

Article

Comprehensive Analysis of Hepatitis B Virus Promoter Region Mutations

Vanessa Meier-Stephenson ^{1,2}, William T.R. Bremner ³, Chimone S. Dalton³, Guido van Marle¹, Carla Coffin^{1,4} and Trushar R. Patel ^{1,2,5,*}

¹ Department of Microbiology, Immunology and Infectious Diseases, Cumming, School of Medicine, University of Calgary, Calgary, Alberta, T2N 1N4, Canada;

² Alberta RNA Research & Training Institute, Department of Chemistry & Biochemistry, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada;

³ Department of Ecosystem & Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, T2N 4N1, Canada;

⁴ Liver Unit, Division of Gastroenterology and Hepatology, Department of Medicine, Calgary, Alberta, T2N 4Z6, Canada;

⁵ DiscoveryLab, Faculty of Medicine & Dentistry, and the Li Ka Shing Institute of Virology, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada;

* Correspondence: trushar.patel@uleth.ca; Tel.: +1-403- 332-4443

Abstract: Over 250 million people are infected chronically with hepatitis B virus (HBV), the leading cause of liver cancer worldwide. HBV persists due in part to its compact, stable minichromosome, the covalently-closed, circular DNA (cccDNA), which resides in the hepatocytes' nuclei. Current therapies target downstream replication products, however, a true virological cure will require targeting the cccDNA. Finding targets on such a small, compact genome is challenging. For HBV, to remain replication-competent, it needs to maintain nucleotide fidelity in key regions, such as the promoter regions, to ensure that it can continue to utilize the necessary host proteins. HBVdb (HBV database) is a repository of HBV sequences spanning all genotypes (A-H) amplified from clinical samples, and hence implying an extensive collection of replication-competent viruses. Here, we analyzed the HBV sequences from HBVdb using bioinformatics tools to comprehensively assess the HBV core and X promoter regions amongst the nearly 70,000 HBV sequences for highly-conserved nucleotides and variant frequencies. Notably, there is a high degree of nucleotide conservation within specific segments of these promoter regions highlighting their importance in potential host protein-viral interactions and thus the virus' viability. Such findings may have key implications for designing antivirals to target these areas.

Keywords: hepatitis B virus (HBV); cccDNA; basal core promoter; X promoter; single nucleotide polymorphisms; logo analyses; genotype alignments

1. Introduction

Hepatitis B virus (HBV) is a partially double-stranded DNA virus belonging to the *Hepadnaviridae* family that chronically infects over 250 million people worldwide[1]. Infection is associated with 25-40% lifetime risk of severe liver disease, including cirrhosis, liver failure, and hepatocellular carcinoma (HCC)[1]. There is no cure for HBV and while current nucleos(t)ide analog treatments act to suppress viral replication, virological relapses often occur once therapy is stopped[2-5]. This is due to the persistence of the virus' minichromosome or covalently closed circular DNA (cccDNA). A therapy against chronic HBV will therefore need to target cccDNA[6-9]. The search for new targets in this compact, challenging virus has led some researchers to consider targeting the involved host proteins[10, 11]. While promising, this approach may harbor some

unwanted off-target effects on the host. An alternative strategy is to analyze the reciprocal interaction for new therapeutic targets, i.e., the viral region(s) that are recognized by the host proteins.

Gene promoter regions are essential sites in DNA recognized by proteins for the downstream processes of replication and transcription[12, 13]. Mutations in promoter regions can often lead to down- or up-regulation of genes that may lead to malignancies[14-18]. For example, in gastric adenocarcinomas, specific mutations in the *AKT1* promoter region have been identified from clinical samples and confirmed *in vitro* to upregulate the Akt/PKB pathway, a central pathway for cancer cell proliferation[17, 18]. As well, *TERT* gene promoter mutations lead to upregulation of hTERT (human telomerase reverse transcriptase), a key step in cell immortalization, and has been found in over 40% of cutaneous melanoma, effectively acting as an independent prognostic factor for disease[19, 20]. Promoter region mutations that occur in the context of properly functioning DNA proofreading mechanisms are corrected or naturally-selected out if not beneficial to the organism, thus maintaining the genome and resulting in conserved sequences.

Similar to the genome organization in higher organisms, pathogens such as viruses and bacteria also contain promoter sites in their genome. However, unlike most living organisms, viruses require the host-machinery to survive. Upon infecting a host cell, viruses manipulate and utilize host proteins to enable its replication and survival. Host proteins are similarly recruited to viral promoter regions for the analogous replicative processes. For example, HIV-1 recruits Sp1 to its COMMD1 promoter region to regulate transcription[21]; and in human papillomavirus (HPV), Sp1, AP1 and NF1 are recruited to early promoters to initiate transcription[22-24]. Thus, for effective infection, those regions that interact with the host proteins would also need to be highly conserved.

Hepatitis B virus' stable genome, the cccDNA, comprises of four genes, C, S, P and X (Figure 1A). The genes are arranged into four overlapping reading frames, the transcripts of which are controlled by four different promoter regions—C, X, PreS1 and PreS2, again overlapping with a segment of the preceding gene[25-27]. As such, HBV has one of the smallest genomes, comprising only 3.2 kb. During chronic infection, up to 10¹² virions can be produced per day, which in combination with the highly error-prone HBV polymerase, can give rise to up to 10¹⁰⁻¹¹ point mutations daily generating innumerable viral quasispecies with varying degrees of fitness[28]. Viruses with reduced fitness would be selected out over time, leaving the most stable, replication-competent viruses to be recovered from clinical infection.

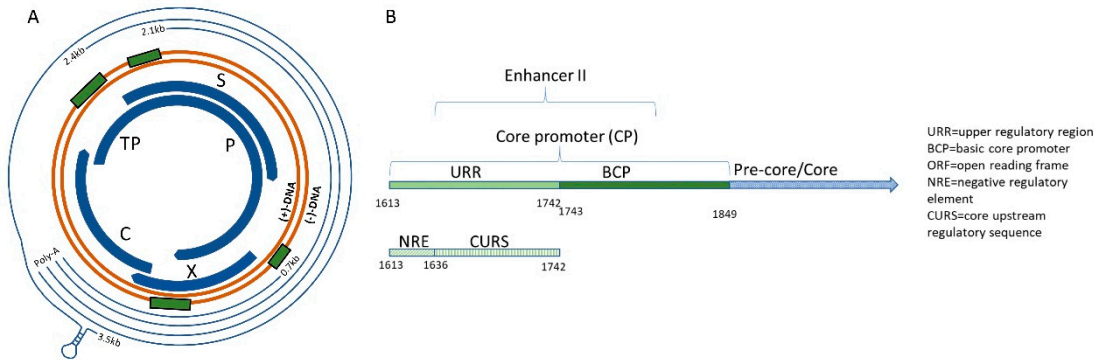


Figure 1. Schematic of the HBV genome. (A) From the (-)-strand DNA, overlapping reading frames code for gene segments C, P, S and X. Transcripts are depicted by the outer thin lines and ending in the poly-A adenylation sequence. The X and Core promoter regions are highlighted by green rectangles preceding the corresponding gene. (B) Core promoter schematic.

In 2012, researchers from France started an international repository of sequenced clinical HBV specimens. The database (HBVdb) is open-access (<https://hbvdb.ibcp.fr/HBVdb/>) and currently has nearly 70 000 reference genomes[29]. Given that these genomes are based on clinical specimens, the repository thus represents an extensive database of *replication-competent* viruses.

In this work, the HBV core and X promoter regions were analyzed across this extensive HBV repository for mutational frequencies in all known genotypes, with the hypothesis that regions exhibiting the highest nucleotide conservation are more likely to be critical in the host protein-viral DNA interaction. Indeed, we find that there is considerable nucleotide conservation throughout the promoter regions, amongst and between genotypes. These conserved sites highlight specific sub-regions as likely essential for maintaining a host protein interaction, and thus may make reasonable targets for antiviral therapy.

2. Materials and Methods

HBV sequences were sourced from the publicly accessible HBV database (<https://hbvdb.ibcp.fr/>)[29] and included 840 sequences of genotype A; 1700 genotype B; 2153 genotype C; 948 genotype D; 255 genotype E; 248 genotype F; 39 genotype G; 26 genotype H. Alignments were imported into Geneious 10.1.3 (www.geneious.com) in FASTA format for downstream analysis. The consensus sequence for the alignment of all available whole genome HBV sequences was extracted and annotated for promoter binding regions previously identified in the core promoter region[30-32]. These annotations included C/EBP-like, HNF4, HNF3, SP1, and TBP binding sites. The same annotations were further applied to individual genotype alignments (HBV genotypes A-H) and extracted as separate alignments. Owing to differences in the alignments derived from the HBV database, each genotype required separate annotation. The genome positions of the relevant binding sites for each genotype are summarized in a table format (Supplementary Table 1). Genotype alignments of promoter binding sites were then converted to sequence logos using WebLogo (<https://weblogo.berkeley.edu/logo.cgi>)[33] and manually stacked for comparison. Variant frequency tables were additionally generated for each alignment converted to a sequence logo using a 0.05% frequency threshold using the ‘find SNP/variants’ function of Geneious and are included in the supplementary material (Supplementary Table 1). Poly-mutational analyses of known high-risk mutations per genotype, including the Fisher exact test were performed using Microsoft Excel.

Table 1. Variant frequencies within the X and core promoter regions. Frequencies between 5-95% are reported. For the full list, please see Supplemental section.

X promoter (nts 1101-1121)		
Genotype	Position (nt)	Variant Frequency
B	1113	91.6% T; 7.7% C
	1116	86.0% A; 11.4% G
C	1103	53.0%; G 46.9% A
	1104	66.0% G; 33.7% C
	1108	94.3% A
E	1108	88.0% A; 10.4% C
	1113	67.6% T; 32.0% C
NREBP (nts 1611-1619)		
A	1612	90.2% A; 9.8% C
B	1613	80.3% G; 19.7% A
C	1613	77.3% C; 22.7% A
D	1613	92.7% G; 7.3% A
E	1613	94.3% G; 5.7% A
G	1617	94.9% G; 5.1% A
C/EBP-HNF4(1) (nts 1644-1672)		
C	1653	87.8% C; 12.2% T
D	1653	90.9% C; 9.1% T
E	1653	93.5% C; 6.5% T
F	1648	94.8% T; 5.2% C
	1658	86.3% G; 12.1% A
	1659	56.3% A; 43.5% C
	1665	83.7% C; 16.1% A
G	1645	66.7% G; 33.3% A
	1647	94.9% C; 5.1% G
	1648	94.9% A; 5.1% T
	1649	94.9% G; 5.1% A
	1652	94.9% A; 5.1% G
	1653	94.9% T; 5.1% C
H	1659	94.9% A
H	1653	92.3% C; 7.7% T
	1661	92.3% G; 7.7% A

HNF3(1)-FTF (nts 1679-1700)		
Genotype	Position (nt)	Variant Frequency
C	1679	90.8% A; 7.7% G
D	1679	92.9% A; 5.1% G
F	1680 1688	45.0% A; 44.4% C; 6.7% T 66.1% G;33.7% C
G	1679 1688 1693 1694 1695 1697	94.9% T; 5.1% A 94.9% A; 5.1% G 94.9% G; 5.1% A 94.9% G; 5.1% C 84.6% G; 10.3% A; 5.1% C 94.9% G; 5.1% T
HNF3(2)-Sp1(1) (nts 1713-1740)		
A	1727 1740	66.7% G; 33.3% A 66.2% C; 33.8% T
B	1726 1727	61.3% C; 38.5% A 94.4% T
C	1719 1721 1727	54.3% T; 45.6% G 77.2% G; 22.8% A 57.6% A; 42.2% G
D	1724 1727	94.6% T;5.4% C 78.2% A; 21.7% G
E	1727	92.2% A; 7.8% G
F	1721 1727 1740	94.8% A; 5.2% G 78.6% G;21.0% A 66.1% C; 33.9% T
G	1721 1725 1726 1727 1730 1736 1739	92.3% T; 5.1% G 94.9% G; 5.1% A 94.9% C; 5.1% A 94.9% T; 5.1% A 94.9% G; 5.1% C 94.9% A; 5.1% G 94.9% A; 5.1% G
H	1721	88.5% A ; 11.5% G
Sp1(2) (nts 1743-1752)		
B	1752	53.6% A; 42.1% G

D	1752	83.8% A; 13.1% C
G	1745	94.9% C; 5.1% G
	1746	94.9% A; 5.1% G
	1748	94.9% T; 5.1% G
	1752	94.9% T; 5.1% A
H	1752	88.5% T; 11.5% C
TBP(1)-HNF4(2)-TBP(2) (nts 1758-1776)		
Genotype	Position (nt)	Variant Frequency
A	1762	71.3% A; 28.5% T
	1764	69.1% G; 30.6% A
	1766	93.3% C; 6.1% T
B	1762	83.7% A; 16.2% T
	1764	84.0% G; 16.0% A
	1773	88.4% C; 11.6% T
C	1758	93.2% T; 6.0% C
	1762	51.9% A; 47.9% T
	1764	50.9% A; 48.9% G
	1766	94.8% C; 5.0% T
	1775	79.4% A; 20.1% G
D	1761	93.4% A; 6.1% C
	1762	76.1% A; 23.3% T
	1764	66.8% G; 25.1% A; 7.4% T
	1766	85.9% C; 9.7% G
	1773	60.2% C; 39.7% T
E	1762	87.1% A; 11.8% T
	1764	87.5% G; 12.2% A
F	1762	76.0% A; 24.0% T
	1764	75.6% G; 24.4% A
	1773	56.9% T; 42.5% C
G	1765	92.3% C; 7.7% T
	1773	94.9% T; 5.1% C
H	1762	92.3% A; 7.7% T
TBP(3) (nts 1788-1795)		
A	1790	86.8% T; 13.0% G

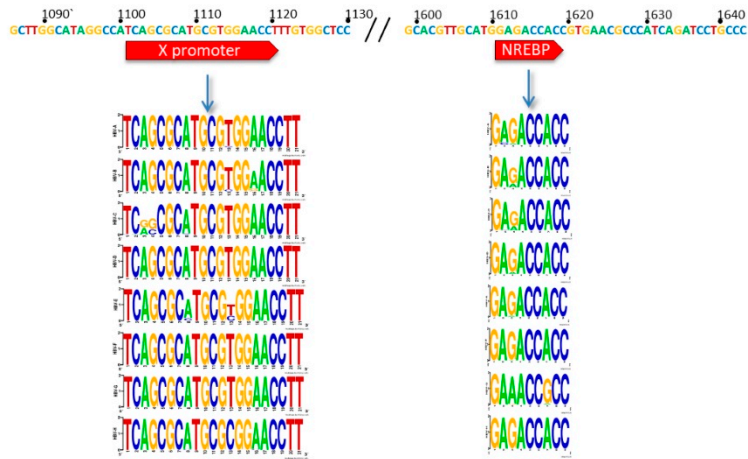
3. Results

3.1. X gene promoter region.

The X gene promoter region extends from nts 1101-1121 (TCAGCGCATGCGTGGAACCTT), that aides the production of the transcript for the small protein, HBx (17 kDa)[34-36]. This protein does not share homology with any other known gene and has been found to play roles in upregulation of the core promoter and thus HBV replication as well as being central to the pathogenesis of HBV-induced hepatocellular carcinoma (HCC)[36-42]. The host protein, nuclear respiratory factor 1 (NRF-1) has been shown to be responsible for X promoter binding and transcription initiation[36]. In our analyses, we note two nucleotide variants at positions 1103 and 1104 of the genotype C sequence logo are unique compared to the other genotypes (Figure 2A). As

well, nt 1113 had some variance between C and T. All other nucleotides had high levels of conservation.

A.



2B.

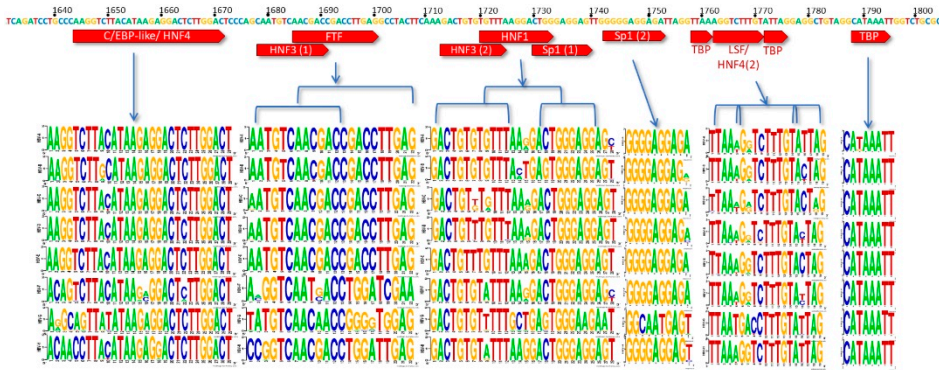


Figure 2. HBV X and Core promoter regions. Upper bar represents the consensus sequence across all genotypes. Highlighted segments refer to previously identified binding sites for various host proteins. Stacked LOGO sequences are displayed per genotype to give a pictorial representation of conserved nucleotides and variants. The complete consensus for the core promoter (nts 1613-1849) is included in the Supplemental Figure 1.

3.2. C gene/HBV Core promoter region.

The Core promoter (CP) spans nts 1611-1847 and can be subdivided into the upper regulatory region (URR) and basic core promoter (BCP) (Figure 1B)[34]. Within the URR is a segment known as the negative regulatory element (NRE) which, when bound by the NRE-binding protein, can suppress the core promoter activity by approximately 10-20-fold [43]. Isolates in the HBV database show a high degree of conservation throughout this region, with the exception of nt 1613 where there were G to A variations amongst all genotypes. Also, genotype G was predominantly A at nt 1617, while all others were G (Figure 1B). The core upstream regulatory sequence (CURS, nts 1634-1740), contains positively regulating regions, including the hepatocyte-enriched transcription factor, HNF-4, which can activate the core promoter approximately 20-fold[44]. This segment has been identified to span approximately nts 1648-1672/1682[30, 32, 44]. In a similar overlapping region, CAAT enhancer-binding protein (C/EBP) has also been shown to bind and increase the promoter activity in a dose-dependent manner[45]. When this region is analyzed, there appears to be high conservation

across genotypes A-E, while several variants are present through nts 1645-1649 of genotypes F, G and H.

Another hepatocyte nuclear factor, HNF-3, has three binding sites: one in Enhancer I (nts 1120-1130—further upstream from the promoter) and two in the CURS region (CURS/Enhancer II; nts 1679-1690 and 1713-1723). Binding of HNF-3 to Enhancer I region increases the promoter activity by ~15 fold[46, 47], whereas to CURS region increases promoter activity in a dose-dependent manner[46, 48]. The first of these two CURS segments show high conservation across genotypes A to E, but has significant variations noted for the equivalent region in genotypes F to H. The second region is nearly completely conserved throughout except for nts 1719 and 1721, with two possible variants present at each site, G and T for nt 1719, and G and A for nt 1721. Adjacent to this region is an HFN1 binding element (nts 1721-1734), that can also greatly upregulate replication. In a cohort of patients with severe HBV-related liver disease, clinical variants were identified containing two HFN1 binding regions[30]. Subsequent primer analyses studies showed that duplication of the HFN1 binding region is enough to compensate for loss of Sp1 binding regions[30]. Across genotypes there is general consensus, but with much more variation than in some of the other binding regions. In particular, nt 1721 has a G to A variability in genotypes C, F and H and a G to T variability in genotype G. Similarly, nts 1726-1727 have the greatest degree of variation, even within the genotype (Figure 2B).

In the basic core promoter (BCP), the minimal essential sequence is composed of a 108-bp fragment (nts 1740-1847) [34]. Within this fragment is the direct repeat 1 (DR1) which is required for reverse transcription[34, 49]. The BCP also contains two Specificity protein 1 (Sp1) binding sites that are critical for the transcription of mRNA from the core promoter[50, 51]. The first of these Sp1 regions has a nucleotide variant present at nt 1740 in genotypes A and F, where C appears in approximately 66% of the genomes in the database (genotypes A to F), as opposed to T, which is nearly entirely conserved at this position across all other genotypes. In the second Sp1 binding site, there is essentially 100% agreement across nts 1743-1751 in all but one genotype, genotype G, where there is notable disruption of the G rich region (nts 1745-1748). The terminal nucleotide of this binding region has a notably different variant profile in genotypes G and H, where T is present in place of A. Further downstream from the Sp1 sites is a region containing a second HNF4 binding site flanked by two TATA-like binding protein (TBP) segments, nts 1758-1776[30, 31]. The stacked Logo representation generated for this region suggests this to be the least conserved amongst those assessed in this study, with notable variants present at three positions, nts 1762, 1764, and 1773. Apart from genotype G, all genotypes share high frequency variants at both nt 1762 (A->T) and 1764 (G->A). While genotypes F, B, D, and G, had nucleotide variants visible in the gene logo at nt 1773 (T->C).

3.3. High-risk HCC mutations

The basal core promoter is the region harboring the greatest number of high-risk HCC mutations. With links of these mutations to HCC development[52] and HCC development to genotype[53], it follows that there is likely a propensity of these mutations to be found more commonly in certain genotypes. Single- and multiple-mutational frequencies by genotype are shown in Figure 3. Genotype C sequences analyzed contain the most mutations overall, with a high percentage of A1762T and G1764A mutants, nearly all of those as the double-mutant. Genotypes B and E have significantly less than the average. Genotypes G and H appear to have consensus mutations C1653T and 1766T, respectively, throughout all the available sequences from the database, while little to no other mutations present.

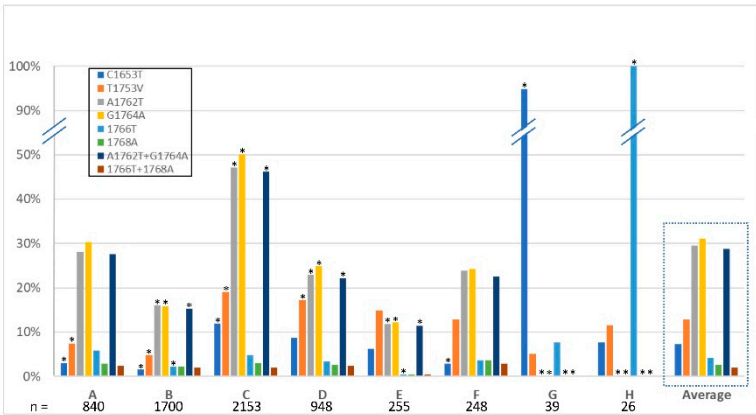


Figure 3. High-risk HCC mutations by genotype. The average of each mutation across all genotypes is outlined by a dashed box. Asterisks refer to significance based on average value.

4. Discussion

Essential to the replication of HBV are the promoter regions of each of the open reading frames (core, preS1, preS2 and X) and enhancers I and II, which interact with host proteins to enable transcription, translation and subsequent viral propagation. Given their key grounding point, promoter regions must maintain a degree of conservation for essential nucleotides in this interaction. Indeed, such is the case found here where key identified binding sites within the promoter regions have an extremely high degree of sequence conservation.

4.1. The X promoter

The X promoter region initiates transcription of the X gene, which creates the 154-amino acid HBx protein. HBx has been shown to enhance replication of HBV and promote its integration into host DNA[54-57]. It also has the ability to manipulate numerous intracellular processes, including DNA repair mechanisms[58], DNA methylation with subsequent downstream cell proliferation, cycling and apoptosis[59, 60], interfere with key cell signaling pathways[61-63] and induce tumor metastases[38, 64, 65]. Given HBx's extensive implications in HBV replication and HBV-induced hepatocellular carcinoma, its promoter may be an important target of HBV therapeutic design[37-42, 66]. In our analysis, apart from a few exceptions, we find highly consistent nucleotide frequencies in the X promoter region throughout all genotypes (Figure 2A).

4.2. Importance of the C gene/HBV Core promoter region.

The Core promoter region is arguably the most important of the HBV promoters as it directs transcription initiation for both the pre-core and pre-genomic RNAs. Mutations in this region would thus have direct impacts on how well the virus is able to replicate.

There are a number of mutations in the core promoter that have been linked with increased risk of HBV-induced HCC. One of the more recognized is that of the natural double mutation in the BCP, A1762T and G1764A. This double mutation prevents the binding of several nuclear receptors while maintaining the binding of HNF4. Simultaneously, this creates a binding site for the transcription factor HNF1 and subsequently affects two amino acids of the HBx protein, the effects of both combinations greatly enhance viral replication[40, 67-69]. The double mutant is also associated with decreased production of HBeAg[70, 71].

Sp1 binding sites are also critical to the transcription of mRNA from the core promoter[50, 51]. In particular, we note the two previously identified Sp1 binding sites and highlight their highly G-rich nature. This region alone harbors a >99% conservation across all but genotype G, which has three fewer G's in this segment (nts 1745, 1746 and 1748) as well as other unique features.

4.2.1. Genotype G has many unique core promoter features.

Genotype G, represented by 39 sequences in the HBV database, has a C1653T nucleotide predominance in the early promoter, possibly creating an additional TATA-binding site (Figure 2). As well, it has much less G-rich Sp1 binding region in the pre-core promoter than the other genotypes, potentially influencing binding of Sp1 in this region (Figure 2). Additionally, it contains a 36-bp insert at the 5' end of the core gene near the ϵ encapsidation signal[72]. This extension results in altered base-pairing, but instead of decreasing efficiency of translation initiation and/or RNA packaging, researchers show that the virus is still replication competent with greatly enhanced core protein production[72]. When the 'extension' was inserted into a similar genotype A plasmid, the corresponding replication markers were also enhanced, supporting the independence of this insertion to enable viral competence[72]. Interestingly, when found clinically, genotype G is nearly always co-infected with genotype A, suggesting a degree of necessity of this genotype for infection, if not maintenance[73-75]. However, given that the occasional monoinfection with genotype G has been described[74, 76, 77], other explanations may be plausible.

Genotype G does not produce e antigen (HBeAg), a protein normally translated from the longer 3.5-kb pre-core mRNA, due to the presence of two stop codons in the pre-core region halting its production[78, 79]. While HBeAg's role in HBV infection remains incompletely elucidated, it has been shown to be heavily linked with numerous immunologic processes and likely contributes to the immune evasion necessary to establish chronic infection [72, 80-82]. Some non-G strains go on to develop an e-antigen negative status over time, but initial phases would otherwise have this. Genotype G is thus likely dependent on other genotypes to establish persistence, but once established, can eventually out-compete its supporting genotype (*i.e.*, genotype A) to create a monoinfection state in a host[79, 83-85].

Studying genotype G's many unique features may give further insight into HBV's replication process. Despite HBeAg production and a lack of a key G-rich pre-core promoter region, genotype G would appear to have further modifications and symbioses that make this strain sustainable.

4.2.2. Other core promoter interactions

Additional factors have been found to influence HBV transcription, and thus inferred interaction with the BCP. These include testicular orphan receptor 4 (TR4) and chicken ovalbumin upstream promoter transcription factor (COUP-TF), which have been shown to competitively inhibit other factors of the core promoter *in vitro* through direct binding at regions described above, but it is unclear as to the degree this influences the *in vivo* situation[32, 35, 86]. Additional factors, peroxisome proliferator-activated receptors (PPAR) and retinoid X receptors (RXR), also appear to influence transcription through the core promoter, though it would appear to be more of an indirect influence on HNF4 and not directly with the promoter nucleotides themselves [32, 87].

4.3. High-risk HCC mutations.

Studies employing clinical isolates with associated clinical data, including disease severity and presence of HCC have linked numerous mutations with that of a greater risk of developing cancer[34, 52, 88]. As well, the link between HCC risk and that of genotype have also been made, *i.e.*, per the Taiwanese REVEAL study, genotype C has a higher chance of developing HCC, with a hazard ratio of 2.99 when the double-mutation A1762T/G1764A is present[53]. While the database lacks clinical outcome data, we are able to quantify relationships between strongly-correlated high-risk HCC mutations and genotype (Figure 3, Supplementary Table 2). The database, a representation of the currently available clinical isolates, is consistent with prior findings, showing a high prevalence of mutations, including the double-mutant, in genotype C. It was also notable that there was a high rate of the triple-mutation T1753V/A1762T/G1764A in genotype C. Further, the analyses also revealed significantly fewer mutations appearing in genotypes B and E.

4.4. Study Limitations.

One limitation of the study is the reliability of the sequences deposited in the HBVdb. While the database has control measures in place to provide quality annotations and sequence information, its sequences are provided by researchers with varying sequencing expertise. Also, while consensus sequences are typically compiled from many reads of a single piece of DNA, our attempt at creating a “consensus” will be skewed by using thousands of sequences across eight genotypes. This will inherently overlook unique genotypic features (i.e., genotype G’s 34-bp insertion) and give stronger weight to the more abundant genotypes (i.e., genotype C), and thus conclusions drawn therein will have to take such biases into consideration. Lastly, given the limited number of sequences available for genotypes G and H, largely due to epidemiologic and endemnicity factors, inferences from the above analyses will also be limited.

4.5. Future perspectives.

Applying the findings of this work to further applications, one may be able to narrow the focus certain host protein-interactions to a more well-defined nucleotide region that persists throughout all competent viruses. Through comprehensive genotype comparisons of HBV, it may be possible to adapt or develop therapeutics for pan-genotypic applications. As well, mutations in host protein binding regions associated with hepatocellular carcinoma could be exploited with structural studies or binding assays to reveal unique pathogenicity features. Lastly, the approach described in this paper may be applied to investigate other viral nucleotide regions involved in host protein interactions.

5. Conclusions

Our frequency analyses of the HBV sequences available in the HBVdb demonstrate a high degree of conservation across all promoters examined. In particular, regions with known host-protein binding have nucleotide conservation of over 99% strongly suggesting the importance of these specific nucleotides in the interaction with the host cell. Such knowledge could be capitalized on, narrowing focus of a study interaction, and importantly aiding the study of how the interaction may differ from that of the host protein’s usual target(s). Further, it can provide insights into regions in the HBV genome that can be targeted for a selective yet pan-genotypic rational drug design for treatment of HBV infection.

Supplementary Materials: The following are available online, Table S1: X and core promoter variant frequencies.

Author Contributions: Conceptualization, VMS, GvM, CSC and TRP; Methodology, VMS, GvM, TRP, Software, VMS, WTRB, CSD; Formal Analysis, VMS, WTRB, CSD; Investigation, CSC, TRP; Data Curation, WTRB; Writing-Original Draft Preparation, VMS, TRP; Writing-Review & Editing, VMS, GvM, CSC, TRP; Supervision, CSC, TRP; Project Administration, VMS, CSC, TRP; Funding Acquisition, VMS, TRP.

Funding: VMS is a recipient of CanHepC Fellowship. TRP is a Canada Research Chair in RNA & Protein Biophysics.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. WHO. *Global Hepatitis Report, 2017*. 2018 [cited 2018 Jan, 2018]; Available from: <http://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/>.
2. Yang, H.C. and J.H. Kao, *Persistence of hepatitis B virus covalently closed circular DNA in hepatocytes: molecular mechanisms and clinical significance*. *Emerg Microbes Infect*, 2014. 3(9): p. e64.
3. Abdelhamed, A.M., et al., *Rebound of hepatitis B virus replication in HepG2 cells after cessation of antiviral treatment*. *J Virol*, 2002. 76(16): p. 8148-60.
4. Tang C. M., Yau T. O., and Y. J., *Management of chronic hepatitis B infection: current treatment guidelines, challenges, and new developments*. *World J Gastroenterol*, 2014. 20: p. 6262–6278.

5. Zoulim, F., F. Lebosse, and M. Levrero, *Current treatments for chronic hepatitis B virus infections*. *Curr Opin Virol*, 2016. **18**: p. 109-116.
6. Arends, J.E., et al., *New Viral and Immunological Targets for Hepatitis B Treatment and Cure: A Review*. *Infectious Diseases and Therapy*, 2017. **6**(4): p. 461-476.
7. Durantel, D. and F. Zoulim, *New antiviral targets for innovative treatment concepts for hepatitis B virus and hepatitis delta virus*. *J Hepatol*, 2016. **64**(1 Suppl): p. S117-S131.
8. Testoni, B., D. Durantel, and F. Zoulim, *Novel targets for hepatitis B virus therapy*. *Liver Int*, 2017. **37** Suppl 1: p. 33-39.
9. Lucifora, J. and U. Protzer, *Attacking hepatitis B virus cccDNA - The holy grail to hepatitis B cure*. *J Hepatol*, 2016. **64**(1 Suppl): p. S41-8.
10. Sarah, C., *Targeting host kinases*. *Nature Reviews Drug Discovery*, 2017. **16**: p. 314.
11. Alonso, S., et al., *Upcoming pharmacological developments in chronic hepatitis B: can we glimpse a cure on the horizon?* *BMC Gastroenterol*, 2017. **17**(1): p. 168.
12. Périer, R.C., T. Junier, and P. Bucher, *The Eukaryotic Promoter Database EPD*. *Nucleic Acids Research*, 1998. **26**(1): p. 353-357.
13. Dreos, R., et al., *EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era*. *Nucleic Acids Research*, 2013. **41**(D1): p. D157-D164.
14. *Promoter Mutations and Cancer*. *Science*, 2013. **339**(6122): p. 882-882.
15. Cuykendall, T.N., M.A. Rubin, and E. Khurana, *Non-coding genetic variation in cancer*. *Current Opinion in Systems Biology*, 2017. **1**: p. 9-15.
16. Heng, J., et al., *Integrated analysis of promoter mutation, methylation and expression of AKT1 gene in Chinese breast cancer patients*. *PLOS ONE*, 2017. **12**(3): p. e0174022.
17. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. *Cell*, 2011. **144**(5): p. 646-74.
18. Wang, M.Y., et al., *A Functional Polymorphism (rs2494752) in the AKT1 Promoter Region and Gastric Adenocarcinoma Risk in an Eastern Chinese Population*. *Sci Rep*, 2016. **6**: p. 20008.
19. Griewank, K.G., et al., *TERT promoter mutation status as an independent prognostic factor in cutaneous melanoma*. *J Natl Cancer Inst*, 2014. **106**(9).
20. Borah, S., et al., *Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer*. *Science*, 2015. **347**(6225): p. 1006-10.
21. Kudo, E., et al., *Transcriptional regulation of HIV-1 host factor COMMD1 by the Sp family*. *Int J Mol Med*, 2018. **41**(4): p. 2366-2374.
22. Chong, T., W.K. Chan, and H.U. Bernard, *Transcriptional activation of human papillomavirus 16 by nuclear factor I, AP1, steroid receptors and a possibly novel transcription factor, PVF: a model for the composition of genital papillomavirus enhancers*. *Nucleic Acids Res*, 1990. **18**(3): p. 465-70.
23. Hubert, W.G., T. Kanaya, and L.A. Laimins, *DNA Replication of Human Papillomavirus Type 31 Is Modulated by Elements of the Upstream Regulatory Region That Lie 5' of the Minimal Origin*. *Journal of Virology*, 1999. **73**(3): p. 1835-1845.
24. Apt, D., et al., *High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter*. *Virology*, 1996. **224**(1): p. 281-91.
25. Seeger, C. and W.S. Mason, *Hepatitis B virus biology*. *Microbiol Mol Biol Rev*, 2000. **64**(1): p. 51-68.
26. Levrero, M., et al., *Control of cccDNA function in hepatitis B virus infection*. *J Hepatol*, 2009. **51**(3): p. 581-92.
27. Nassal, M., *HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B*. *Gut*, 2015. **64**(12): p. 1972-84.
28. Nowak, M.A., et al., *Viral dynamics in hepatitis B virus infection*. *Proceedings of the National Academy of Sciences of the United States of America*, 1996. **93**(9): p. 4398-4402.
29. Hayer, J., et al., *HBVdb: a knowledge database for Hepatitis B Virus*. *Nucleic Acids Research*, 2013. **41**(D1): p. D566-D570.
30. Gunther, S., et al., *Type, prevalence, and significance of core promoter/enhancer II mutations in hepatitis B viruses from immunosuppressed patients with severe liver disease*. *J Virol*, 1996. **70**(12): p. 8318-31.
31. Ishida, H., et al., *Identification of Multiple Transcription Factors, HLF, FTF, and E4BP4, Controlling Hepatitis B Virus Enhancer II*. *Journal of Virology*, 2000. **74**(3): p. 1241-1251.
32. Raney, A.K., et al., *Members of the nuclear receptor superfamily regulate transcription from the hepatitis B virus nucleocapsid promoter*. *J Virol*, 1997. **71**(2): p. 1058-71.
33. Crooks, G.E., et al., *WebLogo: a sequence logo generator*. *Genome Res*, 2004. **14**(6): p. 1188-90.

34. Quarleri, J., *Core promoter: a critical region where the hepatitis B virus makes decisions*. *World J Gastroenterol*, 2014. **20**(2): p. 425-35.
35. Quasdorff, M. and U. Protzer, *Control of hepatitis B virus at the level of transcription*. *J Viral Hepat*, 2010. **17**(8): p. 527-36.
36. Tokusumi, Y., S. Zhou, and S. Takada, *Nuclear respiratory factor 1 plays an essential role in transcriptional initiation from the hepatitis B virus x gene promoter*. *J Virol*, 2004. **78**(20): p. 10856-64.
37. Li, J., et al., *Regulation of hepatocyte nuclear factor 1 activity by wild-type and mutant hepatitis B virus X proteins*. *J Virol*, 2002. **76**(12): p. 5875-81.
38. Jin, Y., et al., *Hepatitis B virus x protein induces epithelial-mesenchymal transition of hepatocellular carcinoma cells by regulating long non-coding RNA*. *Virology*, 2017. **14**(1): p. 238.
39. Sekiba, K., et al., *Hepatitis B virus pathogenesis: Fresh insights into hepatitis B virus RNA*. *World J Gastroenterol*, 2018. **24**(21): p. 2261-2268.
40. Zheng, Y., J. Li, and J.H. Ou, *Regulation of hepatitis B virus core promoter by transcription factors HNF1 and HNF4 and the viral X protein*. *J Virol*, 2004. **78**(13): p. 6908-14.
41. Kew, M.C., *Hepatitis B virus x protein in the pathogenesis of hepatitis B virus-induced hepatocellular carcinoma*. *J Gastroenterol Hepatol*, 2011. **26 Suppl 1**: p. 144-52.
42. Lupberger, J. and E. Hildt, *Hepatitis B virus-induced oncogenesis*. *World J Gastroenterol*, 2007. **13**(1): p. 74-81.
43. Sun, C.T., et al., *Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON*. *J Biol Chem*, 2001. **276**(26): p. 24059-67.
44. Guo, W., et al., *Hepatocyte-specific expression of the hepatitis B virus core promoter depends on both positive and negative regulation*. *Mol Cell Biol*, 1993. **13**(1): p. 443-8.
45. Lopez-Cabrera, M., et al., *Transcriptional factor C/EBP binds to and transactivates the enhancer element II of the hepatitis B virus*. *Virology*, 1991. **183**(2): p. 825-9.
46. Chen, M., et al., *Regulation of hepatitis B virus ENI enhancer activity by hepatocyte-enriched transcription factor HNF3*. *Virology*, 1994. **205**(1): p. 127-32.
47. Cho, E.-y., et al., *Impact of Nucleotide Mutations at the HNF3- and HNF4-Binding Sites in Enhancer 1 on Viral Replication in Patients with Chronic Hepatitis B Virus Infection*. *Gut and Liver*, 2013. **7**(5): p. 569-575.
48. Li, M., et al., *HNF3 binds and activates the second enhancer, ENII, of hepatitis B virus*. *Virology*, 1995. **214**(2): p. 371-8.
49. Yuh, C.H., Y.L. Chang, and L.P. Ting, *Transcriptional regulation of precore and pregenomic RNAs of hepatitis B virus*. *J Virol*, 1992. **66**(7): p. 4073-84.
50. Zhang, P. and A. McLachlan, *Differentiation-specific transcriptional regulation of the hepatitis B virus nucleocapsid gene in human hepatoma cell lines*. *Virology*, 1994. **202**(1): p. 430-40.
51. Zhang, P., A.K. Raney, and A. McLachlan, *Characterization of functional Sp1 transcription factor binding sites in the hepatitis B virus nucleocapsid promoter*. *J Virol*, 1993. **67**(3): p. 1472-81.
52. Liu, S., et al., *Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis*. *J Natl Cancer Inst*, 2009. **101**(15): p. 1066-82.
53. Yang, H.I., et al., *Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma*. *J Natl Cancer Inst*, 2008. **100**(16): p. 1134-43.
54. Bagga, S., et al., *Hepatitis B virus (HBV) X protein-mediated regulation of hepatocyte metabolic pathways affects viral replication*. *Virology*, 2016. **498**: p. 9-22.
55. Bouchard, M.J. and R.J. Schneider, *The Enigmatic X Gene of Hepatitis B Virus*. *Journal of Virology*, 2004. **78**(23): p. 12725-12734.
56. Zhang, X., et al., *Involvement of hepatitis B virus X gene (HBx) integration in hepatocarcinogenesis via a recombination of HBx/Alu core sequence/subtelomeric DNA*. *FEBS Letters*, 2012. **586**(19): p. 3215-3221.
57. Tu, T., et al., *HBV DNA Integration: Molecular Mechanisms and Clinical Implications*. *Viruses*, 2017. **9**(4): p. 75.
58. Cheng, B., et al., *Hepatitis B viral X protein alters the biological features and expressions of DNA repair enzymes in LO2 cells*. *Liver Int*, 2010. **30**(2): p. 319-26.
59. Wei, X., et al., *miR-101 is down-regulated by the hepatitis B virus x protein and induces aberrant DNA methylation by targeting DNA methyltransferase 3A*. *Cell Signal*, 2013. **25**(2): p. 439-46.
60. Tian, Y., et al., *Hepatitis B virus X protein-induced aberrant epigenetic modifications contributing to human hepatocellular carcinoma pathogenesis*. *Mol Cell Biol*, 2013. **33**(15): p. 2810-6.
61. Hsieh, A., et al., *Hepatitis B viral X protein interacts with tumor suppressor adenomatous polyposis coli to activate Wnt/beta-catenin signaling*. *Cancer Lett*, 2011. **300**(2): p. 162-72.

62. Lim, K.H., et al., *HBx-induced NF-kappaB signaling in liver cells is potentially mediated by the ternary complex of HBx with p22-FLIP and NEMO*. PLoS One, 2013. **8**(3): p. e57331.
63. Luo, J., et al., *The hepatitis B virus X protein downregulates NF-kappaB signaling pathways through decreasing the Notch signaling pathway in HBx-transformed L02 cells*. Int J Oncol, 2013. **42**(5): p. 1636-43.
64. Han, J.M., et al., *Peroxisome-localized hepatitis Bx protein increases the invasion property of hepatocellular carcinoma cells*. Arch Virol, 2014. **159**(10): p. 2549-57.
65. Geng, M., et al., *Molecular mechanism of hepatitis B virus X protein function in hepatocarcinogenesis*. World Journal of Gastroenterology : WJG, 2015. **21**(38): p. 10732-10738.
66. Murphy, Christopher M., et al., *Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication*. Cell Reports. **16**(11): p. 2846-2854.
67. Buti, M., et al., *Hepatitis B virus genome variability and disease progression: the impact of pre-core mutants and HBV genotypes*. J Clin Virol, 2005. **34 Suppl 1**: p. S79-82.
68. Parekh, S., et al., *Genome replication, virion secretion, and e antigen expression of naturally occurring hepatitis B virus core promoter mutants*. J Virol, 2003. **77**(12): p. 6601-12.
69. Jammeh, S., et al., *Effect of basal core promoter and pre-core mutations on hepatitis B virus replication*. J Gen Virol, 2008. **89**(Pt 4): p. 901-9.
70. Moriyama, K., et al., *Reduced precore transcription and enhanced core-pregenome transcription of hepatitis B virus DNA after replacement of the precore-core promoter with sequences associated with e antigen-seronegative persistent infections*. Virology, 1996. **226**(2): p. 269-80.
71. Buckwold, V.E., et al., *Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication*. J Virol, 1996. **70**(9): p. 5845-51.
72. Li, K., 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): p. 9202-15.
73. Cornelissen, M., et al., *Widespread hepatitis B virus genotype G (HBV-G) infection during the early years of the HIV epidemic in the Netherlands among men who have sex with men*. BMC Infectious Diseases, 2016. **16**: p. 268.
74. van der Kuyl, A.C., et al., *High prevalence of hepatitis B virus dual infection with genotypes A and G in HIV-1 infected men in Amsterdam, the Netherlands, during 2000-2011*. BMC Infect Dis, 2013. **13**: p. 540.
75. Osioy, C., et al., *Hepatitis B virus genotype G epidemiology and co-infection with genotype A in Canada*. J Gen Virol, 2008. **89**(Pt 12): p. 3009-15.
76. Chudy, M., et al., *Hepatitis B virus genotype G mono-infection and its transmission by blood components*. Hepatology, 2006. **44**(1): p. 99-107.
77. Zaaijer, H.L., et al., *HBsAg-negative mono-infection with hepatitis B virus genotype G*. J Viral Hepat, 2011. **18**(11): p. 815-9.
78. Okamoto, H., 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): p. 8102-10.
79. Kato, H., et al., *Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G*. Hepatology, 2002. **35**(4): p. 922-9.
80. Rehmann, B. and A. Bertolotti, *Immunological aspects of antiviral therapy of chronic hepatitis B virus and hepatitis C virus infections*. Hepatology, 2015. **61**(2): p. 712-21.
81. Lu, B., et al., *Hepatitis B Virus e Antigen Regulates Monocyte Function and Promotes B Lymphocyte Activation*. Viral Immunology, 2017. **30**(1): p. 35-44.
82. Yu, Y., et al., *Hepatitis B Virus e Antigen Activates the Suppressor of Cytokine Signaling 2 to Repress Interferon Action*. Scientific Reports, 2017. **7**(1): p. 1729.
83. Kato, H., et al., *Characteristics of Hepatitis B Virus Isolates of Genotype G and Their Phylogenetic Differences from the Other Six Genotypes (A through F)*. J Virol, 2002. **76**(12): p. 6131-7.
84. Sugiyama, M., et al., *Early dynamics of hepatitis B virus in chimeric mice carrying human hepatocytes mono-infected or co-infected with genotype G*. Hepatology, 2007. **45**(4): p. 929-37.
85. Tran, A., et al., *Emergence of and takeover by hepatitis B virus (HBV) with rearrangements in the pre-S/S and pre-C/C genes during chronic HBV infection*. J Virol, 1991. **65**(7): p. 3566-74.
86. Lin, W.J., et al., *Suppression of hepatitis B virus core promoter by the nuclear orphan receptor TR4*. J Biol Chem, 2003. **278**(11): p. 9353-60.
87. Yu, X. and J.E. Mertz, *Differential regulation of the pre-C and pregenomic promoters of human hepatitis B virus by members of the nuclear receptor superfamily*. J Virol, 1997. **71**(12): p. 9366-74.

- 482 88. Papatheodoridis, G.V., et al., *Incidence of hepatocellular carcinoma in chronic hepatitis B patients receiving*
483 *nucleos(t)ide therapy: a systematic review*. J Hepatol, 2010. 53(2): p. 348-56.