-terminal truncated HBV X protein plays key role in hepatocarcinogenesis. Clin Cancer Res 2008 Aug 15;14(16):5061-5068.