

# Spacer Domain in Hepatitis B Virus Polymerase: Plugging a Hole or Performing a Role?

Caitlin Pley,<sup>a,b</sup> José Lourenço,<sup>c,d</sup> Anna L. McNaughton,<sup>e,f</sup> Philippa C. Matthews<sup>f,g,h</sup>

<sup>a</sup>School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

<sup>b</sup>Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

<sup>c</sup>Department of Zoology, University of Oxford, Oxford, United Kingdom

<sup>d</sup>Biosystems and Integrative Sciences Institute, University of Lisbon, Lisbon, Portugal

<sup>e</sup>Population Health Science, Bristol Medical School, University of Bristol, Bristol, United Kingdom

<sup>f</sup>Nuffield Department of Medicine, University of Oxford Medawar Building, Oxford, United Kingdom

<sup>g</sup>The Francis Crick Institute, London, United Kingdom

<sup>h</sup>Division of Infection and Immunity, University College London, London, United Kingdom

**ABSTRACT** Hepatitis B virus (HBV) polymerase is divided into terminal protein, spacer, reverse transcriptase, and RNase domains. Spacer has previously been considered dispensable, merely acting as a tether between other domains or providing plasticity to accommodate deletions and mutations. We explore evidence for the role of spacer sequence, structure, and function in HBV evolution and lineage, consider its associations with escape from drugs, vaccines, and immune responses, and review its potential impacts on disease outcomes.

**KEYWORDS** hepatitis B virus, HBV, spacer, evolution, diversity, polymorphism, polymerase, genotype, phylogeny

Hepatitis B virus (HBV) is a hepadnavirus responsible for 300 million chronic infections worldwide (1). A better understanding of HBV biology can provide improved insights into viral diversity and epidemiology and underpin refined approaches to treatment and prevention. The tiny genome (~3.2 kb) is organized into four overlapping open reading frames (ORFs) (Fig. 1A), compressing a high density of genetic information (2–4). The HBV replication cycle is dependent on the viral polymerase protein (P), a polyfunctional protein composed of 4 subdomains: terminal protein (TP), spacer, reverse transcriptase (RT), and RNase H (Fig. 1B).

The spacer domain spans amino acids (aa) 184 to 348 of P (based on GenBank reference sequence [X02763](#)), sandwiched between TP and RT (5). Spacer has long been considered a dispensable subdomain, and the structure and function of this region therefore has not been investigated in detail. However, the question arises as to why a conserved spacer domain would have evolved and persisted if it had no specific function. More recent evidence is accumulating to suggest that spacer has a role in the HBV replication cycle, is subject to positive selection (as a result of evasion of therapy, vaccines, and host immune responses), is relevant for distinguishing between viral lineages, and can be associated with different outcomes of HBV infection.

Here, we synthesize evidence for the roles and significance of spacer and highlight gaps in our knowledge about this small but potentially influential region. We consider evidence from the older literature and assess how new insights can contribute to a refined understanding of spacer biology with practical relevance to pathophysiology, genotyping, determination of drug resistance, and sequence-based clinical stratification.

**Editor** Hector C. Aguilar, Cornell University

**Copyright** © 2022 Pley et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

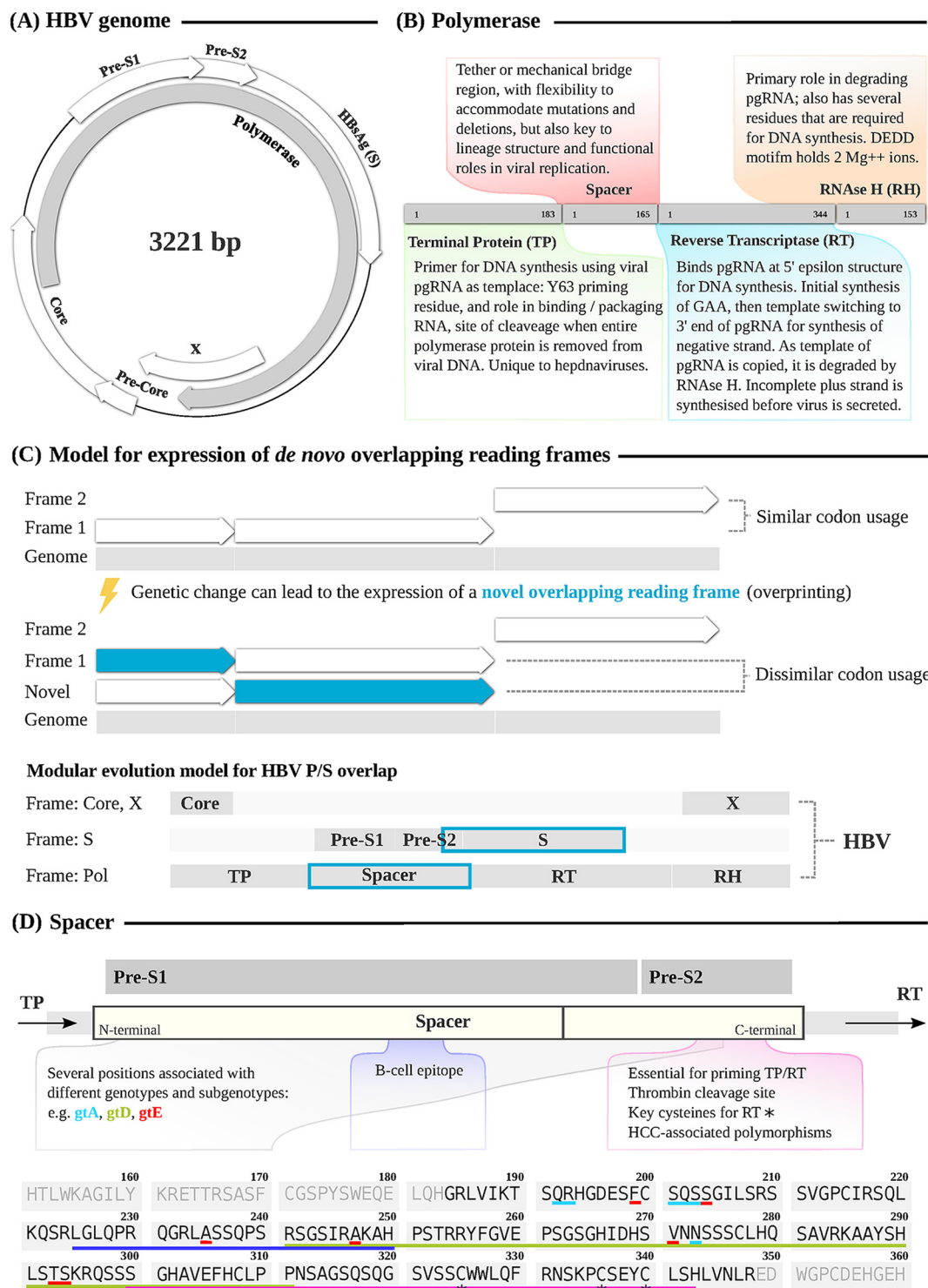
Address correspondence to Philippa C. Matthews, [philippa.matthews@crick.ac.uk](mailto:philippa.matthews@crick.ac.uk).

The authors declare no conflict of interest.

**Received** 10 January 2022

**Accepted** 14 March 2022

**Published** 12 April 2022



**FIG 1** Genomic and functional roles of HBV spacer domain. (A) Structure of overlapping ORFs in the HBV 3.2-kb partially double-stranded circular genome. The entire length of surface (S) (400 aa) is encoded on an alternate frame within the length of Polymerase (P), making it the longest known overlap of any animal virus. (B) HBV P polypeptide, showing subdomains TP (terminal protein), spacer, RT (reverse transcriptase), and RNase H. (C) Model to illustrate overlapping reading frames (ORF) with different codon usage by overlapping genes and modular evolution theory due to imprinting. (D) Annotation of spacer domain (165 aa), illustrating examples of sequence polymorphisms associated with lineage (genotype), enzymatic function, and key functional outcomes. Deletions are best tolerated in the N-terminal 2/3 of the protein, while the C-terminal 1/3 contains the majority of residues with established functional roles. Numbering and sequence are based on genotype A reference genome accession no. [X02763](#). Cysteine residues in the putative zinc finger DNA-binding motif are marked with a star at aa 325, 336, and 340 (equating to previously described positions 312, 323, and 327 in a genotype D strain). Coloured underlining of the sequence matches the annotation in the schematic lay-out above, as follows: turquoise, gtA; green, gtD; red, gtE; purple, B cell epitope; pink, key regions in C-terminus.

## DEFINITIONS AND TERMINOLOGY

The following terms are used here: adaptive evolution, changes in the frequency of beneficial variants and of deleterious variants due to selection;  $dN/dS$  ratio, the ratio of observed substitution rates of nonsynonymous versus synonymous genetic changes ( $dN/dS > 1$  is generally interpreted as a proxy for the past occurrence of positive selection); negative selection, change in the frequency of variants within a population that may be harmful to fitness of the organism; neutral selection, stable frequency of variants within a population that have no effect on the fitness of the organism; nonsynonymous mutations, changes in the nucleotide sequence that result in changes in the amino acid sequence of a protein or introduce a premature stop codon; occult HBV, detectable HBV DNA in the absence of HBV surface antigen (HBsAg) in the serum; open reading frame (ORF), portions of DNA/RNA that contain no stop codons and can be translated to form amino acid sequences; phylogeny, the evolutionary history of a group of organisms or samples of the same organism; positive selection, changes in the frequency of variants within a population that may be beneficial to the fitness of the organism, e.g., replicative capacity, persistence, or transmissibility; synonymous mutations, changes in the nucleotide sequence that cause no change in the amino acids sequence of the protein due to codon redundancy.

## EVOLUTION OF HBV SPACER

Due to the intimate relationship between a pathogen's genetic sequence and protein structure and function, here we consider the evolutionary pressures that may be relevant in driving sequence change in spacer and review the genotype specificity of observed polymorphisms.

**The spacer domain evolved through modular evolution.** Overlapping ORFs typically consist of an ancestral gene (encoding essential proteins) and a *de novo* gene that evolved subsequently through overprinting (6, 7), encoding accessory protein(s) (8, 9). A sequence analysis of 43 genera of RNA viruses infecting eukaryotes showed that most proteins created *de novo* are accessory proteins and are predicted to be fully disordered, similar to the spacer region of HBV P (8).

Methods developed for deltaretroviruses show that codon usage can distinguish ancestral genes (which have codon usage similar to nonoverlapping regions of the viral genome) from *de novo* genes (with codon usage very different from the rest of the genome) with high specificity (6, 10). Within HBV genomes, the entire length of surface (S) (400 aa) is on an alternate frame overlapping P (Fig. 1A) (11). However, analyses of P and S have been inconclusive in determining which of the two genes is ancestral, as both are essential for virus survival (11). Comparing the codon usage of the overlapping P and S ORFs has shown that codon usage of the entire overlapping frame ( $>1,000$  nucleotides [nt]) compared to the nonoverlapping region is not significantly different for P and S (11). However, under a sliding window model, two regions of different codon usage can be identified, one in the 5' third of the overlap and another in the 3' two-thirds; these regions are conserved across hepadnavirus genomes (11). This suggests a modular evolution model for the P/S overlap, according to which the PreS1 domain, most of PreS2, and the RT domain of P are ancestral, while the spacer domain, the C-terminal third of PreS2, and the S domain evolved *de novo* by overprinting (Fig. 1C). This theory is consistent with an understanding of the functions of these domains, with essential roles of PreS1 in infectivity (12, 13) and RT in replication (14).

This proposed model of the primordial structure of the HBV genome may also explain why the P polyprotein contains RT and RNase H domains with retroviral homologues (15) but differs in having TP and a spacer domain. Congruent with the theory that spacer evolved later than other P domains, it is thought to have an intrinsically disordered structure (8, 16) (also see "Structure and Function," below).

**Spacer is under positive selection.** It is generally understood that positive selection in one reading frame should be mirrored by negative selection in the overlapping frame (7, 17–20). Accordingly, calculation of the nonsynonymous to synonymous

mutation ratio ( $dN/dS$ ) of the S, core (C), and P genes of HBV can show that while S holds evidence of positive selection, C and P undergo negative selection (21). However, positive selection in S does not necessitate relaxed or negative selection in P; the frameshift between the two genes allows them to evolve independently, accommodating different evolutionary pressures (22).

Findings that  $dN/dS$  is  $>1$  in significant segments of either P or S in the overlapping region and  $>1$  in both genes in only a few key regions indicate that selection pressures fluctuate throughout the region (22). In the P/S overlap, the S ORF is shifted by 1 nt, meaning that the first codon position of P (p1) overlaps the third codon position of S (s3). Degeneracy within the nucleotide code means that changes to the first and second codon positions more frequently result in amino acid changes. Thus, adaptive evolution in P occurs through p1/s3 substitutions that cause an amino acid change in P but rarely in S, and adaptive evolution in S occurs through p3/s2 substitutions that cause an amino acid change in S but infrequently in P. Substitutions resulting in an amino acid change in both genes (p2/s1) are rare and contribute little to entropy across the sequence (22). Thus, while P and S genes are overall both under negative selection, to preserve their important roles in infectivity and replication, parts of both (particularly PreS and spacer) are subject to positive selection to evade the host immune system as well as to withstand anthropogenic pressures, such as vaccine-induced antibodies and antivirals (22) (discussed further in "Host-virus interaction and outcomes of infection," below).

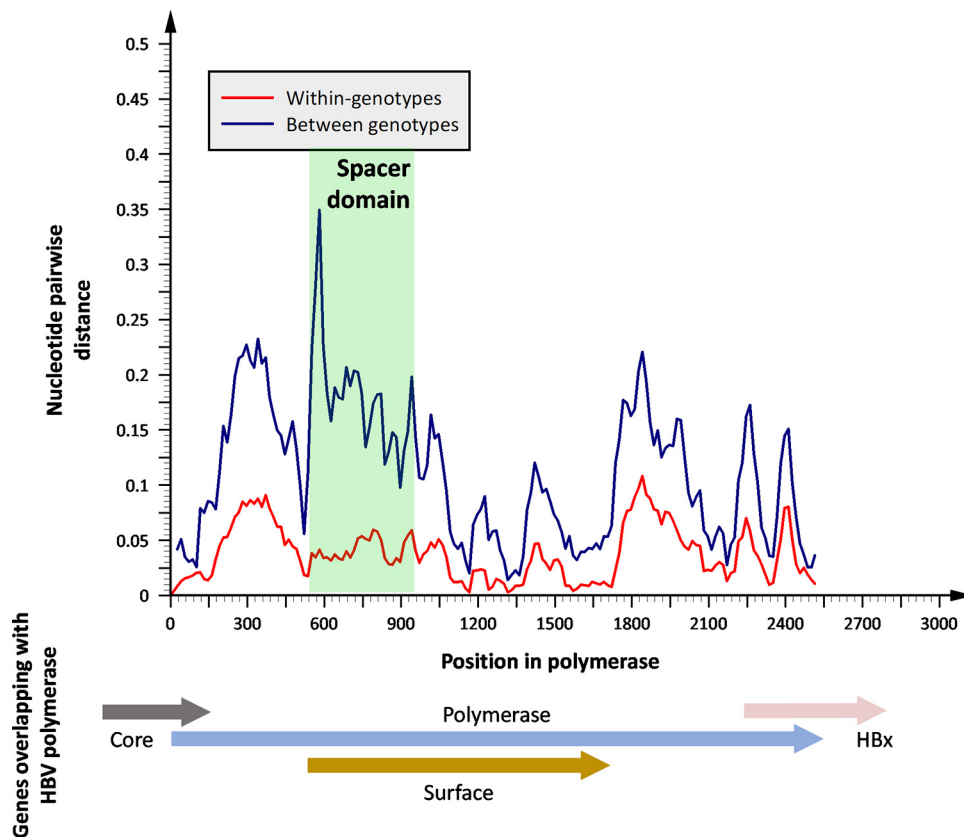
Although a high  $dN/dS$  ratio is observed in spacer compared to other domains in P and S, this does not conclusively determine whether spacer is undergoing true positive selection or is merely tolerant of nonsynonymous mutation (relaxed selection) (23). The variant degree (VD), a parameter used to evaluate the evolutionary selection of a sequence, is thought to negatively correlate with the importance of biological function. The VD of spacer and PreS2 of  $>20\%$  contrasts with PreS1, S, and RT VD values of  $<10\%$  (23), increasing the likelihood that spacer and PreS2 are undergoing relaxed selection (23). At the same time, PreS1, S, and RT are under stricter selection due to their critical roles in hepatocyte binding, cell entry (12, 13, 24–26), and genome replication and packaging, respectively (14, 27, 28).

**Spacer reflects HBV sequence diversity and lineage.** The spacer domain shows a high degree of nucleotide variation between different hepadnaviruses and between genotypes and subgenotypes of HBV (29–33), indicating that despite being evolutionarily newer than other regions of the genome, this domain has undergone extensive divergence over time (32, 33). This striking diversity may be enabled by the domain's disordered structure (16).

Pairwise analysis of nucleotide sequence space in P indicates that spacer holds low within-genotype diversity but high between-genotype diversity, contrasting with other regions of the protein (Fig. 2). If spacer is an inherently plastic region where mutations can accumulate without significantly affecting virus viability, one might expect variants to be stochastically distributed across phylogenies. However, there is accumulating evidence that spacer polymorphisms and deletions are sufficient to distinguish between HBV lineages (Table 1, Fig. 1D). Due to the frameshift between the P and S reading frames, the majority of nucleotide substitutions that result in amino acid changes in spacer will be synonymous in PreS1/2. Therefore, this phenomenon may not be fully explained through the well-recognized selection pressure acting on S (further discussed in "Genetic plasticity of HBV spacer," below), and there may be incompletely understood selection pressures acting on spacer that result in lineage-specific motifs.

## GENETIC PLASTICITY OF HBV SPACER

**Spacer can accommodate insertions and deletions without functional impact.** P can accommodate large insertions and deletions in spacer and can maintain catalytic function, including protein priming, synthesis of the DNA minus strand, removing pre-genomic RNA (pgRNA), and synthesizing plus-strand DNA (34, 35). In one of the first investigations of spacer function, 52 aa were deleted from duck HBV (DHBV) spacer with endogenous polymerase activity largely unaffected and without a quantifiable



**FIG 2** Nucleotide pairwise distances within and between genotypes in HBV polymerase. Representative HBV genotype sequences for genotypes A to I (129) were analyzed using a sliding window size of 100 bp and increments of 25 bp in SSE v1.3 (130). Putative genotype J was excluded, as regions of the genotype are thought to originate from nonhuman HBV sequences.

impact on DNA synthesis (5). Subsequent experiments have deleted large regions of spacer (e.g., amino acids 201 to 292) without affecting endogenous polymerase activity, with only residues 293 to 335 required to maintain enzymatic function (14). Furthermore, naturally occurring spacer variants with large in-frame deletions have replication competence similar to that of wild-type strains (36–38). For example, a sequence with a 69-amino-acid deletion in PreS1/spacer had wild-type (WT)-like polymerase activity, substantiating previous findings that the N-terminal portion of spacer is not essential for replication (36). Similarly, large insertions in spacer have been tolerated without an impact on DNA synthesis activity and RNA packaging capacity (33). Based on this evidence that spacer is capable of withstanding insertion, deletion, and substitution mutations without an impact on viral replication capacity, it is possible to conclude that spacer is a nonessential, or even entirely dispensable, subdomain of P that merely exists as a mechanical tether between TP and RT (5, 14, 23, 27, 33–35, 39–43). Table 2 provides a summary of spacer mutagenesis studies.

#### **Spacer tolerates polymorphisms to accommodate positive selection in PreS.**

Large parts of the S gene undergo positive selection to evade host immune responses, in contrast to the TP, RT, and RNase H domains, which are under negative selection to preserve important replicative functions (21). Most of the variability in S lies in the PreS domains, the “a” (major antigenic) determinant, and the C terminus (44). Positive selection in S occurs almost exclusively in known T and B cell epitopes, suggesting adaptive evolution to evade natural or vaccine-mediated immune responses (44) (also see “Host-virus interaction and outcomes of infection,” below). Due to the frameshift between the overlapping P and S reading frames, many nonsynonymous mutations in PreS (p3s2) result in synonymous mutations in spacer. Thus, spacer may mediate the



**TABLE 1** Examples of evidence that spacer residues can be used to distinguish between HBV genotypes, subgenotypes, and geographic distribution of lineages<sup>a</sup>

HBV genotype	Location	Observation(s)	Reference
A, subgenotype A1	Brazil	Four amino acids in the polymerase, of which two (Pro18 and His90) are in spacer, and one in the core antigen were sufficient to determine sequence clustering with isolates from Asia but not with A1 isolates from Southern and East Africa. These five amino acids supported Bayesian analysis, concluding that the Brazilian isolates formed part of the Asian-American clade of subgenotype A1 (posterior probability value, 0.996).	131
A, subgenotype A1	Zimbabwe	Characteristic shared residues in PreS1, PreS2, and spacer associated with clustering of sequences as part of the South African A1 clade. Spacer was relatively well conserved among the Zimbabwean isolates. Two spacer motifs (Ser9Gln10 and Glu18Ser19Phe20) could distinguish Zimbabwean isolates from clades found in other African and Asian countries where HBV/A1 is endemic but were shared with the South African HBV/A1 clade.	132
B and C	China	Genotype-specific mutations that distinguished genotypes B and C were found to be nonrandomly clustered in spacer. 56 of the 85 genotype-specific mutations were distributed in the 5' half of the P/S overlap, of which 42 mutations led to amino acid changes in spacer, and 23 mutations caused amino acid changes in PreS. Only 29 mutations were present in the 3' half of the overlapping reading frame, of which 13 caused residue changes in RT and 18 caused residue changes in S.	23
D, subgenotype D1/D4	Russia and the Baltic	Amino acids 58–128 of spacer contain subgenotype-specific mutations capable of reliably differentiating between D1 and D4 isolates.	133
E	West Africa	Despite low intragenotype diversity of genotype E, the spacer region contained eight unique amino acid residues (Glu16, His21, Arg52, Asp55, Met64, Lys88, Asn110, and His111).	134
Multiple	All	A conserved genotype-specific signature motif was also discovered in PreS1. Deletions at the start of spacer (predicted to start at nt 2856 in the <a href="#">X02763</a> reference sequence) are highly conserved within genotypes. Known conserved deletions in PreS1, $\Delta$ 2861–3 in genotypes E and G, and $\Delta$ 2854–2886 in genotypes D and J result in deletions at the start of spacer.	129

<sup>a</sup>Locations of the residues are given by numbering the start of the spacer domain as the first amino acid.

conflict between diversifying and constraining forces in S and P, respectively. However, most positive selection sites in P are concentrated in spacer, indicating that spacer is not undergoing merely neutral selection or random evolution but may in fact be important for the HBV replication cycle beyond a role facilitating changes in PreS (44).

## STRUCTURE AND FUNCTION

**Spacer is characterized by a disordered protein structure incorporating protease digestion sites.** Spacer's secondary and tertiary protein structure has not been elucidated, and there is a growing recognition that spacer is likely to be an intrinsically disordered protein region (IDPR) (16), together with overlapping pre-S regions with which it overlaps (45). IDPRs can deliver important biological roles even in the absence of a stable physical structure; indeed, conformational flexibility may itself be a functional attribute, allowing them to perform “impossible tricks” that would not be possible for a molecule with a stable, rigid structure (46). Such regions may have specific characteristics (for example, in diversity, charge, and polarity) and are common in viruses, with potentially key roles in protein-protein interaction (PPI) networks (47, 48).

Crystal structures have not been solved for HBV P, with challenges including the multidomain nature of the polyprotein, the possibility of more than one structural configuration, and its hybrid nature incorporating regions that are structurally ordered (such as RT), combined with IDPRs (46, 49). In addition to spatial flexibility, spacer's disordered structure may render it highly sensitive to protease digestion (42, 50). Spacer has a cleavage site for thrombin, deduced by adding thrombin to Pol proteins of various lengths, and can be cleaved by cellular proteases in rat hepatocytes and human hepatoma cells *in vitro* (51, 52). Critically, these cleavage sites are conserved across hepadnaviridae (52), providing some evidence for a functional role. In theory, protease

**TABLE 2** Summary of HBV mutagenesis studies in spacer and the resultant phenotype<sup>a</sup>

Virus	Spacer residues	Phenotype	Reference
<b>Deletions</b>			
HBV	aa 175–300	No ability to engage in RNA binding, RNA packaging, or protein priming	69
HBV	aa 178–336	Reduced polymerase activity (70% of WT)	65
HBV	aa 179–257	No impact on replication capacity and RNA encapsidation	67
HBV	aa 196–291	Increased expression; decreased protein priming ability	62
HBV	aa 199–300	Incapable of RNA packaging; able to bind RNA and perform protein priming	69
HBV	aa 201–292	Decreased polymerase activity (40% of WT)	14
HBV	aa 201–292	Decreased polymerase activity (40% of WT); no impact on RNA packaging	34
HBV	aa 201–335	Impaired polymerase activity (<0.3% of WT)	14
HBV	aa 258–286	Capable of RNA encapsidation	67
HBV	aa 258–336	Incapable of replication and RNA encapsidation	67
HBV	aa 287–317	Incapable of RNA encapsidation	67
HBV	aa 293–335	Impaired polymerase activity (<0.3% of WT)	14
HBV	aa 293–335	Impaired polymerase activity (<0.5% of WT); decrease in RNA packaging efficiency	34
HBV	aa 300–334	Impaired priming function	39
HBV	aa 318–336	Incapable of RNA encapsidation	67
HBV	nt 2878–129	Increased stability of the P protein	66
DHBV	aa 307–356	No impact on polymerase activity	5
<b>Substitutions</b>			
HBV	C312A, C323A, C327A, C341A	Each mutation individually is incapable of RNA binding, RNA packaging, and protein priming	69
HBV	C312A, C323A, C327A	Each mutation individually is incapable of RNA encapsidation	67
HBV	C326R	No detectable replication activity	71
HBV	E245L	Reduced but still significant level of replication	71
HBV	R219S, R246G, G261R, N270S, A272S, R300H	Decreased replication efficiency	27
<b>Insertions</b>			
DHBV	Bacterial protein A inserted into spacer	No impact on DNA synthesis and RNA packaging activities	33
DHBV	12 nt at 1,212 bp (in frame)	No impact on replication capacity	40
DHBV	4 nt at 1,212 bp (frameshift)	Impaired replication capacity	40

<sup>a</sup>The majority of studies were in human HBV, with a small number also carried out using duck HBV (DHBV) as an animal model. Polymorphism locations are provided as cited in the primary studies, which may differ from reference genotype A sequence [X02763](#). aa, amino acid residue; nt, nucleotide; bp, base pairs; WT, wild type.

digestion could separate functional TP and RT/RNase, facilitating movement of the polymerase down the RNA while TP remains bound at the 5' end (5).

However, while multiple studies have observed smaller molecular forms of P *in vitro* using a variety of assays (5, 52–54), and proteolytic cleavage is a known mechanism of protein activation in other viral proteins (55–57) (including the generation of HBe antigen [HBeAg] through cleavage of the precore/core fusion protein [58]), there is no protease requirement for functional activity of P (59, 60). The intact polyprotein demonstrates RT activity *in vitro* (59), and full-length P can be found intact in virions while carrying out the essential function of genome encapsidation (60). Furthermore, there is no evidence to suggest that protease digestion plays a role in later stages of the HBV replication cycle, such as cccDNA formation (52). A continued linkage model is based on evidence that TP and RT remain connected during replication, explaining second-strand initiation and conversion from a linear to a circular genome (5, 42, 43).

**Spacer provides protein flexibility.** Spacer may be crucial in providing flexibility for the TP and RT domains to assume the structural conformations needed to perform diverse functions, including RNA 5' epsilon binding, pgRNA encapsidation, and DNA synthesis. For these events to occur, P must switch from a stable (inactive) configuration to its active state, the kinetics of which may be enabled by spacer and facilitated by a variety of host cellular chaperones (44), such as heat shock protein 40 (hsp40), hsp70, and hsp90 (61–64), although a direct host-virus interaction has not been elucidated. This flexible, dynamic structure is typical of an IDPR, with interplay between

subdomains facilitated by interaction with other viral and host proteins and by characteristics of the environment, such as pH and temperature (46).

**Role of spacer in P function.** Mutagenesis studies (Table 2) have investigated the role of the spacer domain in the function of P. Performing experiments *in vitro* has enabled researchers to isolate the effect of mutations on P function, preventing any confounding effect that the mutations in the overlapping reading frame have on protein S. Deletion of the spacer domain has generated conflicting evidence, demonstrating both reduced (65) and increased (66) DNA-dependent DNA-polymerase activity. However, increased activity following spacer deletion may relate to higher expression and stability of the P protein without spacer (62, 66). In contrast, replacing the spacer domain of a less efficiently replicating isolate with the domain of a more efficiently replicating isolate has led to substantial increase in replication efficiency (27); the authors explained this increase with a point mutation at residue 300, part of the minimal portion of the spacer domain, although the isolated role of residue 300 remains unclear. Indeed, experiments with spacer deletion mutants have suggested differing significance of the N-terminal and C-terminal regions (Fig. 1D). Deletion of the N-terminal two-thirds of spacer does not affect RNA encapsidation, replicative capacity, or pgRNA availability (34, 67). Conversely, deletion of the C-terminal third of spacer results in a drastic decrease in RNA packaging efficiency (34, 67), with residues 300 to 334 being essential for the priming function of both TP and RT domains (39), suggesting it is required for replication competence (14).

In addition to the C-terminal region, there is also evidence that other upstream residues of spacer play a functional role. In a deletion mutant study, the TP domain did not require spacer to function, but maximal activity of TP required the N-terminal part of spacer (39). Another study deleted the central part of spacer (aa 196 to 291) and demonstrated that although the truncated P was more readily expressed than full-length P, it could not engage in protein priming (62). A successful interaction between truncated P and the RNA epsilon element could, however, be reconstituted, suggesting that while the spacer deletion mutant remains capable of binding RNA, mere physical binding is not sufficient for a functional protein–RNA interaction (62). An intact spacer domain therefore may be needed for conformational changes required for protein priming or interaction with essential cellular chaperones.

Three cysteine residues in the C-terminal region of spacer (positions 312, 323, and 327, based on a genotype D alignment) (Fig. 1D) have been established as essential for RT activity (29, 67–69). Together with a fourth cysteine in the N terminus of RT, these residues form part of a putative zinc finger DNA-binding motif (69, 70) and are conserved across mammalian hepadnaviridae (67). A C327A substitution in the C-terminal portion of spacer is lethal for HBV, abolishing endogenous polymerase activity, suggesting that spacer has a direct role in replication or an essential role in the folding of P to enable replication (71). Nearly 2 decades after this original research, the essential role of these cysteine residues in pgRNA encapsidation and for binding of P to the 5' epsilon stem-loop of pgRNA was confirmed using alanine-scanning mutagenesis (67). Cotransfection of mutant and WT P in a 3:1 ratio did not have a dominant-negative effect: the mutant protein does not bind pgRNA or incorporate into nucleocapsids, and WT P binds and carries out replication (67). Mutating the three C-terminal cysteine residues produces variants that are completely defective in 5' epsilon binding, protein priming, and RNA packaging. Therefore, overall this putative zinc finger domain is essential for RT catalytic activity (29, 68).

It is noteworthy that data from *in vitro* HBV replication models have been limited due to reliance on heavily adapted cell culture systems and with spacer modifications (e.g., green fluorescent protein insertions and various deletions) that may interfere with enzymatic function, making it difficult to cross-compare data from different studies. More recently developed culture approaches offer the potential for studying HBV replication more reliably (72, 73). Furthermore, while findings from animal models should always be extrapolated with caution, DHBV and similar animal models are considered reliable for use in studying viral entry, polymerase function, and immune



response, although less reliable to study the clinical manifestations of HBV, such as cirrhosis and hepatocellular carcinoma (HCC) (74).

### HOST-VIRUS INTERACTION AND OUTCOMES OF INFECTION

Spacer has been determined to contain the highest number of positively selected sites of any domain in the HBV genome. Maximum likelihood estimates of codon selection suggest spacer is at the center of a complex epistatic network coordinating clusters of mutations involved in virulence, immune escape, and drug resistance (16).

**Spacer mutations associated with HCC.** The incidence of HCC, disease progression, and treatment outcomes are influenced by a complex interplay of host, viral, and environmental factors. Certain viral genotypes (including C and F) are associated with a higher lifetime risk of HCC (75–80), and mutations in spacer, as well as other regions, have been associated with HCC (75, 76, 81–83). Next-generation sequencing has identified multiple HCC-associated single-nucleotide variants (SNVs), concentrated in PreS1 and spacer (84). A mutation at nt 31 was significantly associated with higher alpha-feto-protein levels, larger tumor size, and shorter postoperative survival (85). However, the role of these SNVs in hepatocarcinogenesis remains poorly understood, as it is unclear if oncogenic activity is driven by changes to the PreS1 and/or spacer function (84, 86).

PreS deletions, which also result in deletions within the overlapping spacer, have been reported as independent risk factors for HCC (81–83, 87–89), potentially as a result of accumulation of the misfolded mutant surface antigen causing endoplasmic reticulum stress (87, 90–93). A double-spliced (2.2-kb) HBV variant isolated from liver tissue had deletions spanning most of spacer and PreS2, and parts of PreS1, S, RT, and TP (94), and the 2.2-kb variant is increased in HCC tissue compared to peritumor tissue (95). Although the double-spliced variant is not replication competent, adding it to the full-length version resulted in dose-dependent enhancement of replication efficiency (94). Splice variants therefore might contribute to the increased and persistent HBV replication that leads to HCC in some patients (96, 97). As the S and P proteins are translated from different RNA transcripts, functional studies of these deletions have only considered the impact of the deletions in S, but it is possible the corresponding deletions in spacer also contribute to disease progression.

**Spacer mutations associated with OBI.** Mutations in spacer have also been described in the setting of occult HBV infection (OBI), defined as detectable HBV DNA in the absence of HBsAg in the serum (the marker of active ongoing infection) (98). OBI can reactivate to cause hepatitis flares (99, 100), can be associated with HCC development (101–103), and can be a reservoir for transmission (104, 105). In a study reporting 235 OBI-associated mutations, 151 were in P and 27 in spacer, reflecting immune evasion and leading to decreased viral replication and reduced immune activation (106). However, the biological and clinical significance of specific polymorphisms remains to be clearly elucidated.

**Escape from the adaptive immune response and vaccine-mediated immunity.** P is an important target of the cellular and humoral immune response (107–115), harboring epitopes for antibodies and CD4<sup>+</sup> and CD8<sup>+</sup> lymphocytes (30, 116–118). While some regions of P are functionally conserved, variation elsewhere in the polyprotein can contribute to immune evasion. Positive immune selection pressure acts on parts of P, particularly spacer, both in concert with PreS and independent of the overlapping S gene (22, 44). In a study of HBV genotype D, 13 out of 15 sites under positive selection were located in spacer, and some overlap was detected between positively selected sites in spacer and in PreS, indicating that positive selection can collocate in overlapping genes (44).

One B cell epitope in spacer (aa 225 to 250) may explain some of the sequence diversity of the domain (30) (Fig. 1D). Sequence analysis of P and S genes in vaccine escape mutants among HBV-vaccinated children (119) detected amino acid substitutions that affect B and T cell epitopes, including 8 amino acid substitutions in this B cell epitope in spacer (119). However, disaggregating the specific influence of individual mutations in mediating vaccine escape requires further efforts.

**Spacer mutations in the setting of drug resistance and to preserve viral fitness.** There is some evidence that spacer accommodates compensatory mutations to

ameliorate a fitness detriment caused by mutation(s) elsewhere in the viral genome. Due to spacer's proximity to key functional residues in TP and RT, it is possible that amino acid mutations in spacer affect the catalytic activity of P and therefore could fine-tune or restore changes imposed by other mutations (e.g., see reference 44). A study of covariance in genotypes B, C, and D found that polymorphisms were concentrated in the spacer and PreS1, suggesting a coevolutionary relationship between these sites and sites elsewhere in the genome (120).

Lamivudine (3TC) resistance mutations in the YMDD motif of RT, which negatively affect replicative ability, can be restored by compensatory mutations in the putative zinc finger subdomains (121), including spacer's C-terminal cysteines (29, 67–69). Other mutations can also be relevant in driving resistance; for example, RT (A181T) and spacer (S331C) polymorphisms together result in a decrease in 3TC susceptibility (122). A181T, S331C, and A181T+S331C mutants were 82%, 94%, and 96% replication efficient, respectively, compared to the wild type (122). Although the effect is modest, it is possible that the spacer mutation coevolved to restore a minor fitness detriment caused by the RT mutation (122), since residue 331 is within the portion of the spacer important for RT function (39). Furthermore, after 3TC discontinuation (when the mutant strain became undetectable), 3TC readministration has been associated with the reappearance of both mutations (122). A spacer deletion arising in an adefovir-treated patient is of uncertain significance in drug resistance (36), while selection of a mutation (A300E) has also been reported from a patient with clinical evidence of entecavir resistance, although this polymorphism did not alter *in vitro* resistance or replication capacity (123). The sparse evidence base for treatment-associated mutations in spacer is related to the lack of routine investigation of drug resistance in clinical practice and focus only on RT if sequencing is undertaken (29).

A study of duck HBV provided evidence of spacer's potential reactivity to mutations elsewhere in the genome, tolerating changes in its own sequence to overcome a fitness detriment by introducing insertions into the distal PreS (40). *In vivo*, in-frame mutants retained the inserted nucleotides, while frameshift mutants either reverted to wild type or selected a deletion in spacer/PreS that shifted the frame back to normal, restoring infectivity and secondary protein structure and demonstrating compensation for a fitness cost by deleting a specific nonessential portion of the genome (40).

## FUTURE DIRECTIONS AND TRANSLATIONAL APPLICATIONS

HBV sequencing has not been widely applied due to lack of a mandate for genotyping in clinical guidelines (e.g., see reference 124) and the typically low viral loads of many chronic infections (125). However, as sequencing methods are improved and there is increasing interest in the application of sequence data, opportunities will arise to enhance insights into viral diversity and its impact on disease outcomes (4, 126). Sequencing repositories contain an untapped resource for spacer data by way of sequences that have been generated for the S gene but could be mined for analysis of spacer (Fig. 1A).

To date, structural insights for P have been extrapolated from the homologous regions of the HIV polymerase (127). Determination of crystal structures may provide insights into the interactions between protein subdomains, including better understanding the role of spacer in maintaining integrity, supporting the viral replication cycle, and accounting for diverse outcomes of infection. The oncogenic mechanism of viral variants associated with HCC remains to be elucidated, but pathogen genomics may be relevant for future personalized medicine approaches for cancer surveillance or risk assessment. More research is needed to determine the specific protein-protein interactions between spacer and host chaperones and restriction factors. Ultimately, driving this agenda is important to inform clinical practice, for example, through using viral sequence to stratify patients for surveillance and/or treatment, to guide public health interventions, and to inform the development of new therapeutics (128).

## CONCLUSIONS

Spacer is a small but highly diverse and versatile domain, with features of an IDPR. Earlier studies assumed spacer was a dispensable region, and it is doubtless true that there is some redundancy, particularly in the N-terminal portion. However, evidence has also accumulated for important evolutionary and biological roles, more focused in the C-terminal third, which play essential roles in RNA binding and packaging, protein priming, and reverse transcription, with a dynamic structure that supports protein function, including potentially coordinating host-virus interactions. Further research is required to more definitively elucidate the functions of specific spacer residues.

Spacer sequence variation between and within mammalian hepadnaviridae, as well as between genotypes and subgenotypes of HBV, is highly lineage specific, overturning the hypothesis that spacer is diverse because of low functional or evolutionary importance. Rather, spacer may play a crucial role in the evolution of both the P and the overlapping S genes, influenced by pressure from exposure to drugs, vaccines, and the host immune response. Through the increasing use of whole-genome sequencing data and trends in personalized medicine, we are in an era of opportunities to further unravel the role of spacer in HBV evolution, epidemiology, and pathogenesis, with potentially important translational implications.

## ACKNOWLEDGMENTS

P.C.M. is funded by a Wellcome intermediate fellowship, reference no. 110110/Z/15/Z. A.L.M. has received NIHR Research Capability Funding (RCF) from the University of Oxford. This research was funded in whole, or in part, by the Wellcome Trust. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Outside this work, PCM supervises a PhD student who receives co-funding from GSK. Other authors have no competing interests to declare.

## REFERENCES

- Liang TJ. 2009. Hepatitis B: the virus and disease. *Hepatology* 49:513–521. <https://doi.org/10.1002/hep.22881>.
- Miller RH, Kaneko S, Chung CT, Girones R, Purcell RH. 1989. Compact organization of the hepatitis B virus genome. *Hepatology* 9:322–327. <https://doi.org/10.1002/hep.1840090226>.
- Simmonds P. 2001. The origin and evolution of hepatitis viruses in humans. *J Gen Virol* 82:693–712. <https://doi.org/10.1099/0022-1317-82-4-693>.
- McNaughton AL, D'Arienzo V, Ansari MA, Lumley SF, Littlejohn M, Revill P, McKeating JA, Matthews PC. 2019. Insights from deep sequencing of the HBV genome—unique, tiny, and misunderstood. *Gastroenterology* 156:384–399. <https://doi.org/10.1053/j.gastro.2018.07.058>.
- Bartenschlager R, Schaller H. 1988. The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *EMBO J* 7:4185–4192. <https://doi.org/10.1002/j.1460-2075.1988.tb03315.x>.
- Keese PK, Gibbs A. 1992. Origins of genes: “big bang” or continuous creation? *Proc Natl Acad Sci U S A* 89:9489–9493. <https://doi.org/10.1073/pnas.89.20.9489>.
- Pavesi A. 2006. Origin and evolution of overlapping genes in the family Microviridae. *J Gen Virol* 87:1013–1017. <https://doi.org/10.1099/vir.0.81375-0>.
- Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. 2009. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J Virol* 83:10719–10736. <https://doi.org/10.1128/JVI.00595-09>.
- Li F, Ding SW. 2006. Virus counterdefense: diverse strategies for evading the RNA-silencing immunity. *Annu Rev Microbiol*. <https://doi.org/10.1146/annurev.micro.60.080805.142205>.
- Pavesi A, Magiorkinis G, Karlin DG. 2013. Viral proteins originated de novo by overprinting can be identified by codon usage: application to the “gene nursery” of deltaretroviruses. *PLoS Comput Biol* 9:e1003162. <https://doi.org/10.1371/journal.pcbi.1003162>.
- Pavesi A. 2015. Different patterns of codon usage in the overlapping polymerase and surface genes of hepatitis B virus suggest a de novo origin by modular evolution. *J Gen Virol* 96:3577–3586. <https://doi.org/10.1099/jgv.0.000307>.
- Le Seyec J, Chouteau P, Cannie I, Guguen-Guillouzo C, Gripon P. 1999. Infection process of the hepatitis B virus depends on the presence of a defined sequence in the pre-S1 domain. *J Virol* 73:2052–2057. <https://doi.org/10.1128/JVI.73.3.2052-2057.1999>.
- Blanchet M, Sureau C. 2007. Infectivity determinants of the hepatitis B virus pre-S domain are confined to the N-terminal 75 amino acid residues. *J Virol* 81:5841–5849. <https://doi.org/10.1128/JVI.00096-07>.
- Radziwill G, Tucker W, Schaller H. 1990. Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. *J Virol* 64:613–620. <https://doi.org/10.1128/JVI.64.2.613-620.1990>.
- Toh H, Hayashida H, Miyata T. 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature* 305:827–829. <https://doi.org/10.1038/305827a0>.
- Campo DS, Dimitrova Z, Lara J, Purdy M, Thai H, Ramachandran S, Ganova-Raeva L, Zhai X, Forbi JC, Teo CG, Khudyakov Y. 2012. Coordinated evolution of the hepatitis B virus polymerase. *In Silico Biol* 11:175–182. <https://doi.org/10.3233/ISB-2012-0452>.
- Hughes AL, Hughes MAK. 2005. Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes. *Virus Res* 113:81–88. <https://doi.org/10.1016/j.virusres.2005.03.030>.
- Narechania A, Terai M, Burk RD. 2005. Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. *J Gen Virol* 86:1307–1313. <https://doi.org/10.1099/vir.0.80747-0>.
- Guyader S, Ducray DG. 2002. Sequence analysis of potato leafroll virus isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products. *J Gen Virol* 83:1799–1807. <https://doi.org/10.1099/0022-1317-83-7-1799>.
- Hughes AL, Westover K, da Silva J, O'Connor DH, Watkins DI. 2001. Simultaneous positive and purifying selection on overlapping reading frames of the tat and vpr genes of simian immunodeficiency virus. *J Virol* 75:7966–7972. <https://doi.org/10.1128/jvi.75.17.7966-7972.2001>.

21. Van De Klundert MAA, Cremer J, Kootstra NA, Boot HJ, Zaaijer HL. 2012. Comparison of the hepatitis B virus core, surface and polymerase gene substitution rates in chronically infected patients. *J Viral Hepat* 19: e34–e40. <https://doi.org/10.1111/j.1365-2893.2011.01506.x>.
22. Zaaijer HL, Van Hemert FJ, Koppelman MH, Lukashov VV. 2007. Independent evolution of overlapping polymerase and surface protein genes of hepatitis B virus. *J Gen Virol* 88:2137–2143. <https://doi.org/10.1099/vir.0.82906-0>.
23. Zhang D, Chen J, Deng L, Mao Q, Zheng J, Wu J, Zeng C, Li Y. 2010. Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. *Infect Genet Evol* 10:84–88. <https://doi.org/10.1016/j.meegid.2009.10.006>.
24. Ni Y, Sonnabend J, Seitz S, Urban S. 2010. The pre-S2 domain of the hepatitis B virus is dispensable for infectivity but serves a spacer function for L-protein-connected virus assembly. *J Virol* 84:3879–3888. <https://doi.org/10.1128/JVI.02528-09>.
25. Neurath AR, Kent SBH, Parker K, Prince AM, Strick N, Brotman B, Sproul P. 1986. Antibodies to a synthetic peptide from the preS 120–145 region of the hepatitis B virus envelope are virus-neutralizing. *Top Catal* 4:35–37.
26. Pontisso P, Schiavon E, Fraiese A, Pornaro E, Realdi G, Alberti A. 1986. Antibody to the hepatitis B virus receptor for polymerized albumin in acute infection and in hepatitis B vaccine recipients. *J Hepatol* 3: 393–398. [https://doi.org/10.1016/S0168-8278\(86\)80494-9](https://doi.org/10.1016/S0168-8278(86)80494-9).
27. Lin X, Yuan Z-H, Wu L, Ding J-P, Wen Y-M. 2001. A single amino acid in the reverse transcriptase domain of hepatitis B virus affects virus replication efficiency. *J Virol* 75:11827–11833. <https://doi.org/10.1128/JVI.75.23.11827-11833.2001>.
28. Tiollais P, Pourcel C, Dejean A. 1985. The hepatitis B virus. *Nature* 317: 489–495. <https://doi.org/10.1038/317489a0>.
29. Clark DN, Hu J. 2015. Unveiling the roles of HBV polymerase for new antiviral strategies. *Future Virol* 10:283–295. <https://doi.org/10.2217/fvl.14.113>.
30. zu Putlitz J, Lanford RE, Carlson RI, Notvall L, de la Monte SM, Wands JR. 1999. Properties of monoclonal antibodies directed against hepatitis B virus polymerase protein. *J Virol* 73:4188–4196. <https://doi.org/10.1128/JVI.73.5.4188-4196.1999>.
31. Norder H, Ebert JW, Fields HA, Mushahwar IK, Magnusius LO. 1996. Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology* 218:214–223. <https://doi.org/10.1006/viro.1996.0181>.
32. Sprengel R, Kuhn C, Will H, Schaller H. 1985. Comparative sequence analysis of duck and human hepatitis B virus genomes. *J Med Virol* 15: 323–333. <https://doi.org/10.1002/jmv.1890150402>.
33. Chang LJ, Hirsch RC, Ganem D, Varmus HE. 1990. Effects of insertional and point mutations on the functions of the duck hepatitis B virus polymerase. *J Virol* 64:5553–5558. <https://doi.org/10.1128/JVI.64.11.5553-5558.1990>.
34. Bartenschlager R, Junker-Niepmann M, Schaller H. 1990. The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J Virol* 64:5324–5332. <https://doi.org/10.1128/jvi.64.11.5324-5332.1990>.
35. Klingmüller U. 1988. PhD thesis. University of Heidelberg, Heidelberg, Germany.
36. Hong R, Bai W, Zhai J, Liu W, Li X, Zhang J, Cui X, Zhao X, Ye X, Deng Q, Tiollais P, Wen Y, Liu J, Xie Y. 2013. Novel recombinant hepatitis B virus vectors efficiently deliver protein and RNA encoding genes into primary hepatocytes. *J Virol* 87:6615–6624. <https://doi.org/10.1128/JVI.03328-12>.
37. Gerner P, Schäfer HMC, Prange R, Pravitt D, Wirth S. 2003. Functional analysis of a rare HBV deletion mutant in chronically infected children. *Pediatr Res* 53:891–897. <https://doi.org/10.1203/01.PDR.0000064906.63939.72>.
38. Pollicino T, Amaddeo G, Restuccia A, Raffa G, Alibrandi A, Cutroneo G, Favaloro A, Maimone S, Squadrito G, Raimondo G. 2012. Impact of hepatitis B virus (HBV) preS/S genomic variability on HBV surface antigen and HBV DNA serum levels. *Hepatology* 56:434–443. <https://doi.org/10.1002/hep.25592>.
39. Lanford RE, Kim Y-H, Lee H, Notvall L, Beames B. 1999. Mapping of the hepatitis B virus reverse transcriptase TP and RT domains by transcomplementation for nucleotide priming and by protein-protein interaction. *J Virol* 73:1885–1893. <https://doi.org/10.1128/JVI.73.3.1885-1893.1999>.
40. Li JS, Cova L, Buckland R, Lambert V, Deléage G, Trépo C. 1989. Duck hepatitis B virus can tolerate insertion, deletion, and partial frameshift mutation in the distal pre-S region. *J Virol* 63:4965–4968. <https://doi.org/10.1128/JVI.63.11.4965-4968.1989>.
41. Wang GH, Zoulim F, Leber EH, Kitson J, Seeger C. 1994. Role of RNA in enzymatic activity of the reverse transcriptase of hepatitis B viruses. *J Virol* 68:8437–8442. <https://doi.org/10.1128/JVI.68.12.8437-8442.1994>.
42. Radziwill G, Zentgraf H, Schaller H, Bosch V. 1988. The duck hepatitis B virus DNA polymerase is tightly associated with the viral core structure and unable to switch to an exogenous template. *Virology* 163:123–132. [https://doi.org/10.1016/0042-6822\(88\)90239-5](https://doi.org/10.1016/0042-6822(88)90239-5).
43. Bosch V, Bartenschlager R, Radziwill G, Schaller H. 1988. The duck hepatitis B virus P-gene codes for protein strongly associated with the 5′-end of the viral DNA minus strand. *Virology* 166:475–485. [https://doi.org/10.1016/0042-6822\(88\)90518-1](https://doi.org/10.1016/0042-6822(88)90518-1).
44. Chen P, Gan Y, Han N, Fang W, Li J, Zhao F, Hu K, Rayner S. 2013. Computational evolutionary analysis of the overlapped surface (S) and polymerase (P) region in hepatitis B virus indicates the spacer domain in P is crucial for survival. *PLoS One* 8:e60098. <https://doi.org/10.1371/journal.pone.0060098>.
45. Seitz S, Habjanić J, Schütz AK, Bartenschlager R. 2020. The hepatitis B virus envelope proteins: molecular gymnastics throughout the viral life cycle. *Annu Rev Virol* 7:263–288. <https://doi.org/10.1146/annurev-virology-092818-015508>.
46. Uversky VN. 2019. Intrinsically disordered proteins and their “mysterious” (meta)physics. *Front Phys*. <https://doi.org/10.3389/fphy.2019.00010>.
47. Mishra PM, Verma NC, Rao C, Uversky VN, Nandi CK. 2020. Intrinsically disordered proteins of viruses: involvement in the mechanism of cell regulation and pathogenesis. *Prog Mol Biol Transl Sci* 174:1–78. <https://doi.org/10.1016/bs.pmbts.2020.03.001>.
48. Uversky VN. 2016. Paradoxes and wonders of intrinsic disorder: complexity of simplicity. *Intrinsically Disord Proteins* 4:e1135015. <https://doi.org/10.1080/21690707.2015.1135015>.
49. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L. 2015. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* 72:137–151. <https://doi.org/10.1007/s00018-014-1661-9>.
50. Schlicht HJ, Radziwill G, Schaller H. 1989. Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core-polymerase fusion proteins. *Cell* 56:85–92. [https://doi.org/10.1016/0092-8674\(89\)90986-0](https://doi.org/10.1016/0092-8674(89)90986-0).
51. Wu TP, Yee V, Tulinsky A, Chrusciel RA, Nakanishi H, Shen R, Priebe C, Kahn M. 1993. The structure of a designed peptidomimetic inhibitor complex of  $\alpha$ -thrombin. *Protein Eng Des Sel* 6:471–478. <https://doi.org/10.1093/protein/6.5.471>.
52. Lin CG, Yang SJ, Hwang WL, Su TS, Lo SJ. 1995. Demonstration of the presence of protease-cutting site in the spacer of hepatitis B viral Pol protein. *J Virol Methods* 51:61–73. [https://doi.org/10.1016/0166-0934\(94\)00118-z](https://doi.org/10.1016/0166-0934(94)00118-z).
53. Bavand MR, Laub O. 1988. Two proteins with reverse transcriptase activities associated with hepatitis B virus-like particles. *J Virol* 62:626–628. <https://doi.org/10.1128/JVI.62.2.626-628.1988>.
54. Feitelson MA, Clayton MM, Duan LX. 1991. Polymerase-related polypeptides associated with woodchuck hepatitis core antigen (WHcAg) particles. *Virology* 180:430–433. [https://doi.org/10.1016/0042-6822\(91\)90052-d](https://doi.org/10.1016/0042-6822(91)90052-d).
55. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RAF, Scolnick EM, Sigal IS. 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A* 85:4686–4690. <https://doi.org/10.1073/pnas.85.13.4686>.
56. Krausslich HG, Wimmer E. 1988. Viral proteinases. *Annu Rev Biochem* 57: 701–754. <https://doi.org/10.1146/annurev.bi.57.070188.003413>.
57. Craven RC, Bennett RP, Wills JW. 1991. Role of the avian retroviral protease in the activation of reverse transcriptase during virion assembly. *J Virol* 65:6205–6217. <https://doi.org/10.1128/JVI.65.11.6205-6217.1991>.
58. Schodel F, Peterson D, Zheng J, Jones JE, Hughes JL, Milich DR. 1993. Structure of hepatitis B virus core and e-antigen. A single precore amino acid prevents nucleocapsid assembly. *J Biol Chem* 268:1332–1337. [https://doi.org/10.1016/S0021-9258\(18\)54079-5](https://doi.org/10.1016/S0021-9258(18)54079-5).
59. Wang GH, Seeger C. 1992. The reverse transcriptase of hepatitis B virus acts as a protein primer for viral DNA synthesis. *Cell* 71:663–670. [https://doi.org/10.1016/0092-8674\(92\)90599-8](https://doi.org/10.1016/0092-8674(92)90599-8).
60. Bartenschlager R, Schaller H. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J* 11:3413–3420. <https://doi.org/10.1002/j.1460-2075.1992.tb05420.x>.
61. Hu J, Toft D, Anselmo D, Wang X. 2002. In vitro reconstitution of functional hepadnavirus reverse transcriptase with cellular chaperone proteins. *J Virol* 76:269–279. <https://doi.org/10.1128/jvi.76.1.269-279.2002>.
62. Hu J, Flores D, Toft D, Wang X, Nguyen D. 2004. Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. *J Virol* 78:13122–13131. <https://doi.org/10.1128/JVI.78.23.13122-13131.2004>.



63. Wang X, Grammatikakis N, Hu J. 2002. Role of p50/CDC37 in hepatitis virus assembly and replication. *J Biol Chem* 277:24361–24367. <https://doi.org/10.1074/jbc.M202198200>.
64. Stahl M, Beck J, Nassal M. 2007. Chaperones activate hepatitis virus reverse transcriptase by transiently exposing a C-proximal region in the terminal protein domain that contributes to  $\epsilon$  RNA binding. *J Virol* 81:13354–13364. <https://doi.org/10.1128/JVI.01196-07>.
65. Kim Y, Hong Y, Jung G. 1999. Hepatitis B virus: DNA polymerase activity of deletion mutants. *Biochem Mol Biol Int* 47:301–308. <https://doi.org/10.1080/15216549900201323>.
66. Younhee K, Guhung J. 1999. Active human hepatitis B viral polymerase expressed in rabbit reticulocyte lysate system. *Virus Genes* 19:123–130.
67. Kim S, Lee J, 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. <https://doi.org/10.1128/JVI.00332-09>.
68. Jones SA, Hu J. 2013. Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. *Emerg Microbes Infect* 2:e56. <https://doi.org/10.1038/emi.2013.56>.
69. Jones SA, Clark DN, Cao F, Tavis JE, Hu J. 2014. Comparative analysis of hepatitis B virus polymerase sequences required for viral RNA binding, RNA packaging, and protein priming. *J Virol* 88:1564–1572. <https://doi.org/10.1128/JVI.02852-13>.
70. Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, Melino G, Raschella G. 2017. Zinc-finger proteins in health and disease. *Cell Death Discov* 3:17071. <https://doi.org/10.1038/cddiscovery.2017.71>.
71. Faruqi AF, Roychoudhury S, Greenberg R, Israel J, Shih C. 1991. Replication-defective missense mutations within the terminal protein and spacer/intron regions of the polymerase gene of human hepatitis B virus. *Virology* 183:764–768. [https://doi.org/10.1016/0042-6822\(91\)91007-4](https://doi.org/10.1016/0042-6822(91)91007-4).
72. Michailidis E, Pabon J, Xiang K, Park P, Ramanan V, Hoffmann H-H, Schneider WM, Bhatia SN, de Jong YP, Shlomai A, Rice CM. 2017. A robust cell culture system supporting the complete life cycle of hepatitis B virus. *Sci Rep* 7:16616. <https://doi.org/10.1038/s41598-017-16882-5>.
73. Rao S, Hossain T, Mahmoudi T. 2021. 3D human liver organoids: an in vitro platform to investigate HBV infection, replication and liver tumorigenesis. *Cancer Lett* 506:35–44. <https://doi.org/10.1016/j.canlet.2021.02.024>.
74. Zhang X, Wang X, Wu M, Ghildyal R, Yuan Z. 2021. Animal models for the study of hepatitis B virus pathobiology and immunity: past, present, and future. *Front Microbiol* 12:1921. <https://doi.org/10.3389/fmicb.2021.715450>.
75. Fletcher GJ, Eapen CE, Abraham P. 2020. Hepatitis B genotyping: the utility for the clinicians. *Indian J Gastroenterol* 39:315–320. <https://doi.org/10.1007/s12664-019-00995-y>.
76. Lin CL, Kao JH. 2017. Natural history of acute and chronic hepatitis B: the role of HBV genotypes and mutants. *Best Pract Res Clin Gastroenterol* 31:249–255. <https://doi.org/10.1016/j.bpg.2017.04.010>.
77. Rajoriya N, Combet C, Zoulim F, Janssen HLA. 2017. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *J Hepatol* 67:1281–1297. <https://doi.org/10.1016/j.jhep.2017.07.011>.
78. Chan HLY, Hui AY, Wong ML, Tse AML, Hung LCT, Wong VWS, Sung JY. 2004. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. *Gut* 53:1494–1498. <https://doi.org/10.1136/gut.2003.033324>.
79. Chuon C, Takahashi K, Matsuo J, Katayama K, Yamamoto C, Ko K, Hok S, Nagashima S, Akbar SMF, Tanaka J. 2019. High possibility of hepatocarcinogenesis in HBV genotype C1 infected Cambodians is indicated by 340 HBV C1 full-genomes analysis from GenBank. *Sci Rep* 9:1–11. <https://doi.org/10.1038/s41598-019-48304-z>.
80. Wen J, Song C, Jiang D, Jin T, Dai J, Zhu L, An J, Liu Y, Ma S, Qin N, Liang C, Chen J, Jiang Y, Yang L, Liu J, Liu L, Geng T, Chen C, Jiang J, Chen J, Zhu F, Zhu Y, Yu L, Shen H, Zhai X, Xu J, Hu Z. 2015. Hepatitis B virus genotype, mutations, human leukocyte antigen polymorphisms and their interactions in hepatocellular carcinoma: a multi-centre case-control study. *Sci Rep* 5:16489. <https://doi.org/10.1038/srep16489>.
81. Chen CH, Changchien CS, Lee CM, Hung CH, Hu TH, Wang JH, Wang JC, Lu SN. 2008. Combined mutations in pre-S/surface and core promoter/precore regions of hepatitis B virus increase the risk of hepatocellular carcinoma: a case-control study. *J Infect Dis* 198:1634–1642. <https://doi.org/10.1086/592990>.
82. Fang ZL, Sabin CA, Dong BQ, Wei SC, Chen QY, Fang KX, Yang JY, Huang J, Wang XY, Harrison TJ. 2008. Hepatitis B virus pre-S deletion mutations are a risk factor for hepatocellular carcinoma: a matched nested case-control study. *J Gen Virol* 89:2882–2890. <https://doi.org/10.1099/vir.0.2008/002824-0>.
83. Lin CL, Liu CH, Chen W, Huang WL, Chen PJ, Lai MY, Chen DS, Kao JH. 2007. Association of pre-S deletion mutant of hepatitis B virus with risk of hepatocellular carcinoma. *J Gastroenterol Hepatol* 22:1098–1103. <https://doi.org/10.1111/j.1440-1746.2006.04515.x>.
84. Wu IC, Liu WC, Chang TT. 2018. Applications of next-generation sequencing analysis for the detection of hepatocellular carcinoma-associated hepatitis B virus mutations. *J Biomed Sci* 25:51. <https://doi.org/10.1186/s12929-018-0442-4>.
85. Yin F, Xie Y, Fan H, Zhang J, Guo Z. 2017. Mutations in hepatitis B virus polymerase are associated with the postoperative survival of hepatocellular carcinoma patients. *PLoS One* 12:e0189730. <https://doi.org/10.1371/journal.pone.0189730>.
86. Liu WC, Wu IC, Lee YC, Lin CP, Cheng JH, Lin YJ, Yen CJ, Cheng PN, Li PF, Cheng YT, Cheng PW, Sun KT, Yan SL, Lin JJ, Yang JC, Chang KC, Ho CH, Tseng VS, Chang BCH, Wu JC, Chang TT. 2017. Hepatocellular carcinoma-associated single-nucleotide variants and deletions identified by the use of genome-wide high-throughput analysis of hepatitis B virus. *J Pathol* 243:176–192. <https://doi.org/10.1002/path.4938>.
87. Wang HC, Huang W, Der Lai M, Su IJ. 2006. Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis. *Cancer Sci* 97:683–688. <https://doi.org/10.1111/j.1349-7006.2006.00235.x>.
88. Cohen D, Ghosh S, Shimakawa Y, Ramou N, Garcia PS, Dubois A, Guillot C, Deluce NK-N, Tilloy V, Durand G, Voegelé C, Ndow G, d'Alessandro U, Brochier-Armanet C, Alain S, Le Calvez-Kelm F, Hall J, Zoulim F, Mendy M, Thursz M, Lemoine M, Chemin I. 2020. Hepatitis B virus preS2Δ38–55 variants: a newly identified risk factor for hepatocellular carcinoma. *JHEP Rep* 2:100144. <https://doi.org/10.1016/j.jhepr.2020.100144>.
89. Su I-J, Hui-Ching Wang L, Hsieh W-C, Wu H-C, Teng C-F, Tsai H-W, Huang W. 2014. The emerging role of hepatitis B virus pre-S2 deletion mutant proteins in HBV tumorigenesis. <https://doi.org/10.1186/s12929-014-0098-7>.
90. Hsieh YH, Su IJ, Wang HC, Chang WW, Lei HY, Lai MD, Chang WT, Huang W. 2004. Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. *Carcinogenesis* 25:2023–2032. <https://doi.org/10.1093/carcin/bgh207>.
91. Hsieh YH, Su IJ, Wang HC, Tsai JH, Huang YJ, Chang WW, Lai MD, Lei HY, Huang W. 2007. Hepatitis B virus pre-S2 mutant surface antigen induces degradation of cyclin-dependent kinase inhibitor p27Kip1 through c-Jun activation domain-binding protein 1. *Mol Cancer Res* 5:1063–1072. <https://doi.org/10.1158/1541-7786.MCR-07-0098>.
92. Su I-J, Wang H-C, Wu H-C, Huang W-Y. 2008. Ground glass hepatocytes contain pre-S mutants and represent preneoplastic lesions in chronic hepatitis B virus infection. *J Gastroenterol Hepatol* 23:1169–1174. <https://doi.org/10.1111/j.1440-1746.2008.05348.x>.
93. Wang HC, Wu HC, Chen CF, Fausto N, Lei HY, Su IJ. 2003. Different types of ground glass hepatocytes in chronic hepatitis B virus infection contain specific pre-S mutants that may induce endoplasmic reticulum stress. *Am J Pathol* 163:2441–2449. [https://doi.org/10.1016/S0002-9440\(10\)63599-7](https://doi.org/10.1016/S0002-9440(10)63599-7).
94. Ma ZM, Lin X, Wang YX, Tian XC, Xie YH, Wen YM. 2009. A double-spliced defective hepatitis B virus genome derived from hepatocellular carcinoma tissue enhanced replication of full-length virus. *J Med Virol* 81:230–237. <https://doi.org/10.1002/jmv.21393>.
95. Lin X, Wen Y, Wan D, Qian G, Gu J. 2002. Structural and functional analysis of 2.2 kb spliced variant of hepatitis B virus genomes isolated from liver tissues from hepatocellular carcinoma patients. *Chin J Exp Clin Virol* 16:11–15.
96. Lin X, Qian G-S, Lu P-X, Wu L, Wen Y-M. 2001. Full-length genomic analysis of hepatitis B virus isolates in a patient progressing from hepatitis to hepatocellular carcinoma. *J Med Virol* 64:299–304. <https://doi.org/10.1002/jmv.1050>.
97. Lin X, Zhang-Mei M, Yao X, Zhang YP, Wen YM. 2002. Replication efficiency and sequence analysis of full-length hepatitis B virus isolates from hepatocellular carcinoma tissues. *Int J Cancer* 102:487–491. <https://doi.org/10.1002/ijc.10733>.
98. Chevrier MC, St-Louis M, Perreault J, Caron B, Castilloux C, Laroche J, Delage G. 2007. Detection and characterization of hepatitis B virus of anti-hepatitis B core antigen-reactive blood donors in Quebec with an in-house nucleic acid testing assay. *Transfusion* 47:1794–1802. <https://doi.org/10.1111/j.1537-2995.2007.01394.x>.
99. Hui CK, Cheung WWW, Zhang HY, Au WY, Yueng YH, Leung AYH, Leung N, Luk JM, Lie AKW, Kwong YL, Liang R, Lau GKK. 2006. Kinetics and risk of de novo hepatitis B infection in HBsAg-negative patients undergoing cytotoxic chemotherapy. *Gastroenterology* 131:59–68. <https://doi.org/10.1053/j.gastro.2006.04.015>.
100. Coppola N, Tonziello G, Pisaturo M, Messina V, Guastafierro S, Fiore M, Iodice V, Sagnelli C, Stanzione M, Capoluongo N, Pasquale G, Sagnelli E. 2011. Reactivation of overt and occult hepatitis B infection in various immunosuppressive settings. *J Med Virol* 83:1909–1916. <https://doi.org/10.1002/jmv.22199>.



101. Chen CH, Changchien CS, Lee CM, Tung WC, Hung CH, Hu TH, Wang JH, Wang JC, Lu SN. 2009. A study on sequence variations in pre-S/surface, X and enhancer II/core promoter/precore regions of occult hepatitis B virus in non-B, non-C hepatocellular carcinoma patients in Taiwan. *Int J Cancer* 125:621–629. <https://doi.org/10.1002/ijc.24416>.
102. Ikeda K, Kobayashi M, Someya T, Saitoh S, Hosaka T, Akuta N, Suzuki F, Suzuki Y, Arase Y, Kumada H. 2009. Occult hepatitis B virus infection increases hepatocellular carcinogenesis by eight times in patients with non-B, non-C liver cirrhosis: a cohort study. *J Viral Hepat* 16:437–443. <https://doi.org/10.1111/j.1365-2893.2009.01085.x>.
103. Wong DKH, Huang FY, Lai CL, Poon RTP, Seto WK, Fung J, Hung IFN, Yuen MF. 2011. Occult hepatitis B infection and HBV replicative activity in patients with cryptogenic cause of hepatocellular carcinoma. *Hepatology* 54:829–836. <https://doi.org/10.1002/hep.24551>.
104. Levicnik-Stezinar S, Rahne-Potokar U, Candotti D, Lelie N, Allain JP. 2008. Anti-HBs positive occult hepatitis B virus carrier blood infectious in two transfusion recipients. *J Hepatol* 48:1022–1025. <https://doi.org/10.1016/j.jhep.2008.02.016>.
105. Yuen MF, Wong DKH, Lee CK, Tanaka Y, Allain JP, Fung J, Leung J, Lin CK, Sugiyama M, Sugauchi F, Mizokami M, Lai CL. 2011. Transmissibility of hepatitis B virus (HBV) infection through blood transfusion from blood donors with occult HBV infection. *Clin Infect Dis* 52:624–632. <https://doi.org/10.1093/cid/ciq247>.
106. Powell EA, Gededzha MP, Rentz M, Rakgole NJ, Selabe SG, Seleise TA, Mphahlele MJ, Blackard JT. 2015. Mutations associated with occult hepatitis B in HIV-positive South Africans. *J Med Virol* 87:388–400. <https://doi.org/10.1002/jmv.24057>.
107. Lumley S, McNaughton A, Klenerman P, Lythgoe K, Matthews P. 2018. Hepatitis B virus adaptation to the CD8+ T cell response: consequences for host and pathogen. *Front Immunol* 9:1561. <https://doi.org/10.3389/fimmu.2018.01561>.
108. Lumley S, Noble H, Hadley M, Callow L, Malik A, Chua Y, Duffey O, Grolmusova N, Kumar A, Ravenscroft S, Spencer J, Neumann-Haefelin C, Thimme R, Andersson M, Klenerman P, Barnes E, Matthews P. 2016. Hepitopes: a live interactive database of HLA class I epitopes in hepatitis B virus. *Wellcome Open Res* 1:9. <https://doi.org/10.12688/wellcomeopenres.9952.1>.
109. Chang L-J, Dienstag J, Ganem D, Varmus H. 1989. Detection of antibodies against hepatitis B virus polymerase antigen in hepatitis B virus-infected patients. *Hepatology* 10:332–335. <https://doi.org/10.1002/hep.1840100314>.
110. Feitelson MA, Millman I, Duncan GD, Blumberg BS. 1988. Presence of antibodies to the polymerase gene product(s) of hepatitis B and woodchuck hepatitis virus in natural and experimental infections. *J Med Virol* 24:121–136. <https://doi.org/10.1002/jmv.1890240202>.
111. Kann M, Köchel HG, Uy A, Thomssen R. 1993. Diagnostic significance of antibodies to hepatitis B virus polymerase in acutely and chronically HBV-infected individuals. *J Med Virol* 40:285–290. <https://doi.org/10.1002/jmv.1890400406>.
112. Stemler M, Hess J, Braun R, Will H, Schroder CH. 1988. Serological evidence for expression of the polymerase gene of human hepatitis B virus in vivo. *J Gen Virol* 69:689–693. <https://doi.org/10.1099/0022-1317-69-3-689>.
113. Weimer T, Weimer K, Tu ZX, Jung MC, Pape GR, Will H. 1989. Immunogenicity of human hepatitis B virus P-gene derived proteins. *J Immunol* 143:3750–3756.
114. Weimer T, Schödel F, Jung MC, Pape GR, Alberti A, Fattovich G, Beljaars H, van Eerd PM, Will H. 1990. Antibodies to the RNase H domain of hepatitis B virus P protein are associated with ongoing viral replication. *J Virol* 64:5665–5668. <https://doi.org/10.1128/jvi.64.11.5665-5668.1990>.
115. Yuki N, Hayashi N, Kasahara A, Katayama K, Ueda K, Fusamoto H, Kamada T. 1990. Detection of antibodies against the polymerase gene product in hepatitis B virus infection. *Hepatology* 12:193–198. <https://doi.org/10.1002/hep.1840120203>.
116. Mizukoshi E, Sidney J, Livingston B, Ghany M, Hoofnagle JH, Sette A, Rehermann B. 2004. Cellular immune responses to the hepatitis B virus polymerase. *J Immunol* 173:5863–5871. <https://doi.org/10.4049/jimmunol.173.9.5863>.
117. Rehermann B, Fowler P, Sidney J, Person J, Redeker A, Brown M, Moss B, Sette A, Chisari FV. 1995. The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. *J Exp Med* 181:1047–1058. <https://doi.org/10.1084/jem.181.3.1047>.
118. Depla E, Van der Aa A, Livingston BD, Crimi C, Allosery K, De Brabandere V, Krakover J, Murthy S, Huang M, Power S, Babé L, Dahlberg C, McKinney D, Sette A, Southwood S, Philip R, Newman MJ, Meheus L. 2008. Rational design of a multi-epitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections. *J Virol* 82:435–450. <https://doi.org/10.1128/JVI.01505-07>.
119. Lin YM, Jow GM, Mu SC, Chen BF. 2013. Naturally occurring hepatitis B virus B-cell and T-cell epitope mutants in hepatitis B vaccinated children. *Sci World J* 2013:571875. <https://doi.org/10.1155/2013/571875>.
120. Donlin MJ, Szeto B, Gohara DW, Aurora R, Tavis JE. 2012. Genome-wide networks of amino acid covariances are common among viruses. *J Virol* 86:3050–3063. <https://doi.org/10.1128/JVI.06857-11>.
121. Torresi J. 2002. The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. *J Clin Virol* 25:97–106. [https://doi.org/10.1016/s1386-6532\(02\)00049-5](https://doi.org/10.1016/s1386-6532(02)00049-5).
122. Yatsuji H, Noguchi C, Hiraga N, Mori N, Tsuge M, Imamura M, Takahashi S, Iwao E, Fujimoto Y, Ochi H, Abe H, Maekawa T, Tateno C, Yoshizato K, Suzuki F, Kumada H, Chayama K. 2006. Emergence of a novel lamivudine-resistant hepatitis B virus variant with a substitution outside the YMDD motif. *Antimicrob Agents Chemother* 50:3867–3874. <https://doi.org/10.1128/AAC.00239-06>.
123. Marlet J, Lier C, Roch E, Moreau A, Combe B, Handala L, Lefeuvre S, Maugey M, Elkrief L, D'Alteroche L, Potier P, Brand D, Gaudy-Graffin C. 2021. Evolution and phenotypic characterization of whole HBV genome in compliant patients experiencing unexplained entecavir treatment failure. *Antiviral Res* 192:105106. <https://doi.org/10.1016/j.antiviral.2021.105106>.
124. Lampertico P, Agarwal K, Berg T, Buti M, Janssen HLA, Papatheodoridis G, Zoulim F, Tacke F. 2017. EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. *J Hepatol* 67:370–398. <https://doi.org/10.1016/j.jhep.2017.03.021>.
125. Downs L, Vawda S, Bester P, Lythgoe K, Wang T, McNaughton A, Smith D, Maponga T, Freeman O, Várnai K, Davies J, Woods K, Fraser C, Barnes E, Goedhals D, Matthews P. 2020. Bimodal distribution and set point HBV DNA viral loads in chronic infection: retrospective analysis of cohorts from the UK and South Africa. *Wellcome Open Res* 5:113. <https://doi.org/10.12688/wellcomeopenres.15941.2>.
126. McNaughton A, Roberts H, Bonsall D, de Cesare M, Mokaya J, Lumley S, Golubchik T, Piazza P, Martin J, de Lara C, Brown A, Ansari M, Bowden R, Barnes E, Matthews P. 2019. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Sci Rep* 9:7081. <https://doi.org/10.1038/s41598-019-43524-9>.
127. Mokaya J, McNaughton AL, Bester PA, Goedhals D, Barnes EJ, Marsden BD, Matthews PC. 2020. Hepatitis B virus resistance to tenofovir: fact or fiction? A systematic literature review and structural analysis of drug resistance mechanisms. *Wellcome Open Res* 5:151. <https://doi.org/10.12688/wellcomeopenres.15992.1>.
128. Cox AL, El-Sayed MH, Kao J-H, Lazarus JV, Lemoine M, Lok AS, Zoulim F. 2020. Progress towards elimination goals for viral hepatitis. *Nat Rev Gastroenterol Hepatol* 17:533–542. <https://doi.org/10.1038/s41575-020-0332-6>.
129. McNaughton AL, Revill PA, Littlejohn M, Matthews PC, Ansari MA. 2020. Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. *J Gen Virol* 101:271–283. <https://doi.org/10.1099/jgv.0.001387>.
130. Simmonds P. 2012. SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 5:50. <https://doi.org/10.1186/1756-0500-5-50>.
131. Lago BV, Mello FC, Kramvis A, Niel C, Gomes SA. 2014. Hepatitis B virus subgenotype A1: evolutionary relationships between Brazilian, African and Asian isolates. *PLoS One* 9:e105317. <https://doi.org/10.1371/journal.pone.0105317>.
132. Gulube Z, Chirara M, Kew M, Tanaka Y, Mizokami M, Kramvis A. 2011. Molecular characterization of hepatitis B virus isolates from Zimbabwean blood donors. *J Med Virol* 83:235–244. <https://doi.org/10.1002/jmv.21954>.
133. Tallo T, Tefanova V, Priimägi L, Schmidt J, Katargina O, Michailov M, Mukomolov S, Magnius L, Norder H. 2008. D2: major subgenotype of hepatitis B virus in Russia and the Baltic region. *J Gen Virol* 89:1829–1839. <https://doi.org/10.1099/vir.0.83660-0>.
134. Ingasia LAO, Kostaki EG, Paraskevis D, Kramvis A. 2020. Global and regional dispersal patterns of hepatitis B virus genotype E from and in Africa: a full-genome molecular analysis. *PLoS One* 15:e0240375. <https://doi.org/10.1371/journal.pone.0240375>.