Hepatitis B Virus Polymerase-Envelope Fusion Protein: Regulation of Translation, Stability, and Distribution

By
Yuzhou Wang
BS, Cornell University, 2021

Thesis
Submitted in partial fulfillment of the requirements for the Degree of Master of Science
in the Graduate Program of Biotechnology at Brown University

PROVIDENCE, RHODE ISLAND
MAY 2023

AUTHORIZATION TO LEND AND REPRODUCE THE THESIS

As the sole author of this thesis, I authorize Brown University to lend it to other institutions for the purpose of scholarly research.

Date 05/01/2023 Yuzhou Wang

Yuzhou Wang, Author

I further authorize Brown University to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Date 05/01/2023 Yuzhou Wang

Yuzhou Wang, Author

This thesis by (Yuzhou Wang) is accepted in its present form by the Graduate Program in Biotechnology as satisfying the thesis requirements for the degree of Master of Science

Date 5/1/2023 Signature: _____

Dr. Shuping Tong, Advisor

Date 5/1/2023 Signature: _____

Dr. Jisu Li, Advisor

Date 5/1/2023 Signature: _____

Prof. Andrea Sobieraj, Reader

Approved by the Graduate Council

Date _____ Signature: _____

Dr. Thomas Lewis, Dean of the Graduate School

Acknowledgments

I would like to express my sincere gratitude to my thesis advisor, Dr. Shuping Tong, and co-advisor, Dr. Jisu Li, for their guidance, encouragement, and support throughout my research. Their expertise and feedback were invaluable in helping me shape my ideas and refine my arguments. I am also grateful to the members of my thesis committee, Dr. Tong, Dr. Li, and Prof. Andrea Sobieraj, for their insightful comments and suggestions, which helped me improve my work's quality. I would like to extend my gratitude to my co-worker, Wenqing Yuan, who is also graduating this year (Congratulations!) for her collaboration throughout this project. I feel fortunate to have worked with such a supportive and knowledgeable team, and I truly appreciate all they have done to help me succeed.

I would like to thank the Biotechnology Graduate Program directed by Dr. Jacquelyn Schell at Brown University and the Liver Research Center directed by Dr. Jack R. Wands and managed by Rolf Carlson for providing me with the resources and facilities necessary to conduct my research, including access to the library, archives, and other research materials.

Finally, I want to express my appreciation to my family and friends for their unwavering support and encouragement during this process. Their love and encouragement were crucial in keeping me motivated and focused.

Table of Contents

Table of Figures

Table of Tables

Abstract of Hepatitis B Virus Polymerase-Envelope Fusion Protein: Regulation of Translation, Stability, and Distribution, by Yuzhou Wang, ScM, Brown University, May 2023

In the circular 3.2-kb genome of hepatitis B virus (HBV), the polymerase (P) gene overlaps its 5' end with 3' precore/core gene and its central part with the envelope (preS1/preS2/S) gene. Alternative translation initiation from the precore or core AUG codon generates hepatitis B e antigen (HBeAg), a secreted protein or core protein, which assembles into core particles inside cells. Similarly, translation initiated from preS1, preS2, and S region AUGs generates large (L), middle (M), and small (S) envelope proteins, with efficient secretion of S protein inhibited by L protein (containing extra preS1/preS2 domains). At the mRNA level, HBeAg and core protein are respectively translated from 3.5-kb pcRNA and slightly shorter pgRNA. pgRNA also produces low level of P protein through ribosomal leaky scanning or translational termination-reinitiation. All seven HBV proteins are translated from unspliced mRNAs. Splicing of 3.5-kb RNA enables expression of a P-L fusion protein (p43), with N-terminal 47aa from P protein joined by L protein lacking first 18aa. We found that 5' preS1 mutations to prevent L protein expression from pgRNA construct markedly increased extracellular p43, which could be further increased by a C48* nonsense mutation in the core gene. The present study further examined regulation of p43 expression, stability, and distribution. L-minus pcRNA construct produced little p43, which could be overcome by adding C48* core-minus mutation. p43 level could also be increased by other nonsense mutations in the core gene or precore region, and even by a nonsense mutation in the P gene. However, providing core protein or HBeAg in trans did not reduce p43 level. Spliced pgRNA or pcRNA construct produced much less S protein. Co-transfection with an expression construct for S protein markedly increased extracellular p43 from such constructs. In conclusion, p43 expression can be promoted by translational termination-reinitiation including from a naturally

occurring precore mutation. S protein stabilizes p43 and promotes its secretion, while L and core proteins retain p43. RNA splicing to convert P protein into P-L fusion protein (p43) also converts core protein into a core-P fusion protein, which is very likely a destabilizer of p43 and possibly also S protein.

1 Background

1.1 Viral Hepatitis

Viral hepatitis refers to a group of infectious diseases that cause inflammation of the liver. Among the five types of hepatitis virus (A-E), hepatitis A virus (HAV), hepatitis C virus (HCV), and hepatitis E virus (HEV) are RNA viruses, while hepatitis B virus (HBV) and hepatitis D virus (HDV) are DNA viruses. HAV and HEV can cause acute hepatitis through the fecal-oral route, while HBV, HCV, and HDV also cause chronic infection which leads to liver cirrhosis and liver cancer. [1,2]

HBV is highly infectious and can be transmitted through contact with infected blood or bodily fluids, such as through sexual contact, sharing of needles, or from an infected mother to her newborn during delivery. Moreover, co-infection of HDV which requires HBV as its helper virus for virion formation and transmission can increase the severity of HBV infection leading to fulminant hepatitis. Around 0.26% of the U.S. population, mostly the Asian population, live with HBV infection, and only 1/3 of the cases were reported. Preventative vaccines against HBV infection were available since the 1980s, yet less than 40% of adults were vaccinated. This unawareness of HBV prevention contributes to the prevalence of HBV hepatitis and eventually the progression to a life-long chronic infection which is the leading cause of liver cancer. [1–3]

1.2 HBV Overview

HBV is an enveloped DNA virus that infects the human liver. It is the prototype of hepatotropic DNA viruses (*Hepadnaviridae*), with other members infecting woodchucks, ground squirrels, ducks, bats, etc. The 42-nm HBV virions from blood samples were first visualized in 1970 under an electron microscope by David Dane, a British virologist [4]. Therefore, HBV virions are also referred to as Dane particles. Surprisingly, the majority of the particles under the microscope were of smaller size (22nm in diameter), which turned out to be genome-free and non-infectious subviral particles (SVPs).

1.2.1 Genome Structure and Viral Protein Expression

HBV has the smallest genome (3.2kb) among DNA viruses. This genome harbors four genes in the order of precore/core, polymerase (P), preS1/preS2/S (envelope), and X, arranged in an overlapping and circular manner (Fig. 1) [5]. The P gene is the longest (2.5kb). It overlaps its 5' end with the 3' end of the core gene, its central part with the entire envelope gene, and its 3' end with the 5' end of the X gene. However, the reading frames are different, and translated amino acids are also different from such overlapping regions. Following virion entry into hepatocytes, the partially double-stranded relaxed circular DNA (rcDNA) is converted into covalently closed circular DNA (cccDNA). The cccDNA in the nucleus is the template for the transcription of four size forms of mRNAs. These RNAs are unspliced, co-terminal, and polyadenylated because they are driven by four different promoters but share the same polyadenylation signal. [5,6]

The core promoter drives transcription of the 3.5-kb RNA (terminally redundant). Due to variable transcription start sites leading to different functions, the 3.5-kb RNAs are called precore

RNA (pcRNA; the longer version) and pregenomic RNA (pgRNA; the slightly shorter version), respectively. pcRNA has the entire precore region covered at its 5' end and is therefore used for the translation of intact precore/core protein (p25), the precursor to hepatitis B e antigen (HBeAg; p17). pgRNA lacks precore AUG at its 5' end and uses the downstream (core gene) AUG to translate core protein (p21) instead. Core protein forms homodimers for assembly into capsids (core particles), which form the venue for DNA replication. pgRNA is also used for the translation of the P protein (p93) from a downstream AUG at much-reduced efficiency, leading to about a 200:1 ratio between core and P proteins. The P protein consists of four segments in the following order – terminal protein (TP), spacer, reverse transcriptase (RT), and RNase H (RH). Inside capsids, it converts pgRNA into rcDNA to accomplish genome replication. The error-prone reverse transcription by the P protein leads to a high mutation rate of the HBV genome. Selective packaging of pgRNA but not shorter subgenomic RNAs into capsids is driven by a stem-loop structure at its 5' end called the pregenome encapsidation (ε) signal.

The three envelope proteins: L, M, and S are co-terminal and consist of preS1/preS2/S domains, preS2/S domains, and S domain, respectively. They arise by alternative translation initiation. The S domain binds to heparan sulfate proteoglycans (HSPG) on the host cell surface with a low affinity for viral docking, while the preS1 domain binds to sodium taurocholate cotransporting polypeptide (NTCP) with a high affinity for viral internalization [7]. The 2.4-kb subgenomic mRNA, driven by the SPI promoter, covers the entire envelope gene at its 5' end and expresses the L protein. The 2.1-kb mRNA, driven by the SPII promoter, is the most abundant HBV transcript. It has a heterogeneous 5' end spanning the preS2 AUG codon and hence can express either M protein or S protein. The Kozak sequence of the S gene is optimal for translation initiation in contrast to that for preS2 ATG codon, leading to much higher expression of S protein

than M protein [8]. S protein is the major envelope protein. It is also the most abundant protein produced by HBV.

The shortest also least abundant mRNA is the 0.7-kb mRNA. It is driven by the X promoter and used for the expression of HBx protein (p17).

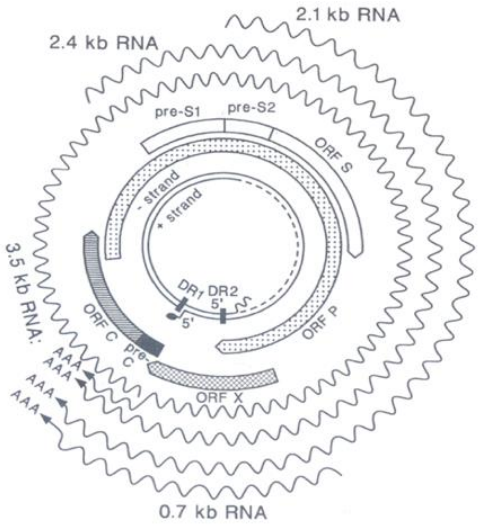

Figure 1. HBV genome, genes, and transcripts [5]***.*** *From inside out are the 3.2-kb relaxed circular HBV genome (rcDNA), the four genes (open reading frames or ORFs), and four size forms of mRNAs. Note that the longest P gene has partial or complete overlap with the other three genes. The four size forms of HBV RNAs are co-terminal, with the 3.5-kb RNA over genome length, and terminally redundant. Such HBV RNA can be transcribed from authentic cccDNA or a cloned overlength HBV genome, such as 1.1mer and 1.3mer constructs used in this work. All four size forms of HBV RNA are unspliced, leading to the expression of seven viral proteins. The 3.5-kb RNA has imprecise transcription start sites. The longer versions (pcRNA) are used for HBeAg expression, while the shorter version (pgRNA) is used for core and P protein expression as well as genome replication.*

1.2.2 HBeAg Seroconversion Selects for Precore or Core Promoter Mutations

According to sequence divergence at the nucleotide level, HBV isolates worldwide can be classified into 10 genotypes (A-J) [9,10]. Most genotypes have a size of 3,215 nucleotides (nt). Genotypes A and D co-circulate in Europe, Africa, and North America. Genotype A has a 6-nt insertion in the 3' core gene to extend core protein from 183 amino acids (aa) to 185aa, while genotype D has a 33-nt deletion at the 5' preS1 region to remove N-terminal 11aa from the preS1 domain of L protein.

Two antigens can be detected in the patient blood: HBeAg and hepatitis B surface antigen (HBsAg). A third antigen, hepatitis B core antigen (HBcAg), is detectable in infected livers. HBcAg corresponds to the core protein. HBeAg is a 17-kDa protein derived from precore/core protein (p25) by removing the N-terminal signal peptide and C-terminal arginine-rich sequence. Due to a large overlap in their primary sequences, HBeAg and HBcAg share antigenic epitopes. However, their immunological functions are the opposite. The particulate HBcAg is highly immunogenic and induces T helper 1 (Th1) cells for immune clearance of HBV-infected hepatocytes, whereas HBeAg activates T helper 2 cells (Th2 cells) to promote immune tolerance for induction of chronic HBV infection [11]. HBsAg corresponds to viral envelope proteins detected by an anti-S antibody. Due to the large excess of the 22-nm noninfectious SVPs relative to 42-nm virions (by 1,000- to 100,000-fold), HBsAg mainly reflects envelope proteins (mostly S protein) on SVPs. Similar to HBeAg, HBsAg in the form of noninfectious SVP has been proposed to induce immune tolerance.

Two seroconversion events occur during chronic HBV infection: HBeAg seroconversion (loss of HBeAg followed by the rise of anti-HBe antibody) and HBsAg seroconversion (loss of HBsAg followed by the rise of anti-HBs antibody). HBeAg seroconversion occurs first, which is

often accompanied by a reduction in viremia titer. HBeAg seroconversion may select for the G1896A nonsense mutation in the precore region (converting codon 28 from TGG to TAG) to prevent HBeAg expression at the translational level [12,13]. At the nucleotide level, that mutation converts a wobble U:G pair (between U1858 and G1896) in the ε signal of pgRNA into a more stable U:A pair. But genotype A has C1858, and the G1896A mutation would disrupt the preexisting C:G pair. That could explain the rare occurrence of G1896A HBeAg-negative mutation in genotype A. When it does occur, it is accompanied by the C1858T co-variation to maintain base pairing in the ε signal [14]. Besides or in addition to G1896A or other mutations to abolish HBeAg expression, mutations in the core promoter (A1762G/T1764A) reduce pcRNA transcription to diminish HBeAg expression but increase pgRNA transcription to enhance genome replication [15]. The phenotype (down-regulation of pcRNA but up-regulation of pgRNA) is dramatically enhanced by the extra T1753C/C1766T mutations, as we previously found in genotype A [16].

1.2.3 Regulation of P Protein Translation from pgRNA by Upstream ORFs

The core gene ATG codon lies 406nt upstream of the ATG codon of the P gene, and the 5' end of the P gene overlaps with the 3' end of the core gene by 146nt. The 3.5-kb pgRNA is unique in translating two proteins: core and P. All other HBV RNAs are monocistronic: one mRNA for one protein (production of both M and S proteins from 2.1-kb RNA is caused by sequence heterogeneity at its 5' end). Since the P gene AUG codon lies about 480nt downstream of the 5' end of pgRNA, P protein translation is inefficient. It has been reported that the majority of P protein is produced via translational termination-reinitiation, and the rest is translated by ribosomal leaky scanning of core gene AUG or backward scanning [17]. Several ORFs upstream of the P gene start codon can regulate P protein translation. uORF of 19 codons spans the 3' precore region and 5'

core gene. It negatively regulates core protein expression while promoting translation reinitiation from the downstream P gene start codon [18,19]. J ORF of 7 codons resides in the middle of the core region and about 120nt upstream of P gene AUG. Translation of J ORF bypasses an internal core gene ATG codon (M93) of favorable Kozak sequence to promote translational termination-reinitiation for P protein expression [17,20]. A 36-nt insertion at the 5' core gene of genotype G creates a hairpin structure 14nt downstream of core gene AUG to significantly increase core protein translation, which leads to reduced P protein translation [21–23].

1.2.4 Cell Culture Systems for Functional Characterization of HBV Biological Properties

Two human liver cancer cell lines, HepG2 and Huh7, are commonly used for transient transfection with cloned HBV genome to characterize viral protein expression, genome replication, and virion production. In this regard, Huh7 cells have higher transfection efficiency than HepG2 cells. Since the 3.5-kb pgRNA required for genome replication is more than the genome length and terminally redundant, HBV genome replication and virion production require transfection with a cccDNA equivalent, such as a cloned tandem dimer or 1.3mer construct. Alternatively, properly designed 1.1mer constructs (with 1.1 copies of HBV genome co-linear with pgRNA cloned downstream of a strong exogenous promoter) can produce a high level of pgRNA to permit robust genome replication and virion production. [24,25]

1.3 p43, a P-L Fusion Protein Generated by Splicing of the 3.5-kb RNA

RNA splicing is important in augmenting eukaryotic gene expression efficiency, and during viral infection, the host splicing machinery is often hijacked by viruses for the same purpose. However, all the seven known HBV proteins: L, M, S, HBeAg, core, P, and HBx, are translated from unspliced HBV RNAs. It has been reported that the 3.5-kb HBV RNA can be spliced into over 20 variants, although their functional consequences are mostly unknown [26]. In 2000 by transfecting Huh7 cells with a 1.1mer pgRNA construct followed by a highly sensitive detection method, Huang et al discovered a truncated P protein of 43kd (p43) [27]. p43 was 10 times more abundant than the full-length P protein of 93kd. It was associated with core particles in cell lysate just like full-length P protein, although in culture supernatant p43 was associated with SVPs and virions. Further analysis revealed p43 to be a translation product of spliced 3.5-kb RNA that deletes nt2454-2940 (numbering according to genotype A) (Fig. 2). The 487-nt deletion causes a frameshift, thus joining the first 147nt of the P gene with the preS1 region lacking the first 87nt. Consequently, for genotype A**,** p43 is a P-L fusion protein with N-terminal 49aa of P protein fused to L protein lacking its first 29aa. In the case of genotype D, which has a 6-nt deletion in the 3' core gene/5' P gene and a 33-nt deletion at 5' preS1 region relative to genotype A, p43 has the N-terminal 47aa of P protein fused to L protein missing its first 18aa.

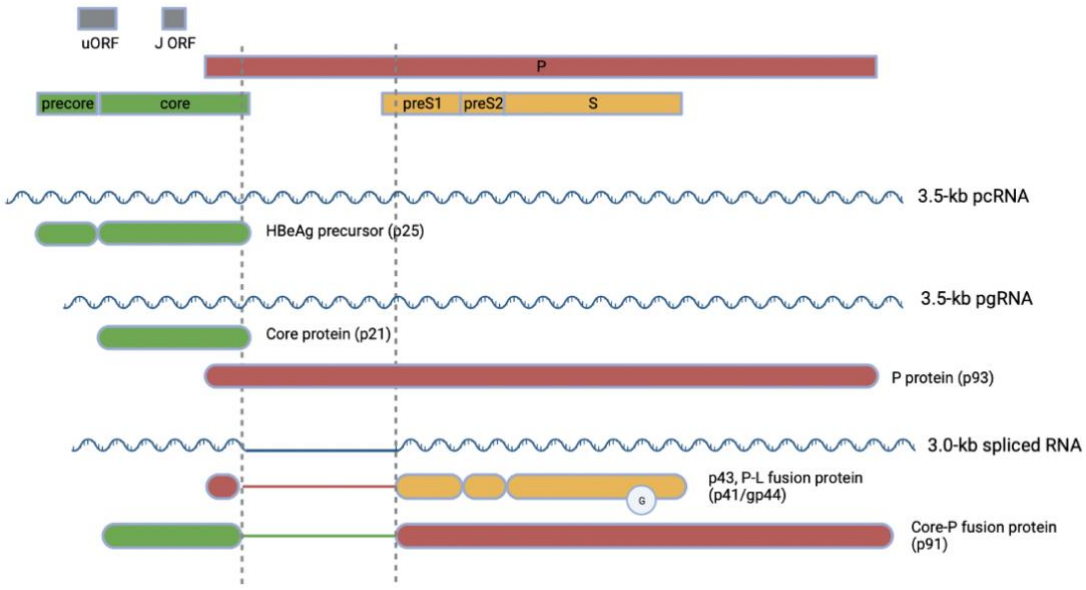

Figure 2. Splicing of the 3.5-kb pgRNA or pcRNA will alter the coding capacity to generate p43, the P-L fusion protein (Created with BioRender.com). *The 3.5-kb pcRNA or pgRNA has all the genetic information of the 3.2-kb HBV genome. However, pcRNA, with entire precore region covered at its 5' end, is only used for translation of precore/core protein (the precursor to HBeAg). pgRNA lacks precore AUG at its 5' end and is primarily used for translation of core protein. A small fraction is used for translation of P protein through ribosomal leaky scanning or translation termination-reinitiation. Deletion of a 487-nt intron (in the case of genotype D) will fuse 5' P gene with the nearly intact envelope gene to express a P-L fusion protein, or p43. It also fuses nearly intact core gene (or precore/core gene in the case of pcRNA) with the 3' 2/3rds of P gene, although such a core-P (or precore/core-P) fusion protein has not been detected.*

1.3.1 5' preS1 Mutations to Prevent L Protein Expression Markedly Increased Extracellular p43

Due to a facultative N-linked glycosylation site in the S domain, the three envelope proteins exist in two size forms: p39/gp42 (L), gp33/gp36 (M), and p24/gp27 (S). We recently attempted

to test the impact of lost L protein expression from the 1.1mer pgRNA construct of genotype A and genotype D by mutating the preS1 ATG codon or by introducing a nonsense mutation at the 5' preS1 region [25]. Surprisingly, Western blot analysis of extracellular L protein by an anti-preS1 antibody rather found a slight upshift of L protein from a doublet of 39kd and 42kd to a doublet of 41kd and 44kd. The level of the 41kd/44kd proteins could be further increased to the level of p39/gp42 from wild-type virus when core protein expression was also abolished (by a C48* nonsense mutation). The 41kd/44kd proteins were also detected from cells transfected with a 1.3mer construct, for which both the 3.5-kb pgRNA and pcRNA were transcribed at their physiological levels through the endogenous core promoter. While p39/gp42 from the genotype D clone migrated slightly faster than the doublet from genotype A due to an 11-aa deletion in the preS1 domain, the doublet of 41kd/44kd from the two HBV genotypes did not show size differences. That feature is consistent with the previously reported p43: while for genotype D p43 is a P-L fusion protein with N-terminal 47aa of P protein fused to L protein lacking its first 18aa, for genotype A p43 has the N-terminal 49aa of P protein fused to L protein missing its first 29aa. Further study confirmed p44 to be the glycosylated form of p41 and moving the preS1 nonsense mutation further downstream to be inside the p43 coding sequence eliminated 41kd/44kd production. Most strikingly, the p41/gp44 doublet could be detected by a monoclonal antibody against the N-terminus of the P protein and abolished by a single-point mutation at the splicing acceptor site for p43. Co-transfecting the p43 producing 1.1mer construct with a 0.7mer expression construct for L/M proteins markedly reduced extracellular p43, which was accompanied by increased intracellular p43 if the 1.1mer construct was both L-minus and core-minus. Therefore, preventing L protein expression promoted p43 secretion and possibly stabilization.

1.3.2 Rationales, Objectives, and Significance of the Present Study

The previous report and most of our recent study were based on 1.1mer constructs to overproduce pgRNA, with lost pcRNA transcription. Although the P protein is preferentially translated from pgRNA, whether the P-L fusion protein (p43) is also better expressed from spliced pgRNA than spliced pcRNA remained an open question. Therefore 1.1mer pcRNA construct will be generated for comparison. Even if p43 is preferentially expressed from spliced pgRNA construct from wild-type HBV, naturally occurring G1896A precore mutation might trigger translational termination (at 3' precore region) – reinitiation (at P gene ATG) to render spliced pcRNA an additional source for p43 expression. That is also relevant to how the C48* nonsense mutation in the core gene increases the p43 protein level: one possibility is translational termination-reinitiation. To directly compare the efficiency of p43 expression from spliced pgRNA vs. spliced pcRNA without complication from the efficiency of splicing, we will generate spliced pgRNA construct and spliced pcRNA construct by deleting the 487-nt intron from the corresponding 1.1mer constructs. That also removes the 5' preS1 region where the L-minus mutations were introduced in our initial study. If p43 could be produced from such spliced pgRNA or pcRNA construct, then artificial L-minus mutation is unnecessary for p43 production.

The 1.1mer constructs are artificial in that pgRNA or pcRNA is driven by a strong exogenous promoter. To mimic natural conditions, the same set of mutations will be examined in the context of a 1.3mer construct, where endogenous core promoter drives transcription of both pcRNA and pgRNA at their proper ratio. Considering that T1753C/A1762G/T1764A/C1766T quadruple core promoter mutation drastically increases pgRNA transcription at the expense of pcRNA [16], they can be used to further verify the contribution of pcRNA vs. pgRNA in p43

translation (following splicing). We will verify whether p43 translation is similar to P protein translation in being regulated by upstream small ORFs.

Taken together, the different types of constructs used (1.1mer pgRNA, 1.1mer pcRNA, and their spliced version; 1.3mer construct, and its spliced version) and different mutations introduced will clarify the regulatory mechanism for p43 translation, stability, and distribution (intracellular retention vs. secretion). p43 represents the 8th HBV protein and is the only HBV protein produced by RNA splicing. If naturally occurring precore and/or core promoter mutations can up regulate p43 expression, then our studies also have clinical relevance. Such mutations have been associated with liver cancer and fulminant hepatitis [6,9]. Our studies may also shed light on the extremely low level of P protein associated with HBV infection: is that indeed caused by poor translation or also attributed to degradation by an unknown mechanism?

2 Methods

2.1 HBV DNA constructs

Table 1 summarizes the different types of HBV DNA constructs used in the present study. These constructs were based on geno5.4 (GenBank accession number KX827293), a genotype A clone, and geno1.2 (GenBank accession number KX827290), a genotype D clone [28]. These two HBV clones were used in our initial characterization of p43 from their L-minus mutants [25]. The 1.1mer pgRNA construct has nt1805-3221/1-1932 of geno5.4 and nt1805-3182/1-1932 of geno1.2 inserted to SacI –HindIII sites of pcDNA3.1 Zeo(-) vector [24,25], which enabled efficient pgRNA transcription under the strong CMV promoter. The vector has an ampicillin resistance gene for transformation into *E. coli* and plasmid DNA preparation. In the present study, we converted 1.1mer pgRNA construct into the 1.1mer pcRNA construct by extending the 5' end of the HBV DNA inserts to nt1775. The 1.3mer construct has been described [25]. It has nt1031-3221/1-1932 of geno5.4 and nt1031-3182/1-1932 of geno1.2 inserted into the MfeI-SacI sites of pcDNA3.1 Zeo(-) vector. That removed a circa 0.8-kb vector sequence upstream of the multiple cloning site (MCS) including the CMV promoter, and the extra 5' HBV sequence permitted transcription of both pcRNA and pgRNA under the endogenous core promoter. The spliced 1.1mer pgRNA construct, spliced 1.1mer pcRNA construct, and spliced 1.3mer construct were generated by deleting nt2454-2940 from geno5.4, or nt2448-2901 from geno1.2. The CMV-p43 construct was derived from a 1.1mer spliced pgRNA construct by moving the 5' end of HBV insert from nt1805 to nt2193, thus rendering the P gene AUG codon (nt2307-2309) much closer to the 5' end of the mRNA for p43 translation.

2.2 Site-directed mutants

Table 2 lists the mutants generated in the present study. Naturally occurring mutations included core promoter mutations (CPMs): the common A1762T/G1764A double mutation (dCPM) and less common T1753C/A1762T/G1764A/C1766T quadruple mutation (qCPM) [16]. Precore nonsense mutations to prevent HBeAg expression include A1814G (Q2*) (called PC-1 here) and G1896A (W28*) (called PC-2 here) [12-14]. The latter was accompanied by C1858T co-variation to maintain base pairing of the ε signal in genotype A (geno5.4) [14]. The "C+" mutation in the core gene was a 36-nt insertion found in genotype G that markedly enhanced core protein translation from pgRNA [21–23]. Artificial mutations included PCM1 and CM1, which prevented translation initiation from the precore or core gene ATG codon by its mutation to GTG. Alternatively, translation from the core gene or precore/core gene was terminated by C-1, C-, C-2, C-3, Q179*, or Q184* nonsense mutations introduced to codon 10, 48, 71, 119, 179, or 184 in the core gene. C2961T (Q219*) was used as the P- mutation to truncate the P protein (and putative core-P fusion protein), whereas the op43 mutations optimized the Kozak sequence for the P gene ATG codon in the CMV-p43 construct. Alternatively, the Kozak sequence was optimized for uORF (u+) or J ORF (J+1 and J+2), whereas the ATG codon for J ORF was mutated to ACG to prevent its translation (J-). The L- mutations converted preS1 codons 12 and 23 of geno5.4 into stop codons [25]. S1, S2, S3, and S4 were used to prevent the expression of full-length S protein by mutating its ATG codon into GCG, ATA, AAG, and TTG, respectively [24].

2.3 Construct making or site-directed mutagenesis

2.3.1 Digestion of parental construct

The original constructs were digested by High-Fidelity Restriction Endonucleases (New England Biolabs, USA). Every 1ug DNA was digested by 10 units of restriction enzymes in a 50ul total reaction volume with an ideal incubation temperature (usually 37°C) for 2 hours. Usually two restriction enzymes, each with a single cleavage site in the plasmid DNA, were used to generate two DNA fragments, with one fragment for replacement.

2.3.2 Polymerase chain reaction (PCR)

PCR was used to amplify desired HBV DNA fragments using Q5® High-Fidelity DNA Polymerase (New England Biolabs, USA). Alternatively, overlap extension PCR was used to introduce mutations or deletions. In that case, the mutagenic antisense primer was paired with an upstream sense primer to generate a PCR product with mutations incorporated into the end of the PCR product. In parallel the mutagenic sense primer was paired with a downstream antisense primer to generate a PCR product with mutations at the beginning. The two PCR products were gel purified and mixed at a 1:1 molar ratio, followed by another round of PCR amplification using the upstream sense primer and downstream antisense primer. The overlap between the original two PCR fragments permitted correct joining followed by extension into a larger fused fragment, with mutations now in the interior of the final PCR product.

2.3.3 Agarose gel electrophoresis and DNA purification

All digested plasmids and PCR products were verified and purified through agarose gel electrophoresis (0.8% agarose, 1ug/ml ethidium bromide) and with QIAquick Gel Extraction Kit (QIAGEN, USA). In the case of the digested plasmid DNA, often the smaller fragment corresponding to the PCR product was discarded while the large fragment was purified for ligation with the PCR product to reconstitute the intact full-length construct, but now with mutations introduced into the region of the smaller fragment.

2.3.4 Ligation or assembly

The digested plasmid DNA (original wild-type clone) and the PCR fragments were either ligated with T4 DNA Ligase (Roche, USA) or assembled with NEBuilder® HiFi DNA Assembly Cloning Kit (New England Biolabs, USA). For the ligation approach, the PCR fragments were also digested with two different restriction enzymes and extracted with the same procedure as the correlated original constructs. The digested plasmid DNA and the digested PCR fragment (molar ratio 1:2) were ligated by 5 units of T4 DNA ligase in a 10ul total reaction volume at 16°C overnight. For assembly, the digested construct and PCR fragments (molar ratio 1:2) were assembled with 5ul HiFi DNA Assembly Master Mix in a 10ul total reaction volume at 50°C for 2 hours.

2.3.5 Transformation and plasmid miniprep

The ligated or assembled products were transformed into *E. coli* competent cells (ratio 1:2) for selection and replication of new constructs. The competent cells were made by Dr. Jisu Li,

stored at -80^{0}C, and thawed on ice before use. The mixture of DNAs and competent cells were incubated on ice for 30min, heat shocked in a 37°C water bath for 1min, chilled on ice for 2min, and then grew in Luria-Bertani (LB) broth in a 37°C shaker at 200rpm for 30min. The cells were spread on LB agar plates with 100ug/ml ampicillin and incubated at 37°C overnight for drug selection. Single colonies were selected to grow in LB broth with 100ug/ml ampicillin in a 37°C shaker at 200rpm overnight for amplification. Plasmid DNAs were extracted with QIAprep Spin Miniprep Kit (QIAGEN, USA) and eluted in TE buffer (10 mM Tris, 1 mM EDTA).

2.3.6 DNA sequencing

For each new construct, the region replaced by restriction enzyme digestion or through DNA assembly was sequenced to confirm the introduction of desired mutations and lack of PCR errors. Sequencing service was provided by The W.M. Keck Biotechnology Resource Laboratory affiliated with Yale University.

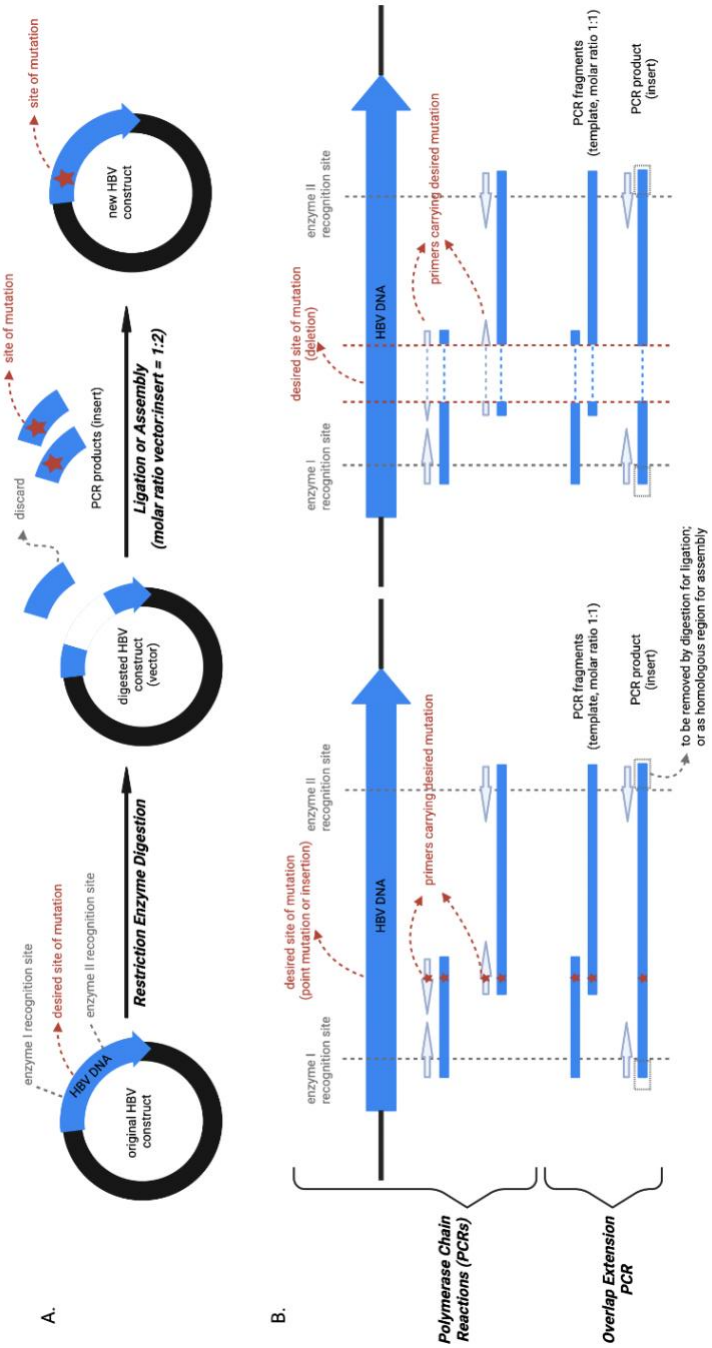

Figure 3. Site-directed mutagenesis (Created with BioRender.com). A. Generation of new constructs. *The site of mutation or replacement in the original HBV construct was removed by restriction enzyme digestion. PCR product carrying desired mutations or region of the HBV genome (insert) was ligated or assembled with the digested HBV construct (vector) at a 2:1 ratio to generate the new HBV construct.* ***B. Generation of PCR products carrying desired mutations (left: point mutation or insertion; right: deletion).*** *The original HBV*

construct was used as the template for PCR. Two sets of primers were used to generate two PCR fragments with the 3' end of the upstream fragment overlapping with the 5' end of the downstream fragment. The reverse primer of the upstream fragment and the forward primer of the downstream fragment were also overlapped and carried the desired mutation. Then a new round of PCR was performed using the two outer primers and the two fragments from the first PCRs as templates to generate a single fragment with internal mutations to serve as the insert for replacement. The product of overlap extension PCR was further digested with restriction enzymes if used for cloning by ligation.

Table 1. Type of constructs used in the present study

construct	**explanation**	**geno5.4 insertion** *(unless specified as geno1.2)*	**sites on pcDNA3.1 Zeo(-) vector**
1.1mer pgRNA	pgRNA driven by CMV promoter	nt1804-3221/1-1932 nt1804-3182/1-1932 (geno1.2)	SacI-HindIII
1.1mer pcRNA	pcRNA driven by CMV promoter	nt1775-3221/1-1932 nt1775-3182/1-1932 (geno1.2)	SacI-HindIII
1.3mer	pgRNA and pcRNA driven by endogenous core promoter	nt1031-3221/1-1932 nt1031-3221/1-1932 (geno1.2)	MfeI-HindIII
1.1mer spliced pgRNA (spgRNA)	spliced pgRNA driven by CMV promoter	nt1804-2453/2941-3221/1-1932 nt1804-2447/2902-3182/1-1932 (geno1.2)	SacI-HindIII
1.1mer spliced pcRNA (spcRNA)	spliced pcRNA driven by CMV promoter	nt1775-2453/2941-3221/1-1932 nt1775-2447/2902-3182/1-1932 (geno1.2)	SacI-HindIII
1.3mer spliced RNA	spliced pgRNA and pcRNA driven by endogenous core promoter	nt1031-2453/2941-3221/1-1932	MfeI-HindIII
CMV-p43	a 0.4-kb deletion from spliced pg/pcRNA, for p43 expression driven by CMV promoter	nt2193-2453/2941-3221/1-1932	SacI-HindIII
CMV-C	Core gene behind CMV promoter	nt1877-2480	NheI-HindIII
CMV-PC	Precore gene behind CMV promoter	nt1788-2480	NheI-HindIII

Table 2. Site-directed mutants used in the present study

region or site of mutation	name	geno5.4 mutations *(unless specified as geno1.2)*	codon change or detailed mutations
core promoter	dCPM	A1726T/G1764A	-
	qCPM	T1753C/A1726T/G1764A/C1766T	-
precore region	PCM1	A1814G	M1V (ATG->GTG)
	PC-1	C1817T	Q2* (CAA->TAA)
	PC-2	C1858T/G1896A	P15P (CCC->CCT), W28* (TGG->TAG)
core gene	CM1	A1901G	M1V (ATG->GTG)
	(DM1)	(A1814G/ A1901G)	(PCM1/CM1)
	C+	1905/ TAGAACAACTTTGCCATATG GCCTTTTTGGCTTAGA/1906	insertion found in genotype G
	C-1	G1928T	G10* (GGA->TGA)
	C-	C2044A	C48* (TGC->TGA)
		T2044A (geno1.2)	C48* (TGT->TGA)
	C-2	G2112A	W71* (TGG->TAG)
	C-3	T2256A	L119* (TTG->TAG)
	Q179*	C2435T	Q179* (CAA->TAA)
	Q184*	C2450T	Q184* (CAA->TAA)
uORF	u+	T1849G/A1850C	Kozak optimization TACATGT->GCCATGT
J ORF	J-	T2164C	M1S (ATG->ACG)
	J+1	T2161C/T2162C	Kozak optimization ATTATGT->ACCATGT
	J+2	T2161C/T2162C/T2166G	Kozak optimization TTATGT->ACCATGG
P gene	op43	A2297C/A2299C/C2301G/A2302 C/C2304A/A2305C/A2306C	Kozak optimization AGACCACCAAATGC->CGCCGCCACCATGC
	P-	C2961T	Q219* (CAA->TAA)
preS1 region	L-	A2887T/T2888A/G2920T	M12* (ATG->TAG), G23* (GGA->TGA)
		C2854T (geno1.2)	Q3* (CAG->TAG)
S region	S1	A155G/T156C	M1A (ATG->GCG)
	S2	G157A	M1I (ATG->ATA)
	S3	T156A	M1K (ATG->AAG)
	S4	A155T	M1L (ATG->TTG)

2.4 Transient transfection

The human hepatoma cell line Huh7 was obtained from ATCC. They were cultured at 37℃ with 5% CO_2, in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2% L-glutamine, and 1% penicillin/streptomycin. Cells were seeded in 6-well plates and transfected the next day with 1ug/well miniprepped plasmid using *Trans*IT®-LT1 Transfection Reagent (Mirus Bio LLC, USA). The pmaxGFP expression plasmid for green fluorescent protein (GFP) was transfected in parallel to check for transfection efficiency. The culture supernatant containing transfection DNAs and reagent was washed off the next day. GFP was visualized under fluorescent microscopy three-day after transfection (Fig. 4). On day 4 post-transfection, cells and culture supernatant were harvested.

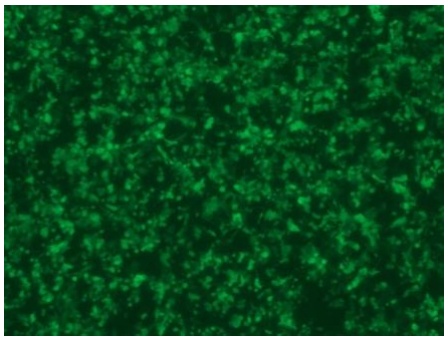

Figure 4. GFP fluorescence on day 3 shows 90% transfection efficiency. *Cell confluence was at 100% on day 3, thus the percentage of area covered by green fluorescence represents the percentage of cells taking up transfected DNA.*

2.5 Cell lysis and PEG precipitation of culture supernatant

Cell pellets were lysed in the lysis buffer (0.5% NP40, 20mM HEPES, pH 7.4) on ice for 10min. The supernatant post centrifugation was collected as cell lysates, and 1/10 of the lysates

were used for Western blotting. Polyethylene glycol (PEG) 8000 was dissolved in PBS buffer (2mM sodium phosphate monobasic, 8mM sodium phosphate dibasic, 0.9% NaCl) and was added to the culture supernatant with a final concentration of 10% PEG for incubation overnight. After centrifugation, the pellets were dissolved in TN buffer (20mM Tris pH7.5, 150mM NaCl) supplemented with 0.5‰ beta-mercaptoethanol. An equivalent to 1/15 volume of original culture supernatant was used for Western blotting.

2.6 Western blotting

0.1% SDS-10-12% polyacrylamide gel electrophoresis (PAGE) was used for protein separation according to size, followed by semi-dry transfer. The anti-preS1 mouse monoclonal antibody (mAb) 7H11 (0.7ug/ml) (targeting aa21-47) was used for detection of L protein or p43 [29,30]; anti-S mouse mAb 45E9 (0.7ug/ml) (targeting aa27-79) was used for probing the S protein [31]; mouse anti-core mAb 2A7 (0.6ug/ml) (targeting aa141-154) was used to detect core protein [31]. The housekeeping protein Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) was probed with MAB374 (0.5ug/ml) (MilliporeSigma, USA) to serve as loading control.

2.7 Enzyme-linked immunosorbent assay (ELISA)

The HBeAg and HBsAg levels in culture supernatant and cell lysates were determined with KHB ELISA HBV Test Kit (Kehua, China) after proper dilution to avoid signal saturation.

3 Results

3.1 L-minus mutations alone failed to confer p43 expression from 1.1mer pcRNA construct of geno5.4

To test whether pcRNA can serve as the source for p43 expression, we generated 1.1mer pcRNA construct of geno5.4, the genotype A clone (see Table 1 for details). Indeed, this construct released very high level of HBeAg in contrast to its 1.1mer pgRNA construct (Fig. 5A, compare lanes 1 & 6). However, p43 was undetectable from cell lysate or culture supernatant of Huh7 cells transfected with its L- mutant in contrast to the L- mutant of the pgRNA construct (Fig. 5C, lanes 2 & 7). Further addition of the core- mutation (C48*) rendered p43 detectable from culture supernatant (but not cell lysate) for the pcRNA construct while increasing its signal for the pgRNA construct (Fig. 5C, lanes 3 & 8). But in the case of geno1.2, the genotype D clone, L- mutation generated extracellular p43 from both the 1.1mer pgRNA construct and 1.1mer pcRNA construct (Fig. 6C, lanes 4 & 9).

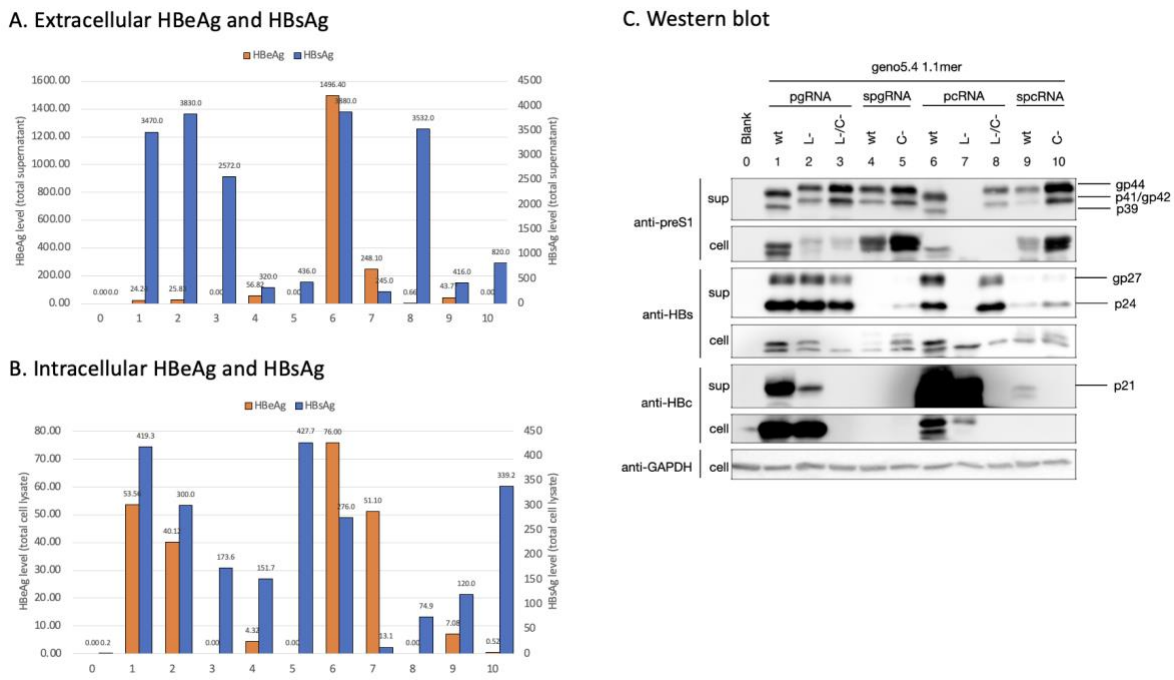

Figure 5. L-minus mutations alone failed to confer p43 expression from 1.1mer pcRNA construct of geno5.4 (genotype A), but bypassing RNA splicing enabled p43 expression from both pgRNA construct and pcRNA construct. *The 1.1mer pcRNA construct was derived from the 1.1mer pgRNA construct by extending the 5'end from nt1804 to nt1775. The spliced pgRNA (spgRNA) and spliced pcRNA (spcRNA) constructs were made by removing nt2454-2940.* ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. PEG precipitated culture supernatant was used for loading. GAPDH was used as loading control.*

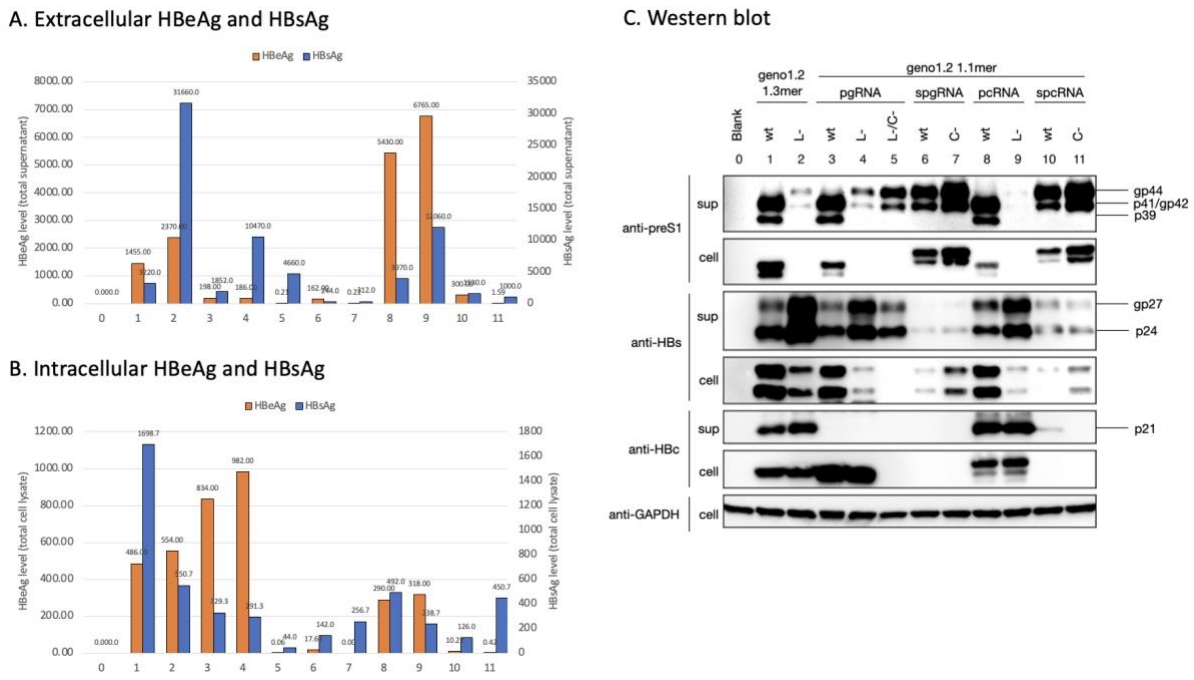

Figure 6. p43 was expressed at a low level from L-minus 1.1mer pcRNA construct of geno1.2 (genotype D), and bypassing RNA splicing increased p43 expression from both pgRNA construct and pcRNA construct. *The 1.1mer pcRNA construct was derived from the 1.1mer pgRNA construct by extending the 5'end from nt1804 to nt1775. The spliced pgRNA (spgRNA) and spliced pcRNA (spcRNA) constructs were made by removing nt2448-2901.* ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. GAPDH was used as loading control.*

3.2 Bypassing RNA splicing enabled p43 expression from 1.1mer pcRNA construct as well

To directly compare spliced pcRNA with spliced pgRNA for p43 expression, we removed the 487-bp intron (nt2454-2940) from both 1.1mer pcRNA construct and 1.1mer pcRNA construct

of geno5.4 (Table 1). The deletion covers the 5' preS1 region where the L- mutations enabled robust p43 production from the 1.1mer pgRNA construct and 1.3mer construct. The deletion would fuse the nearly intact precore/core ORF (for pcRNA construct) or core ORF (for pgRNA construct) with 3' two/thirds of P ORF (Fig. 2). Consequently, HBeAg was nearly lost from culture supernatant for the spliced pcRNA construct (Fig. 5A, compare lanes 6 & 9) while the 21-kd core protein was no longer detectable in cell lysate or culture supernatant for the spliced pgRNA construct (Fig. 5C, compare lanes 1 & 4). p43 could be detected from cell lysate and culture supernatant for both spliced pcRNA construct and spliced pgRNA construct, although the extracellular level not much higher than L- pgRNA construct (Fig. 5C, lanes 2, 4, & 9). Therefore, if RNA splicing occurs, p43 production does not require the artificial introduction of L- mutation. Adding the core- mutation further increased p43 level, more striking for extracellular p43 for the spliced pcRNA construct but intracellular p43 for the spliced pgRNA construct (Fig. 5C, lanes 5 & 10). For the 1.3mer construct, which produces both pcRNA and pgRNA at proper ratio, skipping splicing increased extracellular p43 and made it detectable in cell lysate (Fig. 7C, lanes 2 & 3). For geno1.2 of genotype D, skipping RNA splicing markedly increased intracellular and extracellular p43 in comparison with the original 1.1mer construct, with further increase by the core- mutation (Fig. 6C).

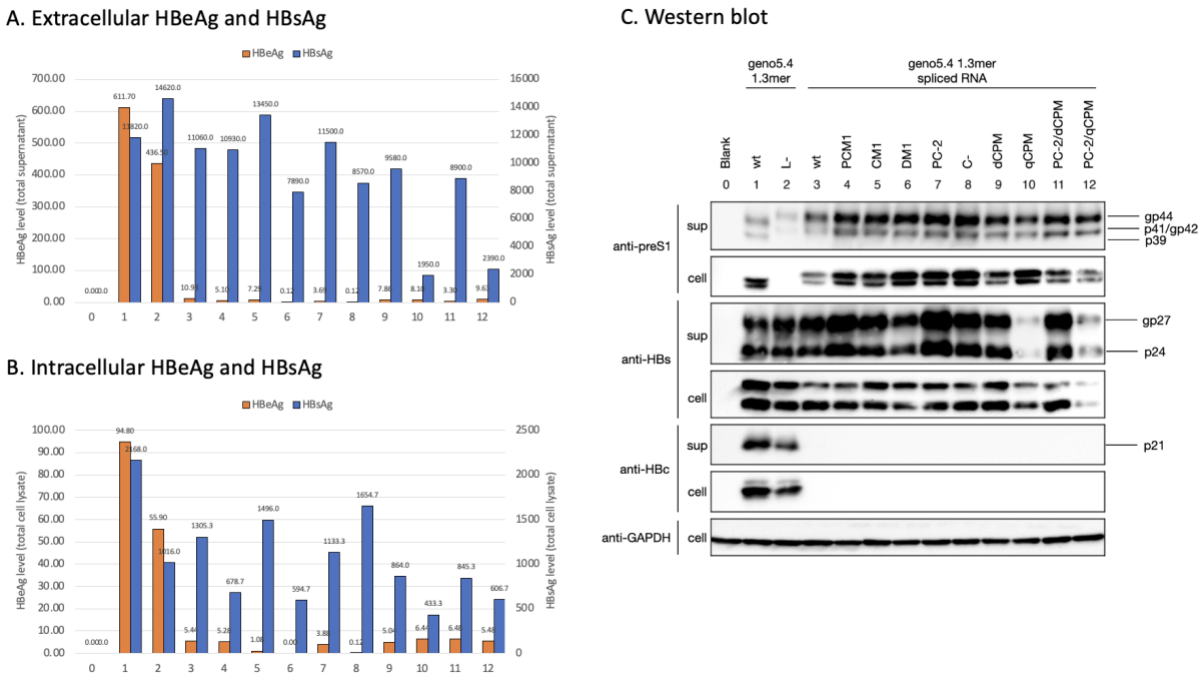

Figure 7. p43 protein level from 1.3mer construct was markedly increased by skipping splicing and could be further increased by abolishing precore and/or core protein expression. *The 1.3mer construct was derived from the 1.1mer pgRNA construct by extending the 5'end from nt1804 to nt1031. The spliced RNA construct was made by removing nt2448-2901. Precore and/or core protein expressions were abolished by either ATG to non-ATG mutations or nonsense mutations.* ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. GAPDH was used as loading control.*

3.3 Providing core protein *in trans* or increasing its expression *in cis* did not reduce p43 protein level

The C48* core- mutation increased p43 protein level for all the six types of constructs tested: 1.1mer pgRNA or pcRNA construct (Fig. 5C), spliced pgRNA or spliced pcRNA construct (Fig. 5C), L- 1.3mer construct (Fig. 8C) or its spliced version (Fig. 7C). That could be attributed to more p43 expression through translational termination – reinitiation from the spliced pgRNA or pcRNA. Alternatively, p43 might be destabilized by core or precore/core protein, or by core-P or precore/core-P fusion protein, but their expression could be terminated by the C48* nonsense mutation. To test the destabilizing effect of precore/core or core protein, their expression construct was co-transfected with the L- 1.3mer construct of geno5.4 with or without the C48* core-mutation. The CMV-PC construct secreted high level of HBeAg (Fig. 8A, lane 9), while the CMV-C construct produced core protein (Fig. 8C, lane 12). Neither construct reduced extracellular p43, with CMV-C construct significantly increasing intracellular p43 (Fig. 8C, compare lanes 2, 4, 7, 8, 10, & 11). In another approach we introduced double or quadruple core promoter mutations (dCPM or qCPM) into the L- 1.3mer construct, which should increase pgRNA transcription while diminishing pcRNA transcription. Consistent with our previous report [16], intracellular core protein was increased by dCPM and much more by qCPM (Fig. 9C, compare lanes 2-4). While extracellular p43 was elevated by both dCPM and qCPM, only qCPM markedly increased intracellular p43. Taken together, our results argue against p43 destabilization by core or precore/core protein, but rather suggest p43 retention by core protein. In the spliced 1.3mer construct, dCPM and qCPM mutations increased intracellular but not extracellular p43 (Fig. 7C, lanes 3, 9, & 10).

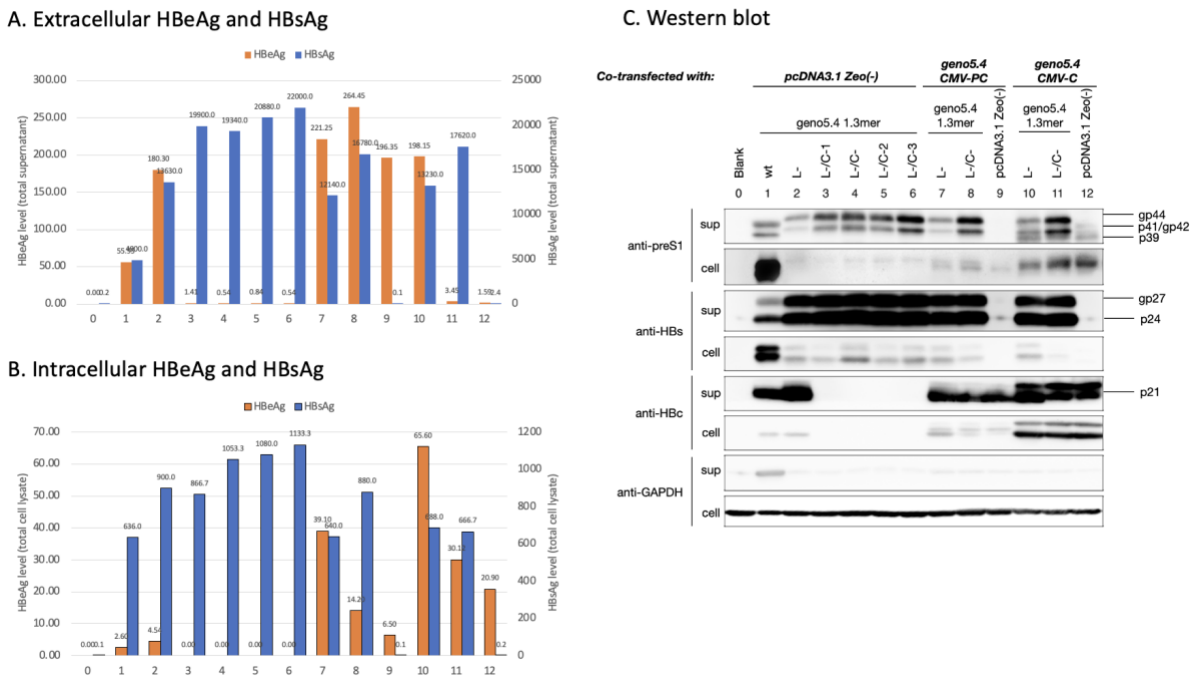

Figure 8. Providing core protein in trans or increasing its expression in cis did not reduce p43 protein level. *The 1.3mer constructs were co-transfected with a precore protein construct (geno5.4 CMV-PC) or a core protein construct (geno5.4 CMV-C). All constructs were based on pcDNA3.1 Zeo(-) which was used as the control for co-transfection.* ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. GAPDH (cell) was used as loading control. Its presence in culture supernatant (sup) indicates bacteria contamination.*

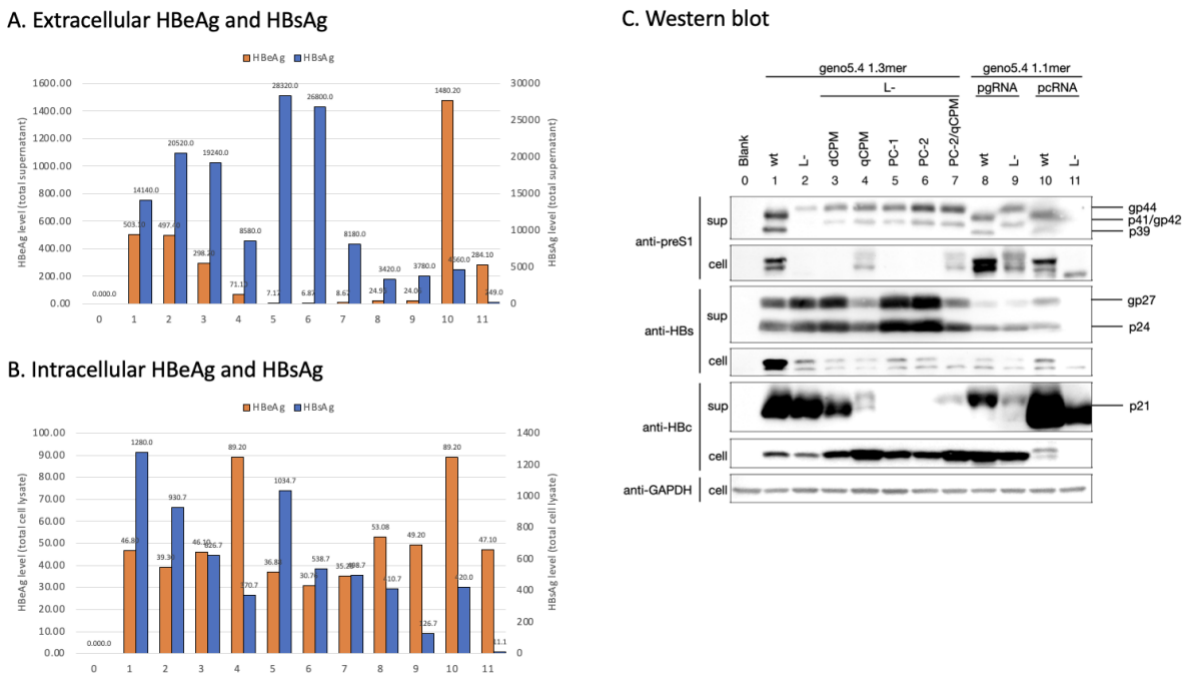

Figure 9. Core promoter mutations (CPMs) and/or precore (PC) mutations increased p43 expression from 1.3mer construct. A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA. *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. GAPDH was used as loading control.*

3.4 p43 translational control: impact of enhancing or preventing translation from two upstream minicistrons or core gene

The uORF spanning 3' precore region and 5' core gene has been identified as a negative regulator of core protein translation [18,19], whereas J ORF within the core gene serves as a positive regulator of P gene translation [17,20]. From the L- 1.1mer pgRNA construct of geno5.4, both intracellular and extracellular levels of p43 were increased by optimizing the Kozak sequences of uORF (u+) or J ORF (J+1 and J+2) but decreased by mutating the ATG codon for J ORF into ACG

(J-; Fig. 10C, compare lane 2 with lanes 4-7). The 36-nt insertion in the 5' core gene of genotype G adds 12aa to core protein. It creates a hairpin structure to dramatically enhance core protein translation from pgRNA, which diminishes translation initiation from the downstream P gene [21–23]. When introduced into L- 1.1mer pgRNA construct, the insertion (C+) increased extracellular level of elongated core protein as expected (Fig. 10C, compare lanes 2 and 3). p43 level was rather reduced in both cell lysate and culture supernatant. From the spliced 1.3mer construct, extracellular p43 was significantly increased by mutating the core gene ATG into GTG, with weaker effect from mutating the precore ATG into GTG (Fig. 7C, compare lanes 3-6). Therefore, p43 expression is diminished by translation of the upstream core gene but can be augmented by translation of the uORF (to bypass core gene AUG codon) and J ORF (to bypass an internal AUG codon in core gene).

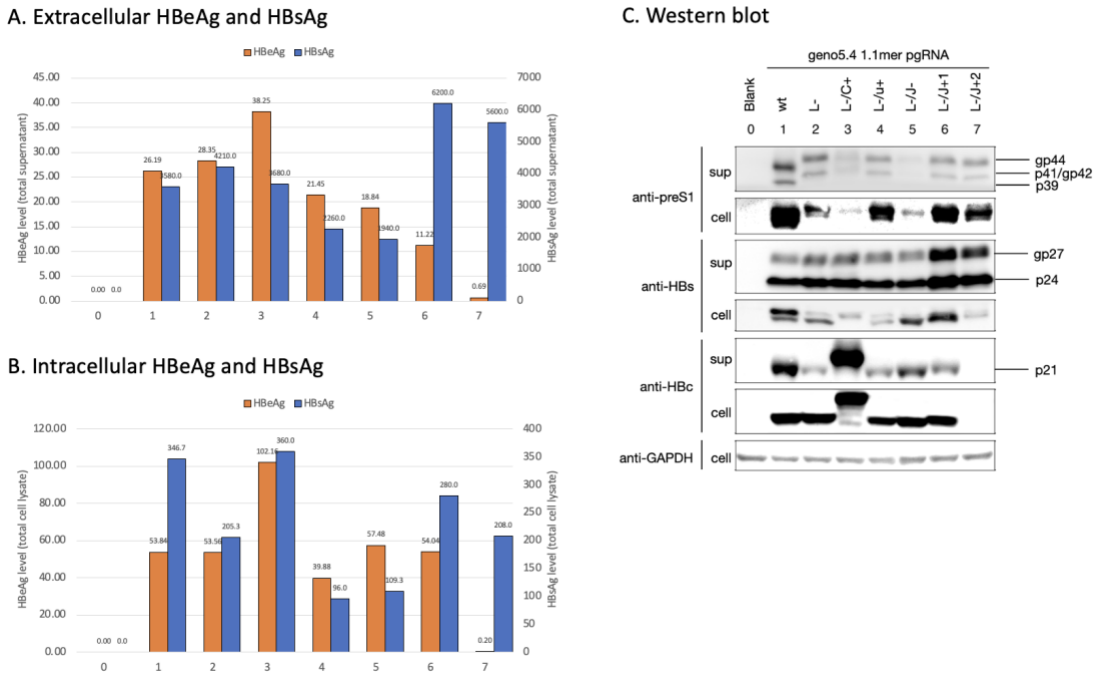

Figure 10. Impact of mutating the uORF, J ORF, or a 36-nt insertion in the core gene on p43 expression from 1.1mer pgRNA construct. *The genotype G insertion at nt1905/1906 (C+) increased core protein by 12aa as well as its expression level. uORF and J ORF expression was enhanced by optimizing the Kozak*

sequences (u+, J+1, J+2) and abolished by ATG to non-ATG mutation (J-). ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. GAPDH was used as loading control.*

3.5 Translational control of p43: impact of nonsense mutations introduced to the precore region, core gene, and P gene

If the C48* nonsense mutation in the core gene increases p43 protein level through translational termination -reinitiation, there should be a positional effect of the nonsense mutation. Naturally occurring Q2* (C1817T) and W28* (G1896A) precore mutations are upstream of the C48* mutation. Both abolished HBeAg expression as expected (Fig. 9A, lanes 2, 5, & 6). They increased extracellular p43 from the L- 1.3mer construct (capable of transcribing both pcRNA and pgRNA), with stronger effect from the W28* mutation (Fig. 9C, compare lanes 2, 5, & 6). For artificial nonsense mutations introduced to the core gene, the G10* (C-1), W71* (C-2), and especially L119* (C-3) nonsense mutations (all upstream of P gene AUG) could also increase extracellular p43 from the L- 1.3mer construct (Table 2; Fig. 8C, lanes 2-6). Further downstream nonsense mutations, Q179* and Q184*, are over 100nt downstream of the P gene AUG codon. They produced nearly full-length core protein (with Q184* missing just last two residues) detectable in Western blot (Fig. 11C, compare lanes 2, 4, & 5). While the Q179* nonsense mutation produced similar p43 phenotype as C48* mutation, Q184* exclusively generated high level of extracellular p43 with little intracellular retention (Fig. 11C, compare lanes 3-5). Surprisingly, a nonsense mutation in the P gene also markedly increased extracellular p43 (Fig.

11C, lane 6). This mutation not only prevents the expression of full-length P protein, but also converts core-P fusion protein into a core protein with its last residue replaced by 6 irrelevant residues. Such a slightly larger core protein was detectable in Western blot (Fig. 11C, compare lane 11 with lanes 2 and 6). In the context of 1.1mer spliced pcRNA construct, the Q179*, Q184*, and P- mutation also markedly increased extracellular p43 similar to the C48* mutation, although highest intracellular p43 was achieved by the C48* mutation (Fig. 11C, lanes 7-11). Interestingly, these three constructs, but not the parental spliced pcRNA construct or its C48* mutant, secreted high level of HBeAg (Fig. 11A, lanes 7-11). Thus, converting the extremely long precore/core-P fusion protein into a version slightly shorter or longer than the original precore/core protein restored proper cleavage at the C-terminus to generate mature HBeAg.

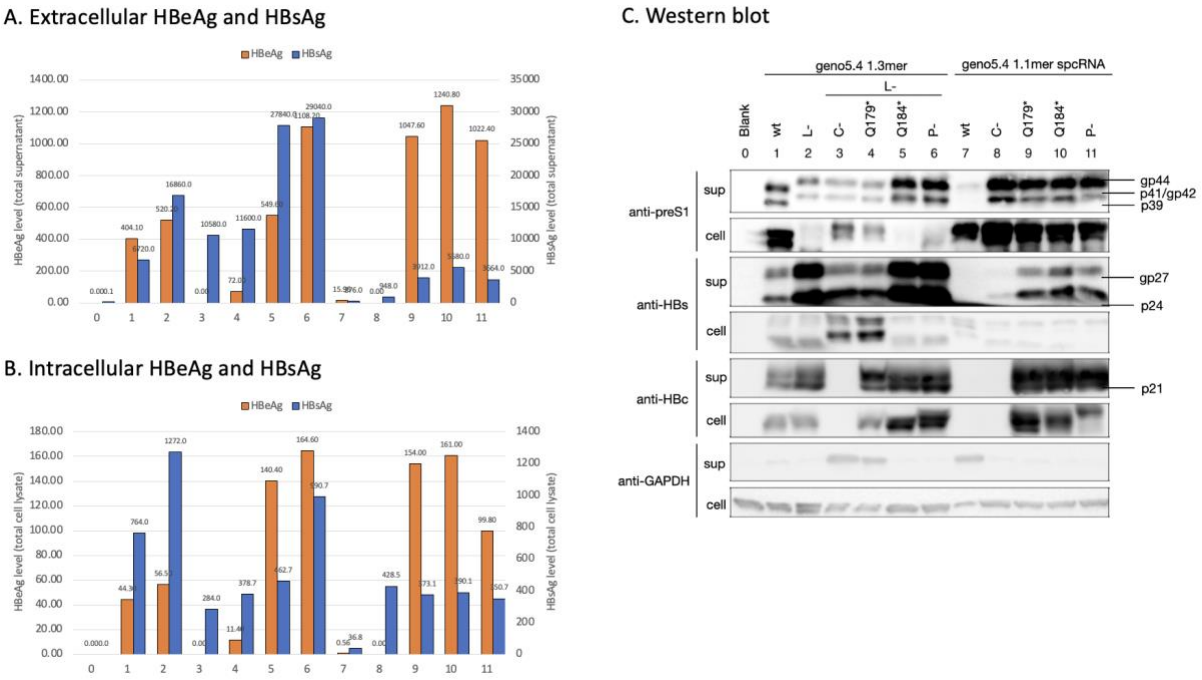

Figure 11. Core gene nonsense mutations downstream of P gene AUG codon or a P gene nonsense mutation increased p43 expression level. A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA. *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43***

(p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21). *GAPDH (cell) was used as loading control. Its presence in culture supernatant (sup) indicates bacteria contamination.*

3.6 Translational control of p43: impact of deleting a 0.4-kb upstream sequence from spliced pgRNA and optimizing the Kozak sequence

To further verify whether p43 expression from the spliced pgRNA is negatively regulated by translation of the core-P fusion protein and/or by the long distance of ribosomal scanning, we moved 5' end of HBV insert in the spliced pgRNA construct from position 1804 to position 2193. The 389-nt deletion removed initiating ATG (M1) of the core gene and two internal ATG codons (M66, M93), uORF, and J ORF. Called CMV-p43, the shortened construct produced much higher extracellular and intracellular levels of p43 even compared with spliced pcRNA construct supplemented with C48* core- mutation (Fig. 12C, compare lanes 3 & 5). Surprisingly, optimizing the P gene Kozak sequence of CMV-p43 from CAAATGC to ACCATGC or ACCATGG dramatically reduced extracellular p43 (Fig. 12C, compare lanes 5 & 6, and data not shown).

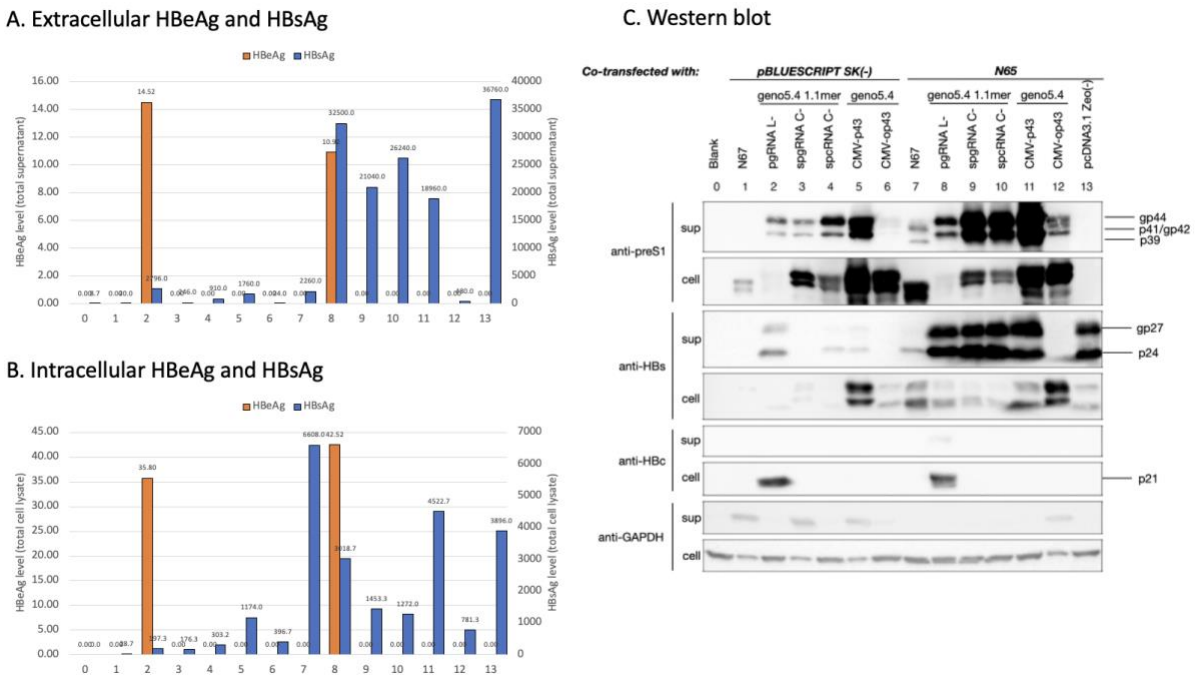

Figure 12. Providing S protein in trans could markedly increase p43 secretion. *The CMV-(o)p43 construct was derived from the 1.1mer spliced pgRNA construct by removing sequences upstream of P gene start codon with or without optimizing the Kozak sequence (o). The 1.1mer construct and CMV-(o)p43 construct were co-transfected with a S protein construct (N65) which was based on pBLUESCRIPT SK(-). Both pBLUESCRIPT SK(-) and pcDNA3.1 Zeo(-) were used as controls for co-transfection.* ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. GAPDH (cell) was used as loading control. Its presence in culture supernatant (sup) indicates bacteria contamination.*

3.7 Providing S protein *in trans* could markedly increase p43 secretion

The CMV-op43 construct with optimized Kozak sequence showed reduced extracellular and intracellular levels of S protein in addition to reduced extracellular p43 (Fig. 12C, compare lanes 5 & 6). Both spliced pgRNA construct and spliced pcRNA construct of geno5.4 also produced little intracellular and especially extracellular S protein (Fig. 5C, compare lanes 1, 4, 6, & 9), with similar reduction in HBsAg titer (Fig. 5A, lanes 1, 4, 6, & 9). Same was true for spliced pgRNA and spliced pcRNA constructs of geno1.2 (Fig. 6C). We therefore tested whether providing extra S protein *in trans* could increase p43 level for such constructs, using N65 as an expression construct for S protein (Tables 1 and 2; Fig. 12, lane 13). Co-transfecting N65 with N67, an expression construct for L protein, increased intracellular L protein level while promoting its secretion as expected [24] (Fig. 12C, lanes 1 & 7). Providing S protein *in trans* also dramatically increased extracellular p43 for the core- version of spliced pgRNA or pgRNA construct, and CMV-core construct (Fig. 12C, lanes 2-5 & 8-11), but not CMV-op43 construct (Fig. 12C, lanes 2 & 6) (but contaminated!).

3.8 p43 could be secreted even in the absence of full-length S protein

To further examine the requirement of S protein co-expression for p43 stability and secretion, the start codon of the S gene in CMV-p43 construct was mutated to GCG, ATA, AAG, or TTG (Table 2). Indeed, no S protein was detected in the culture supernatant, although a truncated S protein was detected in cell lysate (Fig. 13C, lanes 4-7). In this regard mutating the S gene ATG codon will lead to translation initiation from the downstream ATG (M75) to generate an N-terminally truncated and secretion-deficient S protein [24]. M1A (S1), M1I (S2), and MIL (S4)

mutations did not significantly affect p43 abundance (Fig. 13C, lanes 4, 5, & 7), while M1K (S3) mutation dramatically decreased extracellular p43 level (Fig. 13C, lane 6). Thus, different from L protein, the P-L fusion protein can be secreted even without co-expression of full-length S protein.

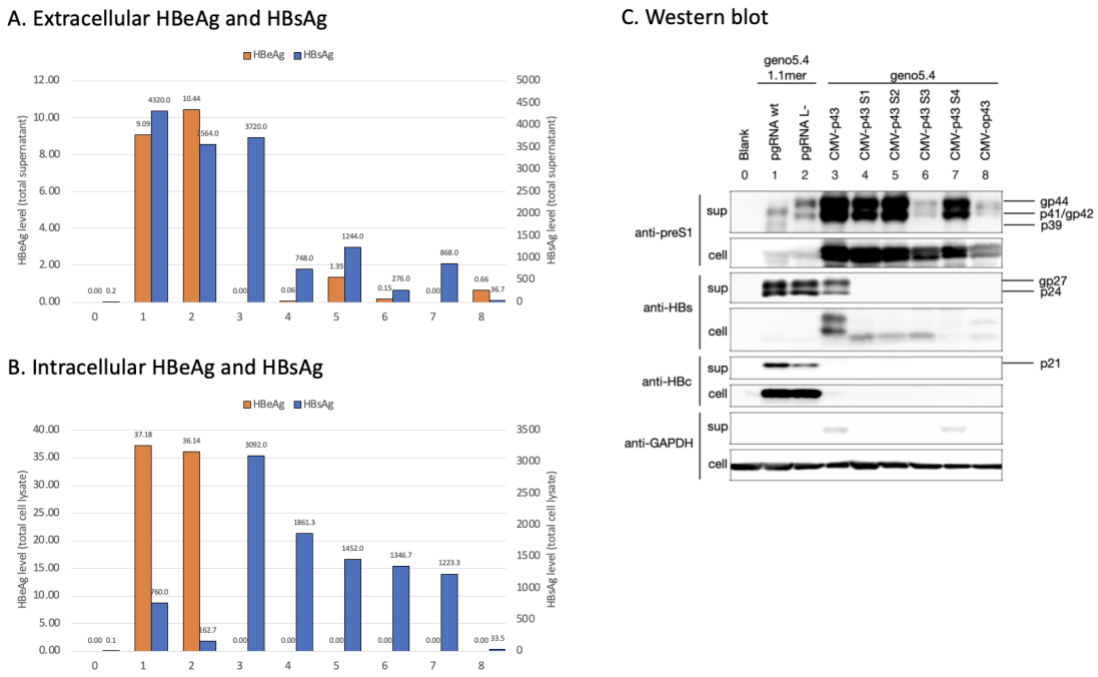

Figure 13. p43 can be secreted even without co-expression of full-length S protein. *Expression of full-length S protein was prevented in the geno5.4 CMV-p43 construct by mutating S gene ATG codon into GCG, ATA, AAG, or TTG (S1-S4).* ***A & B. Extracellular (supernatant) and intracellular (cell lysate) HBeAg and HBsAg levels by ELISA.*** *Numbers represent the adjusted OD450 value in total supernatant or total cell lysate.* ***C. Western blot of extracellular (sup) and intracellular (cell) p43 (p41/gp44), L protein (p39/gp42), S protein (p24/gp27), and core protein (p21)****. The intracellular single band in lanes 4-6 detected by anti-HBs antibody is truncated S protein due to translation initiation from the second in-frame ATG codon (M75). GAPDH (cell) was used as loading control. Its presence in culture supernatant (sup) indicates bacteria contamination.*

4 Discussion and Conclusions

4.1 p43 can be translated from both spliced pgRNA and spliced pcRNA constructs, although L- mutation alone failed to generate extracellular p43 from 1.1mer pcRNA construct of genotype A

P protein is preferentially translated from pgRNA rather than pcRNA. That is at least partly attributed to the stem-loop structure located at the 5' end of pgRNA, the ε signal that enables its packaging inside core particles. The RNA secondary structure retards scanning ribosomes from reaching the core gene AUG codon located near its 3' end. The pcRNA, with 20-30nt extension at the 5' end, rather has that stem-loop structure located downstream of the precore AUG codon to promote translation initiation [32]. That greatly reduces the chance for translation initiation from a downstream AUG. p43 expression requires single splicing (nt2454-2940 for genotype A) of the 3.5-kb pgRNA or pcRNA to fuse the 5' P ORF with nearly intact preS1/preS2/S ORF, which also fuses nearly intact core or precore ORF with 3' 3/4th of the P ORF (Fig. 2). In the present study, different types of constructs and mutants were employed to verify whether p43 is preferentially expressed from spliced pgRNA or spliced pcRNA, in case RNA splicing alters regulation of translational initiation. First, p43 could be expressed from both spliced pcRNA construct and spliced pgRNA construct, with higher p43 level reached by the spliced pgRNA construct in geno5.4 (genotype A) but not geno1.2 (genotype D) (Fig. 5C, lanes 4 & 9; Fig. 6C, lanes 6 & 10). Second, comparison was made between 1.1mer pgRNA construct and 1.1mer pcRNA construct in case the two types of 3.5-kb RNA differ at the efficiency of that particular RNA splicing. In this regard our initial detection of p43 was achieved by introducing L- mutation(s) into the 1.1mer

pgRNA construct [25], which probably promotes p43 secretion to avoid intracellular degradation. But such L- mutations failed to generate extracellular p43 for the 1.1mer pcRNA construct of geno5.4 (Fig. 5C, lanes 2 & 7), while for geno1.2 the L- mutant of 1.1mer pcRNA construct released lower level of p43 than L- pgRNA construct (Fig. 6C, lanes 4 & 9).

4.2 Naturally occurring precore or core promoter mutations can increase p43 expression

Although the 1.1mer constructs are useful to dissect the relative contribution of pgRNA vs. pcRNA in p43 production, they are artificial in that transcription of pgRNA or pcRNA is driven by the strong CMV promoter. In this regard we previously found that p43 could also be expressed from geno5.4 or geno1.2 as L- 1.3mer construct [25], which produces physiological ratio of pcRNA and pgRNA under the endogenous core promoter. Another way to verify the relative contribution of pgRNA vs. pcRNA in p43 expression is to markedly alter their ratio from 1.3mer construct. In this regard the T1753C/A1762G/T1764A/C1766T qCPM dramatically increases pgRNA transcription at the expense of pcRNA transcription [16]. Both dCPM (A1762G/T1764A) and qCPM increased extracellular p43 for the L- 1.3mer construct of geno5.4, with qCPM making p43 detectable in cell lysate (Fig. 9C, lanes 2-4). This finding, together with higher intracellular level of core protein by qCPM than dCPM, not only supports pgRNA (after its splicing) as a source for p43 expression but also suggests p43 retention by core protein. Extracellular (but not intracellular) p43 from L- 1.3mer construct was also increased by Q2* and especially W28* (G1896A) precore nonsense mutations (Fig. 9C, compare lanes 2, 5, & 6). Such mutations most likely increase p43 through translational termination-reinitiation. If so, that can only occur with spliced pcRNA rather

than pgRNA. That probably can explain why combining W28* precore mutation with the qCPM did not further increase intracellular or extracellular p43 (Fig. 9C, compare lanes 4, 6, & 7), because the qCPM would dramatically reduce pcRNA level to minimize the effect of W28* mutation. Translational termination-reinitiation could also explain higher p43 level from W28* mutant than Q2* mutant, as the latter mutation may preferentially promote translational reinitiation from core gene AUG. Overall, findings from qCPM further supports spliced pgRNA as a source for p43 expression, while the common W28* precore mutation could increase p43 expression from spliced pcRNA.

4.3 The C48* nonsense mutation in the core gene probably increases p43 protein level partly through translational termination-reinitiation

Ability of the C48*nonsense mutation to increase p43 protein level was first observed in the L- 1.1mer and 1.3mer constructs [25]. We reproduced such effect here and observed similar effect from 1.1mer pcRNA construct as well as spliced version of pgRNA, pcRNA, or 1.3mer construct (Figs. 5C, 7C, 8C). Since C48* lies upstream of P gene, an immediate possibility is translational termination-reinitiation. That is most effective when the nonsense mutation lies in close proximity to the AUG codon. But C48* is about 260nt upstream of the P gene AUG codon. We therefore compared it with other nonsense mutations in the core gene (G10*, W71*, L119*) in the context of the L- 1.3mer construct of geno5.4 and found L119* (less than 50nt upstream of P gene AUG codon) to be most effective at increasing extracellular p43 (Fig. 8C, lanes 2-6). In this regard the core ORF contains two internal AUG codons: M66 and M93, with M93 having an optimal Kozak sequence (AACATGG). The J ORF, a positive regulator of P gene translation from pgRNA,

bypasses the M93 ATG codon [17,20]. In our hand eliminating the J ORF by mutating its ATG codon also reduced p43 from L- 1.1mer pgRNA construct (Fig. 10C, lanes 2 & 5). Since L119* lies downstream of M93 ATG codon, it probably can redirect translation initiated from either M1 (initiating ATG) or M93 towards P protein expression.

4.4 Neither precore nor core protein destabilizes p43

C48* will terminate the expression of the putative core-P fusion protein from spliced pgRNA construct, and precore-P fusion protein from spliced pcRNA construct, although we failed to detect the core-P fusion protein in Western blot by the anti-core monoclonal antibody. Such fusion proteins can at least partly mediate the increase in p43 protein level if they destabilize p43. Just like much higher expression of core protein than P protein from pgRNA, such fusion protein might be expressed at much higher level than p43 from spliced pgRNA or pcRNA. C48* will further terminate the expression of core protein from 1.1mer pgRNA construct, and precore protein (HBeAg) from 1.1mer pcRNA construct. In this regard p43 is a variant form of L protein, with the N-terminal 18aa replaced by the first 47aa of P protein (in the case of genotype D). L protein initiates virion morphogenesis by interacting with core particles via its matrix domain (at the preS1/preS2 junction) [33], and that sequence is retained in p43. Indeed, initial discovery of p43 was based on immunoprecipitation of core particles from cell lysate followed by Western blot using an antibody against the N-terminus of P protein [27]. In the present study we performed co-transfection experiments between the L- 1.3mer construct and the expression construct for core or precore protein driven by the strong CMV promoter. It turned out that neither precore nor core protein reduced p43 protein level, whether or not the 1.3mer construct harbored C48* mutation to eliminate endogenous precore and core proteins (Fig. 8C, lanes 2, 4, 7, 8, 10, & 11). Rather, the

CMV-core protein construct increased intracellular p43 consistent with its retention of p43. Therefore, if the C48* nonsense mutation increases p43 protein level partly through elimination of a destabilizer, that destabilizer can only be core-P fusion protein or precore-P fusion protein.

4.5 Nonsense mutations in the core or P gene to convert precore-P fusion protein into HBeAg could increase p43 protein from spliced pcRNA construct

Deleting about 0.4-kb sequence from the 5' end of spliced pcRNA or pgRNA construct markedly increased p43 level (Fig. 12C, lanes 3-5). While that deletion makes the P gene AUG much closer to the 5' end of mRNA, it also prevented expression of the precore-P or core-P fusion protein due to loss of all the three in-frame ATG codons in core gene (M1, M66, M93). To further dissect the relative contribution of lost expression of a p43 destabilizer vs. translational termination-reinitiation, we introduced Q179* and Q184* nonsense mutations into the core gene of geno5.4. Over 100nt downstream of the P gene AUG codon, they will shorten the C-terminus of core or precore protein by 7aa and 2aa, respectively, and convert the core-P or precore-P fusion protein into such shortened core or precore protein. Expression of the full-length P protein and coding capacity for p43 will be unaltered. Furthermore, we introduced a Q219* nonsense mutation in the P gene. It will truncate the expression of full-length P protein from pgRNA and convert the core-P or precore-P fusion protein into core or precore protein extended by 5aa. But expression of core or precore protein from unspliced pgRNA or pcRNA will be unaltered, nor will be coding capacity for p43. Remarkably, both the Q184* (core gene) and Q219* (P gene) nonsense mutations increased p43 level, whether from L- 1.3mer construct or spliced pcRNA construct (Fig. 11C). In the spliced pcRNA construct, the Q179*, Q184*, and Q219* nonsense mutations secreted high

titer of HBeAg in contrast to the parental clone (Fig. 11A, lanes 7, 9-11). Therefore, converting precore-P fusion protein into precore protein (HBeAg) can increase p43 level in the absence of translational termination-reinitiation, although higher intracellular p43 from the C48* mutant (Fig. 11C, lanes 8-11) supported added effect of translational termination-reinitiation from this mutation. The Q184* nonsense mutation (Q182* in non-A HBV genotypes), which does not impair core protein function, was detected from Korean HBeAg-positive liver cancer patients. [34]

4.6 Impact of viral genotype and S protein on p43 secretion and stabilization, and possible role of core-P (or precore-P) fusion protein in destabilizing S protein

Our studies on p43 have been focused on geno5.4 of genotype A and geno1.2 of genotype D. In this regard S protein from genotype D is more efficient at secretion than its counterpart from genotype A, whereas L protein from genotype D has a stronger inhibitory effect [28]. Thus in geno1.2 lost L protein expression dramatically increased extracellular S protein for both pgRNA construct and pcRNA construct (Fig. 6C, compare lanes 3 & 4, lanes 8 & 9), whereas in geno5.4 L- mutation did not increase extracellular S protein for the pgRNA construct but abolished already low S protein for the pcRNA construct (Fig. 5C, lanes 1 & 2; lanes 6 & 7). Since L- mutation generated extracellular p43 for geno1.2 but not geno5.4, S protein might stabilize p43 or promote its secretion. Indeed according to the previous report, extracellular p43 was associated with SVPs [27]. While for geno1.2 spliced pgRNA construct and spliced pcRNA construct produced much higher p43 level than their respective unspliced version (Fig. 6C, compare lanes 4 & 6, lanes 9 & 10), for geno5.4 bypassing RNA splicing did not markedly increase extracellular p43 level for the pgRNA

construct but greatly increased intracellular level (Fig. 5C, lanes 2-5). Spliced pgRNA or pcRNA construct is similar to L- pgRNA or pcRNA construct in unable to express L protein, but spliced pcRNA or pgRNA construct produced much less extracellular S protein for geno5.4 and both extracellular and intracellular S protein for geno1.2. That particular RNA splicing does not remove the SPII promoter required for transcription of the 2.1-kb RNA for M and S proteins. So probably the core-P and precore-P fusion proteins can destabilize S protein. Indeed, the C48* core- mutation moderately increased S protein level for the spliced pcRNA construct and spliced pgRNA construct (mostly the intracellular S protein for geno1.2). In the co-transfection experiment based on geno5.4, providing extra S protein *in trans* could increase p43 secretion from both spliced pcRNA and pgRNA constructs with a C48* mutation, and from CMV-p43 construct incapable of expressing core-P fusion protein (Fig. 12C). Whether the same effect can be observed on spliced pcRNA or pgRNA construct without the core- mutation remains to be determined. Thus, S protein probably stabilized p43 and promoted its secretion. In this regard, we previously found that S protein could also stabilize L and M proteins [24]. Since mutating the S gene ATG codon to prevent expression of full-length S protein did not significantly reduce extracellular p43 (Fig. 13C), S protein is not essential for p43 secretion, at least for genotype A. Considering the higher secretion efficiency of S protein from genotype D, similar experiments should be performed in geno1.2 in the future. Ability for p43 to secrete by itself is not surprising since it no longer has the myristoylation signal and a retention sequence, which inhibit L protein secretion [35,36]. Alternatively, continued expression of M protein might facilitate p43 secretion.

4.7 Core protein and p43 could be detected by ELISA kits as HBeAg and HBsAg, respectively

HBeAg and HBsAg secreted to culture supernatant of Huh7 cells transfected with cloned HBV DNA can be detected by ELISA kits. ELISA is highly sensitive, and we could detect weaker HBeAg signals from cell lysate as well (compare Fig. 9B with Fig. 9A). The 1.1mer pgRNA construct is incapable of HBeAg expression, thus releasing little HBeAg signal when compared with the 1.3mer construct (Fig. 9A, lanes 1 & 8). But its intracellular HBeAg titer was twice higher than its extracellular titer and even higher than the intracellular signal from 1.3mer construct (Fig. 9A & 9B, lanes 1 & 8). The Q2* and W28* nonsense mutations in the precore region should abolish HBeAg from the 1.3mer construct, yet high intracellular "HBeAg" titers could still be detected (Fig. 9A & 9B, lanes 5-7). Western blotting established that such intracellular "HBeAg" was actually core protein, with its abundance correlated with intracellular HBeAg titer (Fig. 9C). When the start codon for S protein was mutated in the CMV-p43 construct to prevent the expression of full-length S protein, titer of extracellular HBsAg was reduced but not abolished while intracellular HBsAg was reduced much less significantly (Fig. 13A &13B, lanes 3-7). The intracellular and extracellular HBsAg titers corresponded nicely with the abundance of p43 level according to Western blot (Fig. 13C). Therefore, core protein could contribute to intracellular "HBeAg" signal from ELISA while both intracellular and extracellular p43 can be detected by ELISA as HBsAg.

4.8 Bacterial contamination inhibited secretion of HBV proteins

We observed under microscope bacterial contamination in certain wells of transfected Huh7 cells, which correlated with a high level of GAPDH in the culture supernatant (Fig. 8C, lane 1; Fig. 11C, lanes 3, 4, & 7; Fig. 12C, lanes 1, 3, 4, & 12; Fig. 13C, lanes 3 & 7). The cells were detached from the wells on day 4 post-transfection, but no significant cell damage was observed before then. Under the microscope, floating rod-shaped microorganisms were visualized,

suggesting *E. coli* contamination. The contaminated cells and culture supernatant were collected together and separated through centrifuging at 1,000 rpm for 1 minute. The contaminated cell pellets were smaller than those without contamination. And the contaminated culture supernatant presented a pinkish color (~pH7.8) compared to the yellowish color (~pH7.0) of those without contamination, suggesting the release of intracellular contents (pH7.0-8.0) into the culture supernatant [37,38]. Larger PEG pellets were obtained with contaminated samples, owing to increased proteins released by lysed cells and bacteria in the culture supernatant. Curiously, the HBV protein levels of contaminated samples were decreased in culture supernatant but increased in cell lysate. For example, the 1.3mer wt sample of geno5.4 was contaminated in one set of transfection experiment (Fig. 8C, lane 1). Its extracellular HBeAg and HBsAg titers were significantly lower than those from L- mutant (uncontaminated), but intracellular L and S proteins were much higher (Fig. 8A & C, lanes 1 & 2). For data presented in Fig. 9 where neither sample was contaminated, the extracellular HBeAg and HBsAg levels were similar, and the difference of intracellular L and S protein levels was not as striking (Fig. 9A & C, lanes 1 & 2). In this regard, bacterial contamination in certain wells of transfected cells may complicate data interpretation for rescue of p43 secretion by co-transfection with an S protein construct (Fig. 12C).

4.9 Conclusions

p43, the P-L fusion protein, can be translated from both spliced pgRNA and spliced pcRNA, although whether the pgRNA or pcRNA is more efficient at that particular splicing to generate p43 coding capacity remains to be determined. Naturally occurring G1896A (W28*) precore mutation most likely increases p43 protein level from spliced pcRNA, partly through translational reinitiation and partly through lost precore-P fusion protein. Artificial C48* core gene mutation could increase p43 protein level from either spliced pcRNA or spliced pgRNA, possibly

partly through translational reinitiation and partly through lost core-P or precore-P fusion protein. Nevertheless, nonsense mutations to prevent the expression of precore-P fusion protein from spliced pcRNA construct could still increase p43 protein level even when translational reinitiation seemed unlikely, thus strongly implicating precore-P fusion protein as a p43 destabilizer. On the other hand, co-transfection experiments and phenotypes of the quadruple core promoter mutant ruled out intact core protein or precore protein (HBeAg) as a p43 destabilizer, but suggested p43 retention by core protein. The precore-P or core-P fusion protein may also destabilize S protein to reduce p43 protein level from the pcRNA construct, spliced pcRNA construct, and spliced pgRNA construct of the genotype A clone tested. S protein could stabilize p43 and promote its secretion when provided *in trans*, although abolishing co-expression of full-length S protein from a p43 expression construct did not impair its secretion. Thus N-terminal sequence difference between p43 and L protein may render the former more prone to secretion. Finally, bacterial contamination of transfected cells may complicate data interpretation.

Works Cited

1. What is Viral Hepatitis? | CDC. Published April 8, 2021. Accessed March 3, 2023. https://www.cdc.gov/hepatitis/abc/index.htm

2. Hepatitis (Viral) - NIDDK. National Institute of Diabetes and Digestive and Kidney Diseases. Accessed April 9, 2023. https://www.niddk.nih.gov/health-information/liver-disease/viral-hepatitis

3. Weng MK. Universal Hepatitis B Vaccination in Adults Aged 19–59 Years: Updated Recommendations of the Advisory Committee on Immunization Practices — United States, 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71. doi:10.15585/mmwr.mm7113a1

4. Dane DS, Cameron CH, Briggs M. VIRUS-LIKE PARTICLES IN SERUM OF PATIENTS WITH AUSTRALIA-ANTIGEN-ASSOCIATED HEPATITIS. *The Lancet*. 1970;295(7649):695-698. doi:10.1016/S0140-6736(70)90926-8

5. Seeger C, Mason WS. Molecular biology of hepatitis B virus infection. *Virology*. 2015;479-480:672-686. doi:10.1016/j.virol.2015.02.031

6. Tong S, Revill P. Overview of hepatitis B viral replication and genetic variability. *J Hepatol*. 2016;64(1 Suppl):S4-S16. doi:10.1016/j.jhep.2016.01.027

7. Li W, Urban S. Entry of hepatitis B and hepatitis D virus into hepatocytes: Basic insights and clinical implications. *J Hepatol*. 2016;64(1 Suppl):S32-S40. doi:10.1016/j.jhep.2016.02.011

8. Sheu SY, Lo SJ. Preferential ribosomal scanning is involved in the differential synthesis of the hepatitis B viral surface antigens from subgenomic transcripts. *Virology*. 1992;188(1):353-357. doi:10.1016/0042-6822(92)90764-g

9. Tong S, Li J, Wands JR, Wen Y mei. Hepatitis B virus genetic variants: biological properties and clinical implications. *Emerg Microbes Infect*. 2013;2(1):1-11. doi:10.1038/emi.2013.10

10. Wai CT, Fontana RJ. Clinical significance of hepatitis B virus genotypes, variants, and mutants. *Clin Liver Dis*. 2004;8(2):321-352, vi. doi:10.1016/j.cld.2004.02.006

11. Milich DR, Chen MK, Hughes JL, Jones JE. The secreted hepatitis B precore antigen can modulate the immune response to the nucleocapsid: a mechanism for persistence. *J Immunol Baltim Md 1950*. 1998;160(4):2013-2021.

12. Carman WF, Jacyna MR, Hadziyannis S, et al. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet Lond Engl*. 1989;2(8663):588-591. doi:10.1016/s0140-6736(89)90713-7

13. Tong SP, Li JS, Vitvitski L, Trépo C. Active hepatitis B virus replication in the presence of anti-HBe is associated with viral variants containing an inactive pre-C region. *Virology*. 1990;176(2):596-603. doi:10.1016/0042-6822(90)90030-u

14. Li JS, Tong SP, Wen YM, Vitvitski L, Zhang Q, Trépo C. Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *J Virol*. 1993;67(9):5402-5410. doi:10.1128/JVI.67.9.5402-5410.1993

15. Okamoto H, Tsuda F, Akahane Y, 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;68(12):8102-8110. doi:10.1128/JVI.68.12.8102-8110.1994

16. Parekh S, Zoulim F, Ahn SH, et al. Genome Replication, Virion Secretion, and e Antigen Expression of Naturally Occurring Hepatitis B Virus Core Promoter Mutants. *J Virol*. 2003;77(12):6601-6612. doi:10.1128/JVI.77.12.6601-6612.2003

17. Hwang WL, Su TS. Translational regulation of hepatitis B virus polymerase gene by termination-reinitiation of an upstream minicistron in a length-dependent manner. *J Gen Virol*. 1998;79 (Pt 9):2181-2189. doi:10.1099/0022-1317-79-9-2181

18. Chen A, Kao YF, Brown CM. Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons. *Nucleic Acids Res*. 2005;33(4):1169-1181. doi:10.1093/nar/gki251

19. Zong L, Qin Y, Jia H, et al. Differential regulation of hepatitis B virus core protein expression and genome replication by a small upstream open reading frame and naturally occurring mutations in the precore region. *Virology*. 2017;505:155-161. doi:10.1016/j.virol.2017.02.020

20. Fouillot N, Tlouzeau S, Rossignol JM, Jean-Jean O. Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation. *J Virol*. 1993;67(8):4886-4895. doi:10.1128/JVI.67.8.4886-4895.1993

21. Li K, Zoulim F, Pichoud C, et al. Critical role of the 36-nucleotide insertion in hepatitis B virus genotype G in core protein expression, genome replication, and virion secretion. *J Virol*. 2007;81(17):9202-9215. doi:10.1128/JVI.00390-07

22. Gutelius D, Li J, Wands J, Tong S. Characterization of the pleiotropic effects of the genotype G-specific 36-nucleotide insertion in the context of other hepatitis B virus genotypes. *J Virol*. 2011;85(24):13278-13289. doi:10.1128/JVI.05583-11

23. Zhang J, Zong L, Wang Y, et al. Core gene insertion in hepatitis B virus genotype G functions at both the encoded amino acid sequence and RNA structure levels to stimulate core protein expression. *Virology*. 2019;526:203-213. doi:10.1016/j.virol.2018.11.002

24. Zhang J, Wang Y, Fu S, et al. Role of Small Envelope Protein in Sustaining the Intracellular and Extracellular Levels of Hepatitis B Virus Large and Middle Envelope Proteins. *Viruses*. 2021;13(4):613. doi:10.3390/v13040613

25. Zhang J, Yuan Q, Wang Y, et al. 5′ preS1 Mutations To Prevent Large Envelope Protein Expression from Hepatitis B Virus Genotype A or Genotype D Markedly Increase Polymerase-Envelope Fusion Protein. Ou JHJ, ed. *J Virol*. 2022;96(5):e01723-21. doi:10.1128/jvi.01723-21

26. Kremsdorf D, Lekbaby B, Bablon P, et al. Alternative splicing of viral transcripts: the dark side of HBV. *Gut*. 2021;70(12):2373-2382. doi:10.1136/gutjnl-2021-324554

27. Huang HL, Jeng KS, Hu CP, Tsai CH, Lo SJ, Chang C. Identification and characterization of a structural protein of hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. *Virology*. 2000;275(2):398-410. doi:10.1006/viro.2000.0478

28. Zhang F, Tang X, Garcia T, et al. Characterization of contrasting features between hepatitis B virus genotype A and genotype D in small envelope protein expression and surface antigen secretion. *Virology*. 2017;503:52-61. doi:10.1016/j.virol.2017.01.009

29. Yuan Q, Ge S, Xiong J, et al. A novel immunoassay for PreS1 and/or core-related antigens for detection of HBsAg variants. *J Virol Methods*. 2010;168(1):108-113. doi:10.1016/j.jviromet.2010.04.029

30. Zhang TY, Yuan Q, Zhao JH, et al. Prolonged suppression of HBV in mice by a novel antibody that targets a unique epitope on hepatitis B surface antigen. *Gut*. 2016;65(4):658-671. doi:10.1136/gutjnl-2014-308964

31. Wang SJ, Chen ZM, Wei M, et al. Specific determination of hepatitis B e antigen by antibodies targeting precore unique epitope facilitates clinical diagnosis and drug evaluation against hepatitis B virus infection. *Emerg Microbes Infect*. 2021;10(1):37-50. doi:10.1080/22221751.2020.1862631

32. Hwang WL, Su TS. The encapsidation signal of hepatitis B virus facilitates preC AUG recognition resulting in inefficient translation of the downstream genes. *J Gen Virol*. 1999;80 (Pt 7):1769-1776. doi:10.1099/0022-1317-80-7-1769

33. Bruss V. A short linear sequence in the pre-S domain of the large hepatitis B virus envelope protein required for virion formation. *J Virol*. 1997;71(12):9350-9357. doi:10.1128/JVI.71.12.9350-9357.1997

34. Zhou H, Gewaily D, Ahn SH, et al. Sequence analysis and functional characterization of full-length hepatitis B virus genomes from Korean cirrhotic patients with or without liver cancer. *Virus Res*. 2017;235:86-95. doi:10.1016/j.virusres.2017.03.021

35. Prange R, Clemen A, Streeck RE. Myristylation is involved in intracellular retention of hepatitis B virus envelope proteins. *J Virol*. 1991;65(7):3919-3923. doi:10.1128/JVI.65.7.3919-3923.1991

36. Kuroki K, Russnak R, Ganem D. Novel N-terminal amino acid sequence required for retention of a hepatitis B virus glycoprotein in the endoplasmic reticulum. *Mol Cell Biol.* 1989;9(10):4459-4466. doi: 10.1128/mcb.9.10.4459-4466.1989

37. Carbon dioxide Concentration and pH Control in the Cell Culture Laboratory. Accessed April 5, 2023. https://www.culturecollections.org.uk/news/ecacc-news/co2-concentration-and-ph-control-in-the-cell-culture-laboratory.aspx

38. Pollock AS. Intracellular pH of hepatocytes in primary monolayer culture. *Am J Physiol.* 1984;246(5 Pt 2):F738-744. doi:10.1152/ajprenal.1984.246.5.F738