Asymmetric modification of HBV genomes by an endogenous cytidine deaminase inside HBV cores informs a model of reverse transcription

Smita Nair and Adam Zlotnick*

Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405

* Corresponding author:
Adam Zlotnick
Molecular and Cellular Biochemistry
Indiana University
Simon Hall MSB 220
212 S Hawthorne Dr
Bloomington, Indiana 47405-7003
(812) 856-1925
azlotnic@indiana.edu

Abstract:

Cytidine deaminases inhibit replication of broad range of DNA viruses by deaminating cytidines on single stranded DNA to generate uracil. While several lines of evidence have revealed HBV genome editing by deamination, it is still unclear which nucleic acid intermediate of HBV is modified. Hepatitis B virus has a relaxed circular double-stranded DNA (rcDNA) genome that is reverse transcribed within virus cores from a RNA template. The HBV genome also persists as covalently closed circular DNA (cccDNA) in the nucleus of an infected cell. In the present study, we find that in HBV-producing HepAD38 and Hep2.2.15 cell lines, endogenous cytidine deaminases edited 10-25% of HBV rcDNA genomes, asymmetrically with almost all mutations on the 5' half of the minus strand. This region corresponds to the last half of the minus strand to be protected by plus strand synthesis. Within this half of the genome, the number of mutations peaks in the middle. Over-expressed APOBEC3A and APOBEC3G could be packaged in HBV capsids but did not change the amount or distribution of mutations. We found no deamination on pgRNA indicating that an intact genome is encapsidated and deaminated during or after reverse transcription. The deamination pattern suggests a model of rcDNA synthesis where pgRNA and then newly synthesized minus-sense single stranded DNA are protected from deaminase by interaction with the virus capsid; during plus strand synthesis, when enough dsDNA has been synthesized to displace the remaining minus strand from the capsid surface that single stranded DNA becomes deaminase-sensitive.

Importance

Host-induced mutation of the HBV genome, as by APOBEC proteins, may be a path to clearing the virus. We examined Cytidine to Thymidine mutations in the genomes of HBV particles grown in the presence or absence of overexpressed APOBEC proteins. We found that genomes were subjected to deamination

Journal of Virology
JVI
Journal of Virology

activity during reverse transcription, which takes place within the virus capsid. These observations provide a direct insight into the mechanics of reverse transcription, suggesting that newly synthesized dsDNA displaces ssDNA from the capsid walls making the ssDNA accessible to deaminase activity.

48 **Introduction**

49 Hepatitis B virus causes acute and chronic viral hepatitis. Over 240 million individuals have chronic HBV
50 and HBV claims over 750,000 lives every year. HBV is an enveloped, DNA virus that belongs to
51 Hepadnaviridae family. The genome is a 3.2 kb, partially double stranded circular DNA (relaxed circular
52 DNA, rcDNA), with breaks on both strands, that replicates via an RNA intermediate (1). The virus enters
53 a host cell by using the Na^+ Taurocholate co-transporting polypeptide as a receptor (2). The icosahedral
54 viral core is then transported to the nucleus where it delivers the rcDNA genome. The rcDNA is repaired
55 in the nucleus to form the double stranded covalently closed circular DNA (cccDNA) that is the template
56 for an over-length transcript of the genome- the pregenomic RNA (pgRNA) and sub-genomic RNAs.
57 pgRNA is exported into the cytoplasm where it translates coat protein and polymerase. The core protein
58 encapsidates a pgRNA-polymerase complex forming an immature RNA-filled icosahedral core. The
59 polymerase reverse transcribes the pgRNA to form the rcDNA. This is a complex reaction where a
60 complete minus strand is synthesized using the polymerase as a specific protein primer and most of the
61 pgRNA template is digested, leaving the polymerase covalently bound to the 5' end; the remaining RNA
62 is used to prime plus strand synthesis, a reaction requiring two distinct template transfers. Structural
63 studies of RNA-containing capsids show that the nucleic acid forms a thin shell closely associated with
64 the basic C-terminal domain of the core protein (3-5); the polymerase is perched on the RNA and is
65 hypothesized to perform reverse transcription by traveling on an "RNA track" (4). Mature nucleocapsid
66 core may be enveloped and secreted. Interestingly, about 90% of the secreted virions are genome free
67 (6). Alternatively, the mature nucleocapsid can be recycled back to the nucleus to maintain the cccDNA
68 pool that enables HBV to persist in an infected cell- hallmark of a chronic infection.

69 Present therapies against chronic infection includes interferon-α (IFN-α) treatment and nucleotide
70 analogs (7-11). Nucleotide analogs keep HBV infection at bay but fail to eliminate the resident cccDNA
71 pool (10, 12, 13). IFN-α treatment can lead to sustained virological response (SVR) but only in a small

72 subset of patients: it is most efficacious on those who were infected as adults, have had chronic
73 infection for a relatively short time, and have relatively high liver inflammation (13). Understanding its
74 mechanism of action may make important contributions to HBV therapy. Among the widespread array
75 of IFN-α stimulated genes that are activated, the APOBEC3 family of nucleic acid editing enzymes have
76 gained special interest (14, 15).

77 APOBEC3 (A3) proteins are host restriction factors that can be packaged into the retroviral virions and
78 introduce mutations in the viral genome by deaminating cytosine to uracil, thereby restricting the viral
79 replication (16). The catalytic domain of A3 deaminases is comprised of the hallmark H-X-E active site
80 motif and the Zinc finger motif P-C-$X_{(2-4)}$-C (17). The active site glutamate is involved in proton transfer
81 during catalysis and the zinc ion coordinates a catalytic water molecule (18). A3 deaminases can have
82 one (A3A, A3C, A3H) or two such domains (A3B, A3F, A3G). In the latter case, only one domain is active
83 and the other is involved in packaging into virions (19, 20); A3A is the most potent among all deaminase
84 isoforms. Members of the APOBEC deaminase family are either localized to the cytoplasm or shuttle
85 between the nucleus and the cytoplasm during the cell cycle and can effectively target single stranded
86 DNA (ssDNA) substrates in either compartment (21). Though APOBECs have a preference for ssDNA as
87 substrate, some APOBECs like APOBEC1, APOBEC3A can also deaminate RNA (22, 23).

88 HBV has been reported to be restricted by several members of A3 family of proteins by both
89 deaminase-dependent and -independent mechanisms (24). HBV inhibition by A3G, independent of
90 deaminase activity, was attributed to inhibition of either minus strand synthesis or pgRNA packaging
91 (25-28). Likewise, evidence of HBV genome editing has been reported in infected patient samples and
92 samples from transfected hepatoma cell lines (29-34). It is still unclear which HBV nucleic acid
93 intermediates could be the substrates for APOBEC mediated deamination or which APOBECs catalyze
94 deamination. One report claimed that Interferon- or cytokine-induced activation of nuclear-localized

95 deaminases A3A and A3B in infected cells could specifically degrade cccDNA while being non-genotoxic
96 (14).

97 In the present study, we examined APOBEC packaging in virions and analyzed the possible substrates of
98 APOBEC deamination in HBV. Since HBV nucleic acid intermediates span cellular compartments and
99 comprise of double stranded DNA, RNA and partially double stranded DNA, we analyzed the effect of
100 the potent APOBEC isoform- A3A that is known to be nucleocytoplasmic and edit cytosines in RNA and
101 DNA and methyl cytosines in DNA (21, 23, 35). We found that virus produced from stably transfected
102 cell lines HepAD38 and HepG2.215 packaged overexpressed A3A but that a host-derived deaminase, not
103 A3A, was the source of deaminase activity. Reverse transcribed DNA within HBV capsids is the primary
104 target for deamination. The peculiar pattern of deamination suggests mechanistic features of the
105 reverse transcription reaction.

106

107 **Results:**

108 **Viral encapsidation of APOBEC3A**

109 Intracellular HBV cores can contain ssDNA or rcDNA synthesized during pgRNA reverse transcription; this
110 ssDNA could be an ideal substrate for APOBEC3-catalyzed deamination if APOBEC were to be packaged
111 in the actively transcribing cores. To test if A3A could be packaged in HBV cores, we transiently
112 transfected virus-producing HepAD38 cells with A3A expression plasmid. The virus in supernatant was
113 harvested 4 days post transfection and partially purified through a 20% sucrose cushion. The
114 resuspended viral pellet was assayed for A3A that co-sedimented and was presumably contained inside
115 the HBV cores (Figure 1). As a positive control, we also tested for packaging of A3G, which is reported to
116 be packaged in replication competent HBV cores and inhibit reverse transcription (36). Our results

117 suggest that both A3A and A3G are packaged in cores and thus could have access to the reverse
118 transcription reaction.

119 The basis of A3A packaging is unclear. It could be modulated by nucleic acid, or by interaction with a
120 nucleoprotein complex including viral protein. The zinc finger motif of the pseudo catalytic N-terminal
121 domain of A3G is known to be involved in packaging in HIV (19). We thus tested the packaging of the
122 A3A active site E72A mutant and the A3A zinc finger C101A-C106A double mutant. Both mutants are
123 catalytically inactive, although, E72A mutant retains the protein fold and DNA binding activity and
124 C101A-C106A does not (37, 38). We found about the same level of incorporation of both mutant
125 proteins in the HBV fraction (Figure 1), suggesting that packaging was not dependent on catalytic activity
126 or correct folding of the substrate binding site. While this experiment demonstrates packaging of
127 APOBECs, it does not provide a stoichiometry.

128 **C to T mutations in viral genomes from HepAD38 culture supernatant**

129 To test if the over-expressed A3A led to deaminated HBV nucleic acid inside the nucleocapsid, we
130 analyzed the genomic rcDNA for G to A and C to T mutations. Previously reported mutational analyses
131 have been limited to the region coding for the X protein, so in the present study we sequenced the
132 entire genome. We amplified the genome as two large amplicons – region I (1-1630) and II (1631-3182).
133 During reverse transcription, both regions will transiently be found as a single stranded minus strand.
134 The synthesis of the plus strand will first occlude region II. The plus strand for region I is synthesized
135 later and contains single stranded region owing to incomplete plus strand synthesis (Figure 2). The
136 length of this single stranded region may vary from virus to virus. DNA was extracted from HepAD38
137 culture supernatant pellet and the two regions were PCR amplified with specific primers, using Taq DNA
138 polymerase, and TOPO cloned. This approach allowed us to estimate the frequency of genomes with

defects as well as the frequency of defects in a given amplicon. As the amplicons are replicated DNA, we actually observe C to T and G to A mutations; these are reported as C to T for consistency.

Table 1 reports C to T mutations on the minus strand. The background mutation rate observed for nucleotides other than C to T was 0.02% or less, thus any clone with two or more C to T mutations was considered a possible product of cytidine deamination. On average, C to T mutations were present in about 10% of the clones sequenced, independent of the presence of APOBEC expression (Table 1). No two mutant clones sequenced had the exact same set of mutations suggesting each arose from a unique set of deamination events. There were no C to T mutations on the plus strand beyond background. Although, only 10% of the clones were mutated, the frequency of C to T mutations within affected clones ranged from 8.1-11% in region I, taken as a group 31.3 ± 16.6 out of 358 cytosines were mutated. In region II the frequencies were much lower- about 0.7-1.8 % (4.6 ± 3.4 out of 334 cytosines). Based on Student's unpaired T-test, the mutation rates in regions I and II were statistically significant ($p = 0.0014$). For all genomes sequenced from culture supernatant, about 1% of the total C residues were mutated in region I versus 0.2% in region II (Figure 3b). It should be noted that the total number of Cs or complementary Gs are about the same in the two regions- 358 versus 334. We did not detect any deamination on the plus strand above background indicating that deamination is essentially limited to the minus strand, which is at least transiently single stranded. As an additional control, we also amplified and cloned the A3A coding region from cell lysates of HepAD38 transfected with the A3A plasmid. No mutations were detected in the sequence in either strand, ruling out a major role for sequencing artifacts.

As a working hypothesis, these observations can be explained by the preference of APOBEC3 proteins for single stranded substrates. During reverse transcription the minus strand, without any plus-sense DNA, is an intermediate that transiently accumulates. Region I minus strand DNA is made first and covered up last with a ssDNA stretch that persists after reverse transcription for at least some of the

genomes. This hypothesis is consistent with our observation that there were very few mutations in region II which forms the double stranded region of the genome.

Critically, genomes from HepAD38 culture supernatant produced in the absence of any transfected APOBECs showed a similar frequency of C to T and G to A mutations as genomes from supernatant produced in the presence of APOBECs (Table 1, Figure 3). The results indicate that HBV produced in HepAD38 cells are susceptible to deamination by endogenous cytidine deaminases. Furthermore, while transfected APOBECs were packaged, to our surprise they did not contribute significantly to the amount of deamination. It is notable that while A3G overexpression decreased the nucleocapsid DNA levels by 50% there was no significant change in the nucleocapsid DNA levels with A3A overexpression (data not shown).

C to T mutations in viral genomes from Hep2.2.15 culture supernatant

To rule out the possibility that deamination was peculiar to the HepAD38 cell line, we analyzed viral supernatant from another stably-transfected HBV-producing cell line, HepG2.2.15. HepG2.2.15 cells were transfected with A3A, A3G, or a control plasmid; the media from four days post transfection was pelleted and DNA was extracted from the virus-containing pellet; regions I and II amplified and cloned. Sequencing results of DNA from Hep2.2.15 supernatant, mirrored those from HepAD38 supernatant. C to T mutations were primarily on the minus strand in region I (i.e. G to A mutations were observed on the plus strand). Like the HepAD38 experiments, viral genomes from untransfected Hep2.2.15 culture supernatant also showed C to T mutations at about the same frequency as APOBEC-transfected cells (Figure 4). About 25-30% of genomes sequenced, had mutations on the minus strand (Table 2). In region I, the frequency of mutations in the mutants was about 10% (35.8 ± 15.6 out of 358) versus 1.2% in region II (4.1 ± 2.7 out of 334), which is statistically significant ($p < 0.0001$). Overall, about 3% of the total C residues in region I were mutated while in region II about 0.5% were. The presence of transfected

APOBECs did not modulate the mutation rate. This indicates that these viral genomes were prone to deamination by endogenous cytidine deaminases and that transfected and packaged A3A and A3G contributed insignificantly to the number of mutations. The frequency of mutations in region I was always greater than for region II (Figure 4). These data again suggest that deamination primarily occurs after reverse transcription of the minus strand and may continue in the partially single stranded region in the mature virus.

Genome-Wide Mutation profile

In order to understand the genome wide distribution of the mutations in region I and II, we mapped all C to T mutations identified in genomes from both cell culture supernatants, to the entire length of the HBV genome (Figure. 5). This genome-wide mutation profile reveals a non-uniform and asymmetric distribution between the two halves of the genome and an additional asymmetry within the region I (1-1630). The 5' half of plus strand (1630-3182) showed modest amounts of editing- 50% or more clones remained unedited; mutations were mostly uniformly distributed when compared to the heavily mutated 3' half (1-1630). Within this latter half (1-1630) we observe that the number of mutations that peaks from 550 nts to 1100 nt ca. and surprisingly, tapers off towards the end of the genome. This implies that not all single stranded regions of the genome are accessible to the deaminase.

Trinucleotide Context

APOBEC3G preferentially modifies hot spots comprised of a cytosine-rich cluster in a relatively disordered ssDNA region (39). We examined trinucleotide preferences of mutated sites. Analysis of the sequence context of the deaminated sites in region I of viral genomes from HepAD38 culture supernatant, revealed that 5'CCA/C was the preferred trinucleotide target followed by 5'TCC (Figure 6). In these trinucleotides, it is the central base that is deaminated. Again, for region II there was preference for 5'CCC/T and TCC/T sites (Figure 6). This preference was likewise maintained in deaminated sites in

region I and II of viral genomes from HepG2.2.15 culture supernatant. Seemingly, HBV rcDNA could be deaminated by more than one endogenous APOBEC. Targeted sites are consistent with any of the following APOBEC proteins – A1, A3A, A3F, A3G, and A3H (40-43). Our attempts to detect these deaminases in secreted cores by immunoblots were unsuccessful. This may not be too surprising if there is only one active deaminase per affected virion, only 10% of virions have a deaminase, and only 10% of virions package a genome (6) resulting in only one deaminase per 24,000 copies of core protein. The absence of additional deamination arising from overexpression of A3A and A3G suggest that these are not the source of hyper-mutated genomes in HepAD38 and HepG2.2.15-derived HBV.

C to T mutation screening in total DNA and total RNA

An alternative to our working hypothesis was that pgRNA or integrated viral DNA could be targeted by some nuclear deaminases, like APOBEC3A and 3B, prior to packaging the pgRNA. To test for RNA modification and evidence for modification of integrated DNA we sequenced total RNA from the transfected HepAD38 cells that produced edited genomes in the culture supernatant. We extracted total RNA from this virus producing cell line and reverse transcribed it. The sequencing of the corresponding cDNA revealed no detectable C to T mutations above the background (<0.01% of non C-T) even in the presence of transfected and over-expressed A3A and A3G. The results indicate that pgRNA is not a target of deamination by the endogenous APOBECs or the overexpressed A3A and A3G, that integrated DNA is unmodified, and that an intact genome is packaged for reverse transcription. We note that HepAD38 cells accumulate cccDNA, as evidenced by the appearance of the HBeAg protein; the absence of deaminated RNA in these cells suggest that this pool of cccDNA is not modified by APOBECs. In the absence of mutated RNA, we speculate that cccDNA derived from deaminated rcDNA may be repaired or degraded. Because such cccDNA is relatively rare, we did not pursue sequencing it. The lack of evidence implying editing of nuclear viral DNA in our studies is consistent with findings by Seeger et al that reported cccDNA editing occurred at extremely low frequencies in hepatoma cells (15).

233 To complete our examination of deamination as a function of the stage of the viral lifecycle, we tested
234 intracellular cores for deaminated DNA. We isolated total DNA from HepAD38 cells and sequenced
235 regions I and II. As with the secreted cores, about 10% of the sequenced clones were mutated. In
236 mutated genomes, the frequency of C to T mutations in region I was more than twice that in region II-
237 6.8% versus 2.3% (Table 3, Figure 7). This implies that a major fraction of this mutated DNA was
238 generated during or after reverse transcription.

239

240 **Discussion**

241 The results presented in this paper (and in the literature) provide constraints that allow us to relate HBV
242 reverse transcription, cytidine deamination, and HBV core structure. In this work we observed that
243 cytidine deamination predominantly occurs on the part of the minus strand that is template to the last
244 part of the plus strand to be transcribed; very few mutations are found in the template to the first half
245 of the plus strand (Fig 5). HepAD38 and HepG2.2.15 cells contain large fraction of intracellular cores with
246 single stranded DNA (44, 45). If this ssDNA were to be deaminated by an encapsidated deaminase, one
247 would expect the entire length of genome to be deaminated at similar frequencies. We did not observe
248 C to T mutations attributable to attack on the RNA or plus strand. Why would the ssRNA and first half
249 the minus strand be protected from cytidine deamination? APOBEC proteins prefer a disordered
250 substrate (39), which presumably can conform to their active site. We, and others, previously observed
251 that a genome-length ssRNA or ssDNA can be entirely bound to the capsid interior (3, 46, 47) resulting in
252 a protein-nucleic acid complex that is likely to be protected from enzymatic modification.

253 We suggest that the effect of synthesizing the second DNA strand inside the core, making dsDNA,
254 explains the sensitivity of the second half of the minus strand to modification. dsDNA is a stiff polymer
255 compared to ssDNA or ssRNA and has a persistence length of 50nm compared ~0.7 nm for ssDNA.

256 dsDNA viruses and phages are believed to prefer a coiled organization of their DNA (48, 49) that
257 minimize the energetic costs of bending and electrostatic repulsion between adjacent turns (50, 51). A
258 spool-like organization of DNA would not be visible in an image reconstruction that had been subjected
259 to icosahedral averaging. In HBV, the internal diameter of the core is only ~25 nm (3) and the stress of a
260 full-length dsDNA genome was predicted to be sufficient to destabilize the capsid (52), a prediction
261 borne out by studies of isolated cores (53). We calculated the amount of DNA that would coat the
262 interior surface of the capsid to be 1400 to 1800 base pairs (Figures 8, 9, Methods), considering a range
263 for capsid inner radius of 12.5 to 13.2 nm (3) and spacing between dsDNA strands of 2.6 to 2.9 nm (52).
264 Thus, we propose that as new dsDNA is synthesized and laid down on the interior surface of the capsid it
265 displaces ssDNA template from the interior capsid surface making it susceptible to a cytidine deaminase
266 (Figure 8). This organization leaves the first half of the genome protected by interaction with the capsid
267 interior and then by base pairing. However, it leaves the second half of the genome largely exposed to
268 cytidine deamination. This is consistent with the peak in the frequency of mutations in the latter half of
269 the genome that includes the ssDNA segment. Interestingly, the number of mutations tapers off towards
270 the end (1100-1630 nt) of the genome (Figure 5) implying that not all of the 5' end of region I is
271 accessible to the deaminase; this single stranded region may be partially protected because it is
272 connected to the dsDNA that circularizes the HBV rcDNA genome.

273 Overall frequencies of mutations of HBV genomes were not remarkably high as they were localized to a
274 fraction of the clones: 10% in HepAD38 and 25% in Hep2.2.15 viral supernatant (Tables 1, 2). That is,
275 most genomes are not mutated and those that are mutated have so many defects they would not
276 produce a viable infection; the observed deamination products would not expand genetic diversity, they
277 are dead ends. This is consistent with some observations of deamination from patients that were
278 cirrhotic (33, 34). An important point to be made is that we observed that cytidine deamination is
279 common in cell lines that are not subjected to an inflammatory response. It is not known whether

unmutated virions did not package a deaminase or were resistant. Resistance to APOBEC has been observed in MuMLV and HTLV, where despite APOBEC packaging there is no deamination (54-57). In either case, deamination of rcDNA at the observed frequencies would be expected to have little effect on a chronic infection. This could explain why HBV has no Vif-like protein to counteract the deaminases. Although, it should be noted that X protein was recently shown to downregulate APOBEC3G when co-expressed, while being ineffective against other deaminases (58).

There is a body of literature on APOBEC interaction with HBV. Based on the work of Nguyen and Hu (27), we used A3G overexpression as a control and recapitulated their results. As per earlier results, we found A3G associated with capsids and suppressed virus DNA levels by 50% but did not change in the cytidine deamination pattern (Figures 3, 4) (27). This decrease in the HBV DNA levels could result in apparent lack in the increase of mutation frequency, however, it is notable that we also did not see any decrease in the proportion of mutant genomes. A3G appears to be over-expressed in liver biopsies from HBV-positive cirrhotic patients, where cytidine deamination was observed but could have resulted from other up-regulated cytidine deaminases (33). In a recent study it was found that APOBEC3B also suppresses production of HBV and that knocking down A3B expression decreased DNA deamination (59). Over-expression of A3C led to increased cytidine deamination *(33, 60)*, though interferon treatment did not lead to A3C overexpression and DNA damage (14, 33). To our disappointment, we found that A3A, though it associated with capsids, did not significantly affect HBV expression or deamination (Figures 3, 4). It is of course possible that a relatively small fraction of cores packaged overexpressed APOBECs.

The approach we designed for evaluating deamination, dividing the genome into two replicons which were sequenced in their entirety, is intended to provide context for mutations and minimize bias for their presence or absence. We observed that mutations are clustered on the minus strand of genomes, specifically on the part of the minus strand that corresponds with the last part of the plus strand to be synthesized. A limitation of this approach is that it is not particularly sensitive to rare mutations and is

304 either labor intensive or provides limited population statistics. Deep sequencing of HBV genomes
305 showed a very similar organization of mutations and provides population statistics (34). However,
306 because pyrosequencing is based on short reads, it was not evident if many genomes had a few
307 mutations or a subset had many mutations. 3D-PCR is usually limited to a relatively short segment of
308 nucleic acid, a couple hundred nucleotides, but is extremely sensitive to relatively rare mutations. 3D-
309 PCR has been used extensively to detect cytidine deamination in HBV (14, 33, 59). 3D-PCR has identified
310 evidence of plus strand deamination in HBV that we did not observe in our HepAD38 or HepG2.2.15 cells
311 HBV (29, 59). A concern with 3D-PCR is that by enriching rare sequences it may overemphasize their
312 presence. While no single technique will yield a complete picture, it is clear that cytidine deamination in
313 HBV is not rare.

314 Though virus produced from cell lines HepAD38 and HepG2.2.15 are affected by an endogenous cytidine
315 deaminase, we were unable to identify a specific candidate suggesting that it is not present in high copy
316 number in all particles. Seeger et al and others also reported detecting mutations in viral sequences
317 from these stable cell lines (15). Based on the analysis of the trinucleotide context of the targeted sites
318 (Figure 6) and sequence preference of deaminases, we predict more than one or all of the following
319 deaminases to be involved- A3G, A1, A3A, A3F and A3H. A3B has been proposed as the likely APOBEC
320 based on the effect of knockdowns (59). It has also been observed that a number of unusual cytidine
321 deaminases have been observed to be up-regulated in liver (33). It is noteworthy that our cell lines were
322 not subjected to interferon treatment or activation of the lymphotoxin-beta receptor reported to
323 activate APOBEC3-induced deamination of HBV (14). While several studies have associated both
324 interferon and lymphotoxin-beta mediated upregulation of APOBECs to viral restriction (14, 31, 61, 62),
325 there are also reports that did not find any correlation (15, 63, 64).

326 In summary, HBV mutants screened from total cellular DNA were largely generated during reverse
327 transcription. The cellular DNA pool may also contain deaminated products from integrated HBV

genomes or cccDNA, however we found that an intact pgRNA was encapsidated, implicating a packaged cytidine deaminase and suggesting that any deaminations in the DNA intermediate upstream were likely to be repaired by cellular enzymes. Although, APOBEC family of proteins are important restriction factors against retroviruses and other DNA viruses, our findings reveal they are less effective against HBV.

Materials and Methods

Plasmids: Generation of A3A and mutants. A3A was PCR amplified from pET21a clone (kind gift from Dr. Judith Levin, NIH) using specific primers containing C-terminal HA epitope tag and cloned in pCDNA3.1(+) vector. The mutants described in text were generated by PCR based mutagenesis using overlapping primers and Pfu Turbo DNA polymerase (Agilent). Plasmid expressing A3G (pRR622) was a kind gift from Dr. Alan Rein, NCI-Frederick.

Cell Culture and transfections: HepAD38 cells were maintained on collagen coated flasks in RPMI (Gibco) media supplemented with 10% FBS (Gibco), 10 mM HEPES buffer pH 7.4 (Gibco), 1x Non-Essential amino acids (Gibco) and 1x Antibiotic antimycotic (Sigma), 400 µg/ml of G418 (Clontech), and 3 ng/ml Doxycycline (Sigma). Virus production was induced by removing Doxycycline repression. Hep2.2.15 cells were maintained in DMEM High glucose media (Hyclone) supplemented with 10% FBS (Gibco), 1x Antibiotic antimycotic (Sigma), 250 µg/ml of G418 (Clontech) for routine passages and maintained in DMEM High glucose media supplemented with 2% FBS, 1x of Antibiotic antimycotic for viral production. Transfection were performed using TRANSIT-LT1 (MIRUSBio) transfection reagent as per the manufacturer's protocol using 5 µgs of APOBEC3 plasmids.

Immunoblotting: Virus-containing culture supernatants were collected and passed through 0.22 µm filter to remove cell debris. The virus supernatant was pelleted through a 20% sucrose cushion at 32,000 rpm (SW40Ti rotor) for 6 hours and the pellet was resuspended in 1x PBS. To obtain HBV cores from

cytoplasmic extract, the cells were washed and removed using a cell scraper and resuspended in 1xPBS. The samples were subjected to SDS-PAGE using a 4-15% Tris-Glycine polyacrylamide gel and transferred to Immobilon-P PVDF membrane. The membrane was probed with 1:1000 diluted Rabbit Anti Core antibodies (Dako) and 1:2000 Goat Anti Rabbit-HRP secondary antibodies for detecting HBV core protein. APOBEC3 proteins were detected with 1:2000 diluted Mouse Anti HA epitope-tag 2.2.14 antibodies (ThermoScientific) and 1:2000 Goat anti mouse-HRP antibodies. The protein bands were detected by chemiluminescence on Bio-RAD ChemiDoc XRS+ imaging system.

Screening for C to T hypermutations. Virus-containing culture supernatants were harvested and passed through 0.22 µm filter to remove cell debris. The virus supernatant was pelleted through a 20% sucrose cushion at 32,000 rpm (SW40Ti rotor) for 6 hours and the pellet was resuspended in 1x PBS. The viral resuspension was treated with Turbo DNAse (Ambion) to remove any non-core associated DNA. DNA was extracted from viral pellet using the Nucleospin "Genomic DNA from blood" kit (Macherey-Nagel). Total RNA was extracted from transfected cells using Trizol reagent (Invitrogen). The cDNA was generated using one step RT cDNA kit (NEB). Viral DNA and cDNA (generated from total RNA) were used as templates to amplify HBV regions corresponding to 1-1630 and 1631-3182 nt using Taq DNA polymerase (ThermoScientific). The PCR products were cloned in pCR4 TOPO vector using TOPO TA cloning kit (Invitrogen) and analyzed by Sanger sequencing using BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). As an additional control, we also amplified and cloned the A3A coding region from cell lysates of HepAD38 transfected with the A3A plasmid.

Calculations: To determine the number of base pairs of dsDNA bound to the inner surface of the capsid we assume B-form DNA will be wrapped in a helix and that the interior of the capsid can be approximated by a smooth sphere. The adjustable parameters then become the average inner radius of the capsid (R_{max}), the spacing between adjacent turns of DNA (ds), and the minimum radius the DNA will

adopt (Figure 9). The maximum radius of the capsid is estimated at 12.5 to 13.2 nm. The smaller value is based on the radius of the nucleic acid shell in RNA-filled capsids (3, 4) and the larger value is from examination of empty capsids although they lack the last 34 residues of the core protein, 2G33 and 1QGT (65, 66). Values of ds range from 2.6 to 2.9 nm. The smaller value is typical for dsDNA in bacteriophages (48, 50). The larger value is from a calculation of DNA packaging in HBV that took into account the small diameter of the capsid (compared to most phages and relative to the 50 nm persistence length of DNA) and the positive surface charge (52). In that HBV calculation, the inner diameter of the DNA for a 3200 bp HBV genome was 4.05 nm, which we shall use as a soft limit in our calculations for this paper.

The radius for a turn is a function of its distance from the equator of the capsid perpendicular to the DNA spool. At the equator, the radius is R_{max}. Distal to the equator, the radius is function of R_{max}, the number of turns (n), and the angle (α) described by the isosceles triangle between two adjacent turns (Figure 9):

1) $R_n = R_{max} \cos(n\ \alpha)$

Calculations for the number of nucleotides in the first layer of DNA adsorbed to the capsid wall are presented in Table 4.

Acknowledgements

This work was funded by a grant from NIH-NIAID R01-AI077688 to AZ and a supported research grant from Assembly BioSciences. AZ has an interest in Assembly Biosciences.

394 **References**

395 1. **Seeger C, Mason WS.** 2015. Molecular biology of hepatitis B virus infection. Virology **479-**
396 **480:**672-686.

397 2. **Tong S, Li J.** 2014. Identification of NTCP as an HBV receptor: the beginning of the end or the
398 end of the beginning? Gastroenterology **146:**902-905.

399 3. **Wang JC, Dhason MS, Zlotnick A.** 2012. Structural organization of pregenomic RNA and the
400 carboxy-terminal domain of the capsid protein of hepatitis B virus. PLoS pathogens **8:**e1002919.

401 4. **Wang JC, Nickens DG, Lentz TB, Loeb DD, Zlotnick A.** 2014. Encapsidated hepatitis B virus
402 reverse transcriptase is poised on an ordered RNA lattice. Proc Natl Acad Sci U S A **111:**11329-
403 11334.

404 5. **Nassal M.** 2008. Hepatitis B viruses: reverse transcription a different way. Virus Res **134:**235-
405 249.

406 6. **Ning X, Nguyen D, Mentzer L, Adams C, Lee H, Ashley R, Hafenstein S, Hu J.** 2011. Secretion of
407 genome-free hepatitis B virus--single strand blocking model for virion morphogenesis of para-
408 retrovirus. PLoS pathogens **7:**e1002255.

409 7. **Jordheim LP, Durantel D, Zoulim F, Dumontet C.** 2013. Advances in the development of
410 nucleoside and nucleotide analogues for cancer and viral diseases. Nat Rev Drug Discov **12:**447-
411 464.

412 8. **Scaglione SJ, Lok AS.** 2012. Effectiveness of hepatitis B treatment in clinical practice.
413 Gastroenterology **142:**1360-1368 e1361.

414 9. **Lampertico P, Liaw YF.** 2012. New perspectives in the therapy of chronic hepatitis B. Gut **61**
415 **Suppl 1:**i18-24.

416 10. **Buti M.** 2014. HBeAg-positive chronic hepatitis B: why do i treat my patients with Nucleos(t)ide
417 analogs? Liver Int **34 Suppl 1:**108-111.

11. **Kao JH.** 2014. HBeAg-positive chronic hepatitis B: why do I treat my patients with pegylated interferon? Liver Int **34 Suppl 1:**112-119.

12. **Zoulim F, Locarnini S.** 2009. Hepatitis B virus resistance to nucleos(t)ide analogues. Gastroenterology **137:**1593-1608 e1591-1592.

13. **Gish R, Given BD, Lai C-L, Locarnini S, Lau JYN, Lewis DL, Schluep T.** 2015. Chronic hepatitis B: natural history, virology, current management and a glimpse at future opportunities. Antiviral research **in press**.

14. **Lucifora J, Xia Y, Reisinger F, Zhang K, Stadler D, Cheng X, Sprinzl MF, Koppensteiner H, Makowska Z, Volz T, Remouchamps C, Chou WM, Thasler WE, Huser N, Durantel D, Liang TJ, Munk C, Heim MH, Browning JL, Dejardin E, Dandri M, Schindler M, Heikenwalder M, Protzer U.** 2014. Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA. Science **343:**1221-1228.

15. **Seeger C, Sohn JA.** 2016. Complete Spectrum of CRISPR/Cas9-induced Mutations on HBV cccDNA. Mol Ther **24:**1258-1266.

16. **Sheehy AM, Gaddis NC, Choi JD, Malim MH.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418:**646-650.

17. **LaRue RS, Andresdottir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jonsson SR, Landau NR, Lochelt M, Malik HS, Malim MH, Munk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu XF, Yuhki N, Harris RS.** 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. J Virol **83:**494-497.

18. **Betts L, Xiang S, Short SA, Wolfenden R, Carter CW, Jr.** 1994. Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. J Mol Biol **235:**635-656.

19. **Navarro F, Bollman B, Chen H, Konig R, Yu Q, Chiles K, Landau NR.** 2005. Complementary function of the two catalytic domains of APOBEC3G. Virology **333:**374-386.

20. **Hakata Y, Landau NR.** 2006. Reversed functional organization of mouse and human APOBEC3 cytidine deaminase domains. J Biol Chem **281:**36624-36631.

21. **Lackey L, Law EK, Brown WL, Harris RS.** 2013. Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. Cell Cycle **12:**762-772.

22. **Blanc V, Kennedy S, Davidson NO.** 2003. A novel nuclear localization signal in the auxiliary domain of apobec-1 complementation factor regulates nucleocytoplasmic import and shuttling. J Biol Chem **278:**41198-41204.

23. **Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, Baysal BE.** 2015. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun **6:**6881.

24. **He X, Li J, Wu J, Zhang M, Gao P.** 2015. Associations between activation-induced cytidine deaminase/apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like cytidine deaminase expression, hepatitis B virus (HBV) replication and HBV-associated liver disease (Review). Mol Med Rep **12:**6405-6414.

25. **Turelli P, Mangeat B, Jost S, Vianin S, Trono D.** 2004. Inhibition of hepatitis B virus replication by APOBEC3G. Science **303:**1829.

26. **Seppen J.** 2004. Unedited inhibition of HBV replication by APOBEC3G. J Hepatol **41:**1068-1069.

27. **Nguyen DH, Gummuluru S, Hu J.** 2007. Deamination-independent inhibition of hepatitis B virus reverse transcription by APOBEC3G. J Virol **81:**4465-4472.

28. **Mao R, Zhang J, Jiang D, Cai D, Levy JM, Cuconati A, Block TM, Guo JT, Guo H.** 2011. Indoleamine 2,3-dioxygenase mediates the antiviral effect of gamma interferon against hepatitis B virus in human hepatocyte-derived cells. J Virol **85:**1048-1057.

29. **Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP.** 2005. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc Natl Acad Sci U S A **102:**8321-8326.

30. **Noguchi C, Ishino H, Tsuge M, Fujimoto Y, Imamura M, Takahashi S, Chayama K.** 2005. G to A hypermutation of hepatitis B virus. Hepatology **41:**626-633.

31. **Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, Candinas D, Sommer P, Wain-Hobson S, Vartanian JP, Greeve J.** 2006. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. Hepatology **43:**1364-1374.

32. **Rosler C, Kock J, Kann M, Malim MH, Blum HE, Baumert TF, von Weizsacker F.** 2005. APOBEC-mediated interference with hepadnavirus production. Hepatology **42:**301-309.

33. **Vartanian JP, Henry M, Marchio A, Suspene R, Aynaud MM, Guetard D, Cervantes-Gonzalez M, Battiston C, Mazzaferro V, Pineau P, Dejean A, Wain-Hobson S.** 2010. Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. PLoS Pathog **6:**e1000928.

34. **Beggel B, Munk C, Daumer M, Hauck K, Haussinger D, Lengauer T, Erhardt A.** 2013. Full genome ultra-deep pyrosequencing associates G-to-A hypermutation of the hepatitis B virus genome with the natural progression of hepatitis B. J Viral Hepat **20:**882-889.

35. **Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, Leonard B, Shandilya SM, Bohn MF, Schiffer CA, Brown WL, Harris RS.** 2012. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J Biol Chem **287:**34801-34808.

36. **Nguyen DH, Hu J.** 2008. Reverse transcriptase- and RNA packaging signal-dependent incorporation of APOBEC3G into hepatitis B virus nucleocapsids. J Virol **82:**6852-6861.

37. **Byeon IJ, Ahn J, Mitra M, Byeon CH, Hercik K, Hritz J, Charlton LM, Levin JG, Gronenborn AM.** 2013. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. Nat Commun **4:**1890.

38. **Mitra M, Hercik K, Byeon IJ, Ahn J, Hill S, Hinchee-Rodriguez K, Singer D, Byeon CH, Charlton LM, Nam G, Heidecker G, Gronenborn AM, Levin JG.** 2014. Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. Nucleic Acids Res **42:**1095-1110.

39. **Holtz CM, Sadler HA, Mansky LM.** 2013. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. Nucleic Acids Res **41:**6139-6148.

40. **Armitage AE, Katzourakis A, de Oliveira T, Welch JJ, Belshaw R, Bishop KN, Kramer B, McMichael AJ, Rambaut A, Iversen AK.** 2008. Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. J Virol **82:**8743-8761.

41. **Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS.** 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol **337:**585-596.

42. **Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS.** 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol **85:**11220-11234.

43. **Liddament MT, Brown WL, Schumacher AJ, Harris RS.** 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr Biol **14:**1385-1391.

44. **Ladner SK, Otto MJ, Barker CS, Zaifert K, Wang GH, Guo JT, Seeger C, King RW.** 1997. Inducible expression of human hepatitis B virus (HBV) in stably transfected hepatoblastoma cells: a novel

system for screening potential inhibitors of HBV replication. Antimicrob Agents Chemother **41:**1715-1720.

45. **Sells MA, Chen ML, Acs G.** 1987. Production of hepatitis B virus particles in Hep G2 cells transfected with cloned hepatitis B virus DNA. Proc Natl Acad Sci U S A **84:**1005-1009.

46. **Zlotnick A, Cheng N, Stahl SJ, Conway JF, Steven AC, Wingfield PT.** 1997. Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: implications for morphogenesis and organization of encapsidated RNA. ProcNatlAcadSciUSA **94:**9556-9561.

47. **Newman M, Chua PK, Tang FM, Su PY, Shih C.** 2009. Testing an electrostatic interaction hypothesis of hepatitis B virus capsid stability by using an in vitro capsid disassembly/reassembly system. J Virol **83:**10616-10626.

48. **Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC.** 1997. Encapsidated conformation of bacteriophage T7 DNA. Cell **91:**271-280.

49. **Purohit PK, Kondev J, Phillips R.** 2003. Mechanics of DNA packaging in viruses. Proc Natl Acad Sci U S A **100:**3173-3178.

50. **Tzlil S, Kindt JT, Gelbart WM, Ben-Shaul A.** 2003. Forces and pressures in DNA packaging and release from viral capsids. Biophys J **84:**1616-1627.

51. **Ben-Shaul A.** 2013. Entropy, energy, and bending of DNA in viral capsids. Biophys J **104:**L15-17.

52. **Dhason MS, Wang JC, Hagan MF, Zlotnick A.** 2012. Differential assembly of Hepatitis B Virus core protein on single- and double-stranded nucleic acid suggest the dsDNA-filled core is spring-loaded. Virology **430:**20-29.

53. **Cui X, Ludgate L, Ning X, Hu J.** 2013. Maturation-associated destabilization of hepatitis B virus nucleocapsid. J Virol **87:**11494-11503.

54. **Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, Munk C, Nymark-McMahon H, Landau NR.** 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell **114:**21-31.

55. **Rulli SJ, Jr., Mirro J, Hill SA, Lloyd P, Gorelick RJ, Coffin JM, Derse D, Rein A.** 2008. Interactions of murine APOBEC3 and human APOBEC3G with murine leukemia viruses. J Virol **82:**6566-6575.

56. **Sasada A, Takaori-Kondo A, Shirakawa K, Kobayashi M, Abudu A, Hishizawa M, Imada K, Tanaka Y, Uchiyama T.** 2005. APOBEC3G targets human T-cell leukemia virus type 1. Retrovirology **2:**32.

57. **Nair S, Sanchez-Martinez S, Ji X, Rein A.** 2014. Biochemical and biological studies of mouse APOBEC3. J Virol **88:**3850-3860.

58. **Chen R, Zhao X, Wang Y, Xie Y, Liu J.** 2017. Hepatitis B virus X protein is capable of down-regulating protein level of host antiviral protein APOBEC3G. Sci Rep **7:**40783.

59. **Chen Y, Hu J, Cai X, Huang Y, Zhou X, Tu Z, Hu J, Tavis JE, Tang N, Huang A, Hu Y.** 2018. APOBEC3B edits HBV DNA and inhibits HBV replication during reverse transcription. Antiviral Res **149:**16-25.

60. **Baumert TF, Rosler C, Malim MH, von Weizsacker F.** 2007. Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. Hepatology **46:**682-689.

61. **Komohara Y, Yano H, Shichijo S, Shimotohno K, Itoh K, Yamada A.** 2006. High expression of APOBEC3G in patients infected with hepatitis C virus. J Mol Histol **37:**327-332.

62. **Sarkis PT, Ying S, Xu R, Yu XF.** 2006. STAT1-independent cell type-specific regulation of antiviral APOBEC3G by IFN-alpha. J Immunol **177:**4530-4540.

63. **Jost S, Turelli P, Mangeat B, Protzer U, Trono D.** 2007. Induction of antiviral cytidine deaminases does not explain the inhibition of hepatitis B virus replication by interferons. J Virol **81:**10588-10596.

64. **Meier MA, Suslov A, Ketterer S, Heim MH, Wieland SF.** 2017. Hepatitis B virus covalently closed circular DNA homeostasis is independent of the lymphotoxin pathway during chronic HBV infection. J Viral Hepat **24:**662-671.

65. **Bourne C, Finn MG, Zlotnick A.** 2006. Global structural changes in hepatitis B capsids induced by the assembly effector HAP1. J Virol **80:**11055-11061.

66. **Wynne SA, Crowther RA, Leslie AGW.** 1999. The crystal structure of the human hepatitis B virus capsid. Mol Cell **3:**771-780.

67. **Abraham TM, Loeb DD.** 2006. Base pairing between the 5' half of epsilon and a cis-acting sequence, phi, makes a contribution to the synthesis of minus-strand DNA for human hepatitis B virus. J Virol **80:**4380-4387.

68. **Oropeza CE, McLachlan A.** 2007. Complementarity between epsilon and phi sequences in pregenomic RNA influences hepatitis B virus replication efficiency. Virology **359:**371-381.

69. **Petrov AS, Boz MB, Harvey SC.** 2007. The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. Journal of structural biology **160:**241-248.

70. **Petrov AS, Harvey SC.** 2008. Packaging double-helical DNA into viral capsids: structures, forces, and energetics. Biophysical journal **95:**497-502.

71. **Petrov AS, Locker CR, Harvey SC.** 2009. Characterization of DNA conformation inside bacterial viruses. Phys Rev E Stat Nonlin Soft Matter Phys **80:**021914.

72. **Rau DC, Lee B, Parsegian VA.** 1984. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. ProcNatlAcadSciUSA **81:**2621-2625.

73. **Rau DC, Parsegian VA.** 1992. Direct measurement of the intermolecular forces between counterion condensed dna double helices. BiophysJ **61:**246-259.

579

580

581 **Figure 1. Encapsidation of APOBEC3s in HBV.** HepAD38 cells (tet-) were transfected with plasmids
582 carrying HA-tagged A3A, HA-tagged A3G or an empty vector. Culture supernatant was harvested 4 days
583 post transfection and was pelleted through a 20% sucrose cushion at 32,000 rpm (SW40Ti rotor) for 6
584 hours. The pellet was resuspended in 1x PBS and analyzed by western blot probing for capsid protein
585 using polyclonal anti-CP (Dako) and for APOBECs using anti-HA tag monoclonal antibodies. A3A and A3G
586 co-pelleted with secreted HBV core protein. Also, the A3A active site point mutant A3A-E72A and the
587 zinc finger disrupting A3A-C101A/C106A mutant co-pelleted with core protein.

588

589 **Figure 2. Schematic diagram of the HBV rcDNA genome**. The black solid marker denotes the
590 polymerase, covalently associated with the 5' end of the minus strand; polymerase primes minus strand
591 synthesis, catalyzes reverse transcription, and digests the template pgRNA. The plus strand (inner ring) is
592 incomplete (dashed line and gap), leaving some of the minus strand single stranded. In the standard
593 numbering convention, residue 1 denotes a unique *Eco*RI site. The primers used for PCR amplification
594 correspond to region I (1-1630, right side of the map) and region II (1631-3182, left side of the map).

595

596 **Figure 3 C to T mutations in viral genomes from HepAD38 culture supernatant.** HepAD38 cells were
597 transfected with A3A and A3G expressing plasmids. The viral DNA from culture supernatant was
598 extracted and regions 1-1630 and 1631-3182 were PCR amplified, TOPO cloned and sequenced. Table 1
599 shows the total number of clones sequenced, the number of mutant clones carrying C to T mutations,
600 the total number of C residues in all the sequenced clones, the number of C to T mutations in mutant
601 clones and the frequencies of mutations from genomes secreted into cell media. Mutation frequency
602 was calculated as ratio of the number of C to T mutations and total number of C residues in mutant

603 clones; Overall C to T Mutation frequency is ratio of number of C to T mutations and total number of C
604 residues in all sequenced clones. A histogram compares the frequency of C to T mutations (ratio of total
605 C to T mutations to total Cs sequenced) between regions I (1-1630) and II (1631-3182). The average
606 highly mutated genome had mutations at 31.3 ± 16.6 out of 358 cytosines in region I and at 4.6 ± 3.4 of
607 334 cytosines in region II, a statistically significant difference ($p = 0.0014$).

608

609 **Figure 4. C to T mutations in viral genomes from Hep2.2.15 culture supernatant.** These data are
610 derived from viral supernatant produced in Hep2.2.15 cells in the presence or absence of A3A and A3G.
611 Table 2 shows the number of mutant clones and C to T mutations from genomes secreted into cell
612 media. The histogram compares the frequency of C to T mutation between regions I (1-1630) and II
613 (1631-3182). A larger fraction of clones had mutations in HepG2.2.15 cells than in HepAD38 cells (Figure
614 3). As with HepAD38 cells, the majority of mutations were in region I. The average highly mutated
615 genome had mutations at 35.8 ± 15.6 out of 358 cytosines in region I and at 4.1 ± 2.7 of 334 cytosines in
616 region II, a statistically significant difference ($p < 0.0001$).

617

618 **Figure 5. Genome-wide Mutation Profile.** All C to T mutations identified in genomes from culture
619 supernatants of **(a)** HepAD38 and **(b)** HepG2.2.15 cells are mapped to the entire length of the HBV
620 genome. The histogram is broken into 18 bins of ca.180 nucleotides each. Each bar (red) denotes the
621 average number of mutations in the corresponding region. Each dot is a single highly mutated clone. The
622 data are compiled from 18 HepAD38 clones (403 mutations) and 33 HepG2.2.15 clones (547 mutations).
623 For HepAD38, bins in regions 1-1630 show for 12 clones and regions 1631-3182 shows mutations in 6
624 clones. For HepG2.2.15 bins for regions 1-1630 shows mutations in 13 clones and regions 1631-3182
625 show mutations for 20 clones. The order of the bins corresponds to the plus strand in 5' to 3' direction

using the standard numbering scheme. For both cell lines, more than 70% of clones showed no modifications.

Figure 6. Trinucleotide context of the deaminated cytosine. Sequences of all deaminated sites from region I and region II were analyzed and the 5' to 3' trinucleotide context of the target cytosine determined. The first two bases of the trinucleotide are listed below the histogram, e.g. CC for CCX and the central base is the target cytosine.

Figure 7. C to T mutations in total viral DNA from HepAD38 cells. Viral DNA isolated from HepAD38 cell lysates shows a similar pattern of mutation observed in viral genomes from culture supernatant. (Figure 3). Lysates were isolated from cells that were transfected with an empty vector. Table 3 shows the number of mutant clones and C to T mutations identified in screening cell lysate. The histogram compares the frequency of C to T mutation between regions I (1-1630) and II (1631-3182).

Figure 8. The pattern of cytidine deamination constrains a model of reverse transcription. (**a)** A capsid immediately after initiating second strand (plus strand) synthesis. The ssDNA minus strand (gray) is adsorbed to the basic interior surface of the capsid. The RNA primer and first part of the second strand (red) connect the 5' and 3' ends of the minus strand to circularize the genome. Complementary sequences near the 5' and 3' ends of the minus strand facilitate the transfer of the polymerase to the 3' end of the minus strand needed to circularize the nascent rcDNA (67, 68). (**b)** A capsid where synthesis of the plus strand is about 30% complete. We predict that newly synthesized dsDNA (red and gray) will be arranged in coils that are adsorbed to the basic inner surface of the capsid. At this stage of synthesis,

much of the remaining minus strand is still adsorbed to the capsid inner surface. Double stranded and capsid adsorbed regions of the minus strand are protected from cytidine deamination (see Figure 5). The growing dsDNA will displace a proportional amount of the minus strand. Topologically, for dsDNA synthesis to continue unimpeded, the displaced ssDNA will be in the lumen of the spool, rendering it sensitive to cytidine deamination. From an energetic perspective, it is likely that initial segments of dsDNA are localized to the equator of the capsid to minimize bending energy (69-71). For the same reason, it is likely that the dsDNA spool will remain centered in the capsid and move "down" as new dsDNA is synthesized and added to the "top" of the spool. **(c)** A mature core. The capsid is lined with two to three layers of dsDNA (52). The outermost layer has about 1400 to 1800 bp, depending on spacing between DNA strands, how closely the DNA gets to the inner surface of the capsid, and how many turns of DNA can be fit before the DNA starts a second layer.

Figure 9. Calculating the amount of dsDNA adsorbed to the interior surface of an HBV capsid. This is a stylized cross-section of a capsid showing five turns of DNA. The interior of the capsid is treated as a spherical surface of radius R_{max}. The small circles represent DNA cross sections of diameter d_s; d_s includes hydration and counterions making it larger than the molecular diameter of DNA (72, 73). To simplify the calculations, we treat this model as a series of stacked concentric circles. To calculate the radius of given layer of dsDNA, the angle α is based on the distance to the DNA layer, $R_{max} - d_s/2$, and the distance between adjacent strands of dsDNA, d_s. The smallest radius for a layer of DNA is R_{min}; note that the radius for dsDNA in that layer is $R_{min} + 2_{ds}$.

1

Table 1. C to T MUTATIONS IN VIRAL GENOMES FROM HepAD38 CULTURE SUPERNATANT							
Region	**Expression conditions**	**Clones sequenced**	**Highly mutated genomes**	**Total Cs**	**Total C → T mutations**	**C → T in highly mutated genomes (%)**	**Overall C →T Mutation Frequency (%)**
1-1630	Virus alone	60	6	21480	174	8.1	0.8
	Virus+A3A	38	4	13604	122	8.5	0.9
	Virus+A3G	20	2	7160	79	11.0	1.1
1631-3182	Virus alone	12	1	4008	5	1.5	0.1
	Virus+A3A	27	2	9018	5	0.7	0.05
	Virus+A3G	26	3	8684	18	1.8	0.2

2

3

1

Table 2. C to T MUTATIONS IN VIRAL GENOMES FROM Hep2.2.15 CULTURE SUPERNATANT							
Region	**Expression conditions**	**Clones sequenced**	**Highly mutated genomes**	**Total Cs**	**Total C → T mutations**	**C → T in highly mutated genomes (%)**	**Overall C →T Mutation Frequency (%)**
1-1630	Virus alone	20	5	6980	187	10.4	2.6
	Virus+A3A	20	5	7160	166	9.3	2.3
	Virus+A3G	10	3	3580	113	10.5	3.1
1631-3182	Virus alone	20	7	6180	33	1.4	0.5
	Virus+A3A	20	6	6513	26	1.3	0.4
	Virus+A3G	20	7	6350	22	0.9	0.3

2

3

1

2

Table 3. C → T MUTATIONS IN TOTAL DNA from HepAD38 CELLS							
Region	**Expression conditions**	**Clones sequenced**	**Highly mutated genomes**	**Total Cs**	**Total C → T mutations**	**C → T in highly mutated genomes (%)**	**Overall C →T Mutation Frequency (%)**
1-1630	Virus only	30	3	10560	75	6.8	0.7
1631-3182	Virus only	30	4	9853	32	2.3	0.3

3

Table 4. The number of nucleotides in the first layer of DNA bound to the inner surface of the capsid.

Each table shows the calculation of bound nucleotides for one hemisphere. The maximum inner radius (R_{max} from Figure 9) and DNA strand diameter (d_s for Figure 9) are listed in the header for each table. Turn 1 is equatorial in each case here. The value for alpha is based on radius and spacing between DNA strands. The elevation of each turn is thus an integral coefficient (n) of alpha, allowing calculation of the radius and circumference for that layer from which the number of base pairs in the layer are calculated.

Maximum radius = 12.5 nm
DNA radius (d_s) = 2.6 nm
alpha =13.3°

turn	n	radius (nm)	circum (nm)	R_{in} (nm)	bp / layer
1	0*	12.5	70.4	9.9	207.0
2	1	12.2	68.3	9.6	200.8
3	2	11.2	62.0	8.6	182.4
4	3	9.6	52.0	7.0	153.0
5	4	7.5	38.7	4.9	114.0

Bp in partial capsid = 857.0
whole capsid = 1507 bp
number of turns = 9

Maximum radius= 13.2 nm
DNA radius (d_s) = 2.6 nm
alpha = 12.5°

turn	n	radius	circum	R_{in}	bp / layer
1	0*	13.2	74.8	10.6	219.9
2	1	12.9	72.8	10.3	214.1
3	2	12.0	66.9	9.4	196.9
4	3	10.5	57.5	7.9	169.2
5	4	8.5	45.0	5.9	132.2
6	5	6.1	29.8	3.5	87.8

Bp in partial capsid = 932.3
whole capsid = 1820 bp
number of turns = 11

Maximum radius = 12.5 nm
DNA radius (d_s) = 2.9 nm
Alpha = 15.1°

turn	n	radius	circum	R_{in}	bp / layer
1	0*	12.5	70.4	9.9	207.0
2	1	12.1	67.7	9.5	199.0
3	2	10.8	59.7	8.2	175.7
4	3	8.8	47.1	6.2	138.6
5	4	6.2	30.7	3.6	90.4

Bp in partial capsid = 810.7
whole capsid = 1414
number of turns = 9

Maximum radius = 13.2 nm
DNA radius (d_s) = 2.9 nm
Alpha = 14.2°

turn	n	radius	circum	R_{in}	bp / layer
1	0*	13.2	74.8	10.6	219.9
2	1	12.8	72.2	10.2	212.5
3	2	11.6	64.8	9.0	190.6
4	3	9.7	52.9	7.1	155.7
5	4	7.2	37.4	4.6	109.9

Bp in partial capsid = 888.6
whole capsid = 1557
number of turns = 9

* Equatorial turn.