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 a 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.
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.
Introduction

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

Present therapies against chronic infection includes interferon-α (IFN-α) treatment and nucleotide analogs (7-11). Nucleotide analogs keep HBV infection at bay but fail to eliminate the resident cccDNA pool (10, 12, 13). IFN-α treatment can lead to sustained virological response (SVR) but only in a small
subset of patients: it is most efficacious on those who were infected as adults, have had chronic infection for a relatively short time, and have relatively high liver inflammation (13). Understanding its mechanism of action may make important contributions to HBV therapy. Among the widespread array of IFN-α stimulated genes that are activated, the APOBEC3 family of nucleic acid editing enzymes have gained special interest (14, 15).

APOBEC3 (A3) proteins are host restriction factors that can be packaged into the retroviral virions and introduce mutations in the viral genome by deaminating cytosine to uracil, thereby restricting the viral replication (16). The catalytic domain of A3 deaminases is comprised of the hallmark H-X-E active site motif and the Zinc finger motif P-C-X_{12-40}-C (17). The active site glutamate is involved in proton transfer during catalysis and the zinc ion coordinates a catalytic water molecule (18). A3 deaminases can have one (A3A, A3C, A3H) or two such domains (A3B, A3F, A3G). In the latter case, only one domain is active and the other is involved in packaging into virions (19, 20); A3A is the most potent among all deaminase isoforms. Members of the APOBEC deaminase family are either localized to the cytoplasm or shuttle between the nucleus and the cytoplasm during the cell cycle and can effectively target single stranded DNA (ssDNA) substrates in either compartment (21). Though APOBECs have a preference for ssDNA as substrate, some APOBECs like APOBEC1, APOBEC3A can also deaminate RNA (22, 23).

HBV has been reported to be restricted by several members of A3 family of proteins by both deaminase-dependent and -independent mechanisms (24). HBV inhibition by A3G, independent of deaminase activity, was attributed to inhibition of either minus strand synthesis or pgRNA packaging (25-28). Likewise, evidence of HBV genome editing has been reported in infected patient samples and samples from transfected hepatoma cell lines (29-34). It is still unclear which HBV nucleic acid intermediates could be the substrates for APOBEC mediated deamination or which APOBECs catalyze deamination. One report claimed that Interferon- or cytokine-induced activation of nuclear-localized
deaminases A3A and A3B in infected cells could specifically degrade cccDNA while being non-genotoxic (14).

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

Results:

Viral encapsidation of APOBEC3A

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

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

C to T mutations in viral genomes from HepAD38 culture supernatant

To test if the over-expressed A3A led to deaminated HBV nucleic acid inside the nucleocapsid, we analyzed the genomic rcDNA for G to A and C to T mutations. Previously reported mutational analyses have been limited to the region coding for the X protein, so in the present study we sequenced the entire genome. We amplified the genome as two large amplicons – region I (1-1630) and II (1631-3182). During reverse transcription, both regions will transiently be found as a single stranded minus strand. The synthesis of the plus strand will first occlude region II. The plus strand for region I is synthesized later and contains single stranded region owing to incomplete plus strand synthesis (Figure 2). The length of this single stranded region may vary from virus to virus. DNA was extracted from HepAD38 culture supernatant pellet and the two regions were PCR amplified with specific primers, using Taq DNA 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
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).
To complete our examination of deamination as a function of the stage of the viral lifecycle, we tested intracellular cores for deaminated DNA. We isolated total DNA from HepAD38 cells and sequenced regions I and II. As with the secreted cores, about 10% of the sequenced clones were mutated. In mutated genomes, the frequency of C to T mutations in region I was more than twice that in region II—6.8% versus 2.3% (Table 3, Figure 7). This implies that a major fraction of this mutated DNA was generated during or after reverse transcription.

Discussion

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

We suggest that the effect of synthesizing the second DNA strand inside the core, making dsDNA, explains the sensitivity of the second half of the minus strand to modification. dsDNA is a stiff polymer compared to ssDNA or ssRNA and has a persistence length of 50nm compared ~0.7 nm for ssDNA.
dsDNA viruses and phages are believed to prefer a coiled organization of their DNA (48, 49) that minimize the energetic costs of bending and electrostatic repulsion between adjacent turns (50, 51). A spool-like organization of DNA would not be visible in an image reconstruction that had been subjected to icosahedral averaging. In HBV, the internal diameter of the core is only ~25 nm (3) and the stress of a full-length dsDNA genome was predicted to be sufficient to destabilize the capsid (52), a prediction borne out by studies of isolated cores (53). We calculated the amount of DNA that would coat the interior surface of the capsid to be 1400 to 1800 base pairs (Figures 8, 9, Methods), considering a range 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).

Thus, we propose that as new dsDNA is synthesized and laid down on the interior surface of the capsid it displaces ssDNA template from the interior capsid surface making it susceptible to a cytidine deaminase (Figure 8). This organization leaves the first half of the genome protected by interaction with the capsid interior and then by base pairing. However, it leaves the second half of the genome largely exposed to cytidine deamination. This is consistent with the peak in the frequency of mutations in the latter half of the genome that includes the ssDNA segment. Interestingly, the number of mutations tapers off towards the end (1100-1630 nt) of the genome (Figure 5) implying that not all of the 5’ end of region I is accessible to the deaminase; this single stranded region may be partially protected because it is connected to the dsDNA that circularizes the HBV rcDNA genome.

Overall frequencies of mutations of HBV genomes were not remarkably high as they were localized to a fraction of the clones: 10% in HepAD38 and 25% in Hep2.2.15 viral supernatant (Tables 1, 2). That is, most genomes are not mutated and those that are mutated have so many defects they would not produce a viable infection; the observed deamination products would not expand genetic diversity, they are dead ends. This is consistent with some observations of deamination from patients that were cirrhotic (33, 34). An important point to be made is that we observed that cytidine deamination is 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
either labor intensive or provides limited population statistics. Deep sequencing of HBV genomes showed a very similar organization of mutations and provides population statistics (34). However, because pyrosequencing is based on short reads, it was not evident if many genomes had a few mutations or a subset had many mutations. 3D-PCR is usually limited to a relatively short segment of nucleic acid, a couple hundred nucleotides, but is extremely sensitive to relatively rare mutations. 3D-PCR has been used extensively to detect cytidine deamination in HBV (14, 33, 59). 3D-PCR has identified evidence of plus strand deamination in HBV that we did not observe in our HepAD38 or HepG2.2.15 cells HBV (29, 59). A concern with 3D-PCR is that by enriching rare sequences it may overemphasize their presence. While no single technique will yield a complete picture, it is clear that cytidine deamination in HBV is not rare.

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

In summary, HBV mutants screened from total cellular DNA were largely generated during reverse 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 \( \mu \)m 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 (\( R_{\text{max}} \)), the spacing between adjacent turns of DNA (\( d_s \)), and the minimum radius the DNA will...
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 $R_{\text{max}}$. Distal to the equator, the radius is function of $R_{\text{max}}$, the number of turns ($n$), and the angle ($\alpha$) described by the isosceles triangle between two adjacent turns (Figure 9):

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

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.


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

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

Figure 3 C to T mutations in viral genomes from HepAD38 culture supernatant. HepAD38 cells were transfected with A3A and A3G expressing plasmids. The viral DNA from culture supernatant was extracted and regions 1-1630 and 1631-3182 were PCR amplified, TOPO cloned and sequenced. Table 1 shows the total number of clones sequenced, the number of mutant clones carrying C to T mutations, the total number of C residues in all the sequenced clones, the number of C to T mutations in mutant clones and the frequencies of mutations from genomes secreted into cell media. Mutation frequency was calculated as ratio of the number of C to T mutations and total number of C residues in mutant.
clones; Overall C to T Mutation frequency is ratio of number of C to T mutations and total number of C residues in all sequenced clones. A histogram compares the frequency of C to T mutations (ratio of total C to T mutations to total Cs sequenced) between regions I (1-1630) and II (1631-3182). The average highly mutated genome had mutations at 31.3 ± 16.6 out of 358 cytosines in region I and at 4.6 ± 3.4 of 334 cytosines in region II, a statistically significant difference (p = 0.0014).

Figure 4. C to T mutations in viral genomes from Hep2.2.15 culture supernatant. These data are derived from viral supernatant produced in Hep2.2.15 cells in the presence or absence of A3A and A3G. Table 2 shows the number of mutant clones and C to T mutations from genomes secreted into cell media. The histogram compares the frequency of C to T mutation between regions I (1-1630) and II (1631-3182). A larger fraction of clones had mutations in Hep2.2.15 cells than in HepAD38 cells (Figure 3). As with HepAD38 cells, the majority of mutations were in region I. The average highly mutated 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 region II, a statistically significant difference (p < 0.0001).

Figure 5. Genome-wide Mutation Profile. All C to T mutations identified in genomes from culture supernatants of (a) HepAD38 and (b) HepG2.2.15 cells are mapped to the entire length of the HBV genome. The histogram is broken into 18 bins of ca.180 nucleotides each. Each bar (red) denotes the average number of mutations in the corresponding region. Each dot is a single highly mutated clone. The data are compiled from 18 HepAD38 clones (403 mutations) and 33 HepG2.2.15 clones (547 mutations). For HepAD38, bins in regions 1-1630 show for 12 clones and regions 1631-3182 shows mutations in 6 clones. For HepG2.2.15 bins for regions 1-1630 shows mutations in 13 clones and regions 1631-3182 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 $R_{\text{max}}$. 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 $\alpha$ is based on the distance to the DNA layer, $R_{\text{max}} - d_s/2$, and
the distance between adjacent strands of dsDNA, $d_s$. The smallest radius for a layer of DNA is $R_{\text{min}}$ note
that the radius for dsDNA in that layer is $R_{\text{min}} + 2d_s$. 
### Table 1. C to T MUTATIONS IN VIRAL GENOMES FROM HepAD38 CULTURE SUPERNATANT

<table>
<thead>
<tr>
<th>Region</th>
<th>Expression conditions</th>
<th>Clones sequenced</th>
<th>Highly mutated genomes</th>
<th>Total Cs</th>
<th>Total C → T mutations</th>
<th>C → T in highly mutated genomes (%)</th>
<th>Overall C → T Mutation Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1630</td>
<td>Virus alone</td>
<td>60</td>
<td>6</td>
<td>21480</td>
<td>174</td>
<td>8.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Virus+A3A</td>
<td>38</td>
<td>4</td>
<td>13604</td>
<td>122</td>
<td>8.5</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Virus+A3G</td>
<td>20</td>
<td>2</td>
<td>7160</td>
<td>79</td>
<td>11.0</td>
<td>1.1</td>
</tr>
<tr>
<td>1631-3182</td>
<td>Virus alone</td>
<td>12</td>
<td>1</td>
<td>4008</td>
<td>5</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Virus+A3A</td>
<td>27</td>
<td>2</td>
<td>9018</td>
<td>5</td>
<td>0.7</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Virus+A3G</td>
<td>26</td>
<td>3</td>
<td>8684</td>
<td>18</td>
<td>1.8</td>
<td>0.2</td>
</tr>
</tbody>
</table>
### Table 2. C to T MUTATIONS IN VIRAL GENOMES FROM Hep2.2.15 CULTURE SUPERNATANT

<table>
<thead>
<tr>
<th>Region</th>
<th>Expression conditions</th>
<th>Clones sequenced</th>
<th>Highly mutated genomes</th>
<th>Total Cs</th>
<th>Total C → T mutations</th>
<th>C → T in highly mutated genomes (%)</th>
<th>Overall C → T Mutation Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1630</td>
<td>Virus alone</td>
<td>20</td>
<td>5</td>
<td>6980</td>
<td>187</td>
<td>10.4</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>Virus+A3A</td>
<td>20</td>
<td>5</td>
<td>7160</td>
<td>166</td>
<td>9.3</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Virus+A3G</td>
<td>10</td>
<td>3</td>
<td>3580</td>
<td>113</td>
<td>10.5</td>
<td>3.1</td>
</tr>
<tr>
<td>1631-3182</td>
<td>Virus alone</td>
<td>20</td>
<td>7</td>
<td>6180</td>
<td>33</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Virus+A3A</td>
<td>20</td>
<td>6</td>
<td>6513</td>
<td>26</td>
<td>1.3</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Virus+A3G</td>
<td>20</td>
<td>7</td>
<td>6350</td>
<td>22</td>
<td>0.9</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Table 3. C → T MUTATIONS IN TOTAL DNA from HepAD38 CELLS

<table>
<thead>
<tr>
<th>Region</th>
<th>Expression conditions</th>
<th>Clones sequenced</th>
<th>Highly mutated genomes</th>
<th>Total Cs</th>
<th>Total C → T mutations</th>
<th>C → T in highly mutated genomes (%)</th>
<th>Overall C → T Mutation Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1630</td>
<td>Virus only</td>
<td>30</td>
<td>3</td>
<td>10560</td>
<td>75</td>
<td>6.8</td>
<td>0.7</td>
</tr>
<tr>
<td>1631-3182</td>
<td>Virus only</td>
<td>30</td>
<td>4</td>
<td>9853</td>
<td>32</td>
<td>2.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>
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 \( R_{\text{max}} \) 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.

<table>
<thead>
<tr>
<th>Turn</th>
<th>( n )</th>
<th>Radius (nm)</th>
<th>Circum (nm)</th>
<th>( R_{\text{in}} ) (nm)</th>
<th>bp / layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0*</td>
<td>12.5</td>
<td>70.4</td>
<td>9.9</td>
<td>207.0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12.2</td>
<td>68.3</td>
<td>9.6</td>
<td>200.8</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>11.2</td>
<td>62.0</td>
<td>8.6</td>
<td>182.4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>9.6</td>
<td>52.0</td>
<td>7.0</td>
<td>153.0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7.5</td>
<td>38.7</td>
<td>4.9</td>
<td>114.0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Turn</th>
<th>( n )</th>
<th>Radius (nm)</th>
<th>Circum (nm)</th>
<th>( R_{\text{in}} ) (nm)</th>
<th>bp / layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0*</td>
<td>13.2</td>
<td>74.8</td>
<td>10.6</td>
<td>219.9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12.9</td>
<td>72.8</td>
<td>10.3</td>
<td>214.1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>12.0</td>
<td>66.9</td>
<td>9.4</td>
<td>196.9</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>10.5</td>
<td>57.5</td>
<td>7.9</td>
<td>169.2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>8.5</td>
<td>45.0</td>
<td>5.9</td>
<td>132.2</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>6.1</td>
<td>29.8</td>
<td>3.5</td>
<td>87.8</td>
</tr>
</tbody>
</table>

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

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

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

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

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

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

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

* Equatorial turn.