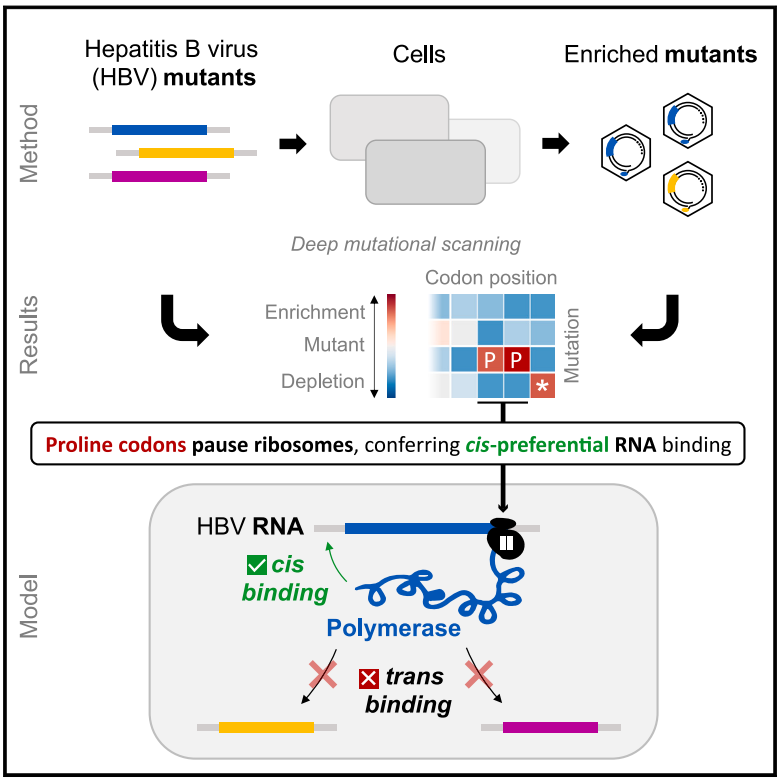
Cell

Article

Deep mutational scanning of hepatitis B virus reveals a mechanism for *cis*-preferential reverse transcription

Graphical abstract



Highlights

- Deep mutational scanning (DMS) provides a high-resolution fitness map of HBV polymerase
- Initiating HBV replication with RNA uncouples *cis*- and *trans*-acting protein functions
- Ribosome pausing tethers HBV Pol to RNA, enforcing *cis*-preferential reverse transcription
- DMS of the HBV genome supports a leaky ribosome scanning model for polymerase translation

Authors

Yingpu Yu, Maximilian A. Kass, Mengyin Zhang, ..., Debora S. Marks, Charles M. Rice, William M. Schneider

Correspondence

ricec@rockefeller.edu (C.M.R.), wschneider@rockefeller.edu (W.M.S.)

In brief

Deep mutational scanning reveals that conserved proline codons at the 3′ end of the hepatitis B virus polymerase open reading frame stall ribosomes. As a result, the ribosome physically tethers the nascent polymerase to its encoding RNA, which enforces *cis*-preferential genome RNA packaging and reverse transcription.

Yu et al., 2024, Cell *187*, 2735–2745
May 23, 2024
https://doi.org/10.1016/j.cell.2024.04.008

CellPress

Cell

CellPress

Article

Deep mutational scanning of hepatitis B virus reveals a mechanism for *cis*-preferential reverse transcription

Yingpu Yu,[1,7] Maximilian A. Kass,[1,2,7] Mengyin Zhang,[1] Noor Youssef,[3,4] Catherine A. Freije,[1] Kelly P. Brock,[3,4,8] Lauren C. Aguado,[1] Leon L. Seifert,[1,5] Sanjana Venkittu,[1] Xupeng Hong,[1] Amir Shlomai,[1,9] Ype P. de Jong,[1,6] Debora S. Marks,[3,4] Charles M. Rice,[1,10,*] and William M. Schneider[1,*]

[1]Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, NY 10065, USA
[2]Department of Infectious Diseases, Molecular Virology, Heidelberg University, Medical Faculty Heidelberg, Heidelberg, Germany
[3]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[4]Department of Organismic and Evolutionary Biology, Broad Institute of MIT and Harvard, Harvard University, Cambridge, MA 02138, USA
[5]Center for Clinical and Translational Science, The Rockefeller University, New York, NY 10065, USA
[6]Division of Gastroenterology and Hepatology, Weill Cornell Medical College, New York, NY 10065, USA
[7]These authors contributed equally
[8]Present address: Kernal Biologics, Cambridge, MA 02142, USA
[9]Present address: Department of Medicine D and the Liver Institute, Belinson Hospital & the Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[10]Lead contact
*Correspondence: ricec@rockefeller.edu (C.M.R.), wschneider@rockefeller.edu (W.M.S.)
https://doi.org/10.1016/j.cell.2024.04.008

SUMMARY

Hepatitis B virus (HBV) is a small double-stranded DNA virus that chronically infects 296 million people. Over half of its compact genome encodes proteins in two overlapping reading frames, and during evolution, multiple selective pressures can act on shared nucleotides. This study combines an RNA-based HBV cell culture system with deep mutational scanning (DMS) to uncouple *cis*- and *trans*-acting sequence requirements in the HBV genome. The results support a leaky ribosome scanning model for polymerase translation, provide a fitness map of the HBV polymerase at single-nucleotide resolution, and identify conserved prolines adjacent to the HBV polymerase termination codon that stall ribosomes. Further experiments indicated that stalled ribosomes tether the nascent polymerase to its template RNA, ensuring *cis*-preferential RNA packaging and reverse transcription of the HBV genome.

INTRODUCTION

Hepatitis B virus (HBV) is a small, enveloped DNA virus that chronically infects 296 million people worldwide. Chronic infection can cause liver cirrhosis and hepatocellular carcinoma, resulting in nearly one million deaths annually.[1] Despite the availability of effective treatments that suppress virus replication, a complete cure is rare.

HBV is transmitted through bodily fluids and infects human hepatocytes. Virions entering cells contain an incomplete form of the genome known as relaxed circular DNA (rcDNA) characterized by a full-length minus-strand DNA covalently bound to the polymerase protein (Pol) and an incomplete plus-strand DNA covalently bound to an RNA primer. Host enzymes convert rcDNA to the stable nuclear episomal form known as covalently closed circular DNA (cccDNA), which is then transcribed by host RNA polymerase II to produce all HBV RNAs. HBV rcDNA is reverse transcribed from one such transcript, the pregenomic RNA (pgRNA), which is also the template for translating the Core and Pol proteins. The persistence of cccDNA in hepatocytes is believed to be a major challenge in achieving a cure.[2,3]

The 3.2 kilobase HBV genome is highly compact, and over half of the nucleotides encode proteins in overlapping reading frames. Consequently, in these overlapping regions, selective pressures that influence protein sequence in one reading frame will affect the protein sequence in the other.[4] This feature constrains HBV evolution and limits its ability to acquire resistance to therapeutics. A better understanding of how HBV manages or possibly benefits from these constraints may provide insights into its life cycle and vulnerabilities to exploit therapeutically.

Here, we combined a recently described RNA-based method for studying HBV replication[5] with deep mutational scanning (DMS) to interrogate selective pressures acting on the HBV genome. This enabled us to assess nearly all single codon variants in the Core and Pol open reading frames (ORFs) in high-throughput pooled assays. The Core protein can function in *trans*

and complement pgRNAs encoding defective Core ORFs.[5,6] By contrast, the Pol protein functions predominantly in *cis*, preferentially reverse transcribing the pgRNA from which it was translated.[7] Exploiting the natural *trans*- and *cis*-acting functions of the HBV Core and Pol proteins enabled us to uncouple selective pressures acting on shared nucleotides in these two reading frames. This approach identified *cis*-acting elements in the Core ORF that regulate Pol translation, and the results support an intricate yet elegant leaky ribosome scanning model whereby the scanning 40S ribosomal subunit bypasses one or more initiation codons before initiating translation at a downstream initiation codon. Further, this approach enabled us to define clear boundaries for the Pol spacer domain, identify two putative zinc-finger motifs in the reverse-transcriptase (RT) domain, and uncover the molecular basis of how HBV Pol preferentially packages and reverse transcribes the pgRNA template from which it was translated (*cis* preference).

RESULTS

HBV pgRNA transfection elucidates *cis*-acting but not *trans*-acting functions

All the proteins encoded by the HBV genome overlap with the Pol protein (Figure 1A). We reasoned that initiating genome replication by transfecting cells with HBV pgRNA might enable us to uncouple selective pressures acting on Core, the surface protein (HBsAg), and the multifunctional HBx protein from selective pressures acting on Pol. To test this, we used a series of HBV mutant pgRNAs with early termination codons in either the Core, HBsAg, or HBx ORFs and quantified HBV DNA by qPCR 2 days post-transfection (Figure 1B). HBsAg and HBx functions are not needed for reverse transcription,[8,9] so as expected, pgRNAs harboring truncated HBsAg and HBx produced wild-type (WT) levels of HBV DNA. By contrast, the Core protein is essential for reverse transcription because this process occurs within capsids comprising 240 Core subunits.[10] Accordingly, HBV pgRNA harboring an early termination codon in the Core ORF failed to produce HBV DNA, as did the catalytically inactive Pol mutant (YMHH) control. These results indicated that initiating HBV genome replication with pgRNA transfection and quantifying the HBV DNA output could enable us to uncouple Pol sequence requirements from that of HBsAg and HBx in regions of overlap.

Although the Core mutant pgRNA described above failed to produce HBV DNA, we hypothesized it might still be possible to uncouple the sequence requirements for Core and Pol in their overlapping region when using a pooled library of pgRNA sequence variants because pgRNAs encoding defective HBV Core proteins can be encapsidated in *trans* by functional Core proteins within the same cell.[5,6] By contrast, because Pol has a *cis* preference for reverse transcribing the pgRNA from which it was translated, non-functional Pol sequences will be depleted from the DNA population.[7] This concept is illustrated in Figure 1C, and if this holds true, transfection will deliver many pgRNA variants per cell, thereby uncoupling the sequence requirements of Core from Pol on any given pgRNA.

To assess *trans* complementation vs. *cis* preference for Core and Pol in our assay, we mixed WT and mutant pgRNAs at defined ratios before transfection and then sequenced the resulting HBV DNA purified from cell lysates 2 days later. With an absolute *cis* preference, we should only recover WT DNA. By contrast, with no *cis* preference (absolute *trans* activity), we should recover DNA with the same ratio of WT and mutant as the input pgRNA mixture. As shown in Figure 1D, even when a high percentage (>90%) of the transfected pgRNA in the cell encodes a catalytically dead Pol (YMHH), nearly all the pgRNAs converted to DNA encoded the WT sequence, indicating that Pol exhibits a strong *cis* preference. By contrast, the percentage of WT and Core early termination mutant DNA sequences recovered was nearly identical to that of input, as expected since Core acts in *trans*. This suggested that it would be possible to uncouple Core sequence requirements from Pol sequence requirements in their overlapping sequence.

DMS in the Core-encoding region reveals *cis* elements that control Pol translation

The ability to uncouple selective pressures on Pol from Core provides a unique opportunity to selectively assess *cis*-acting sequences in the Core region that influence Pol translation. A model proposed for Pol translation suggests leaky ribosomal scanning and two ribosome termination-reinitiation events are required for Pol translation (Figure 2A).[11–13] To assess this model, we utilized DMS to identify known or potentially unknown *cis*-acting elements in the Core region that influence Pol translation. The plasmid libraries used as templates for pgRNA synthesis were constructed as described in the STAR Methods, and a schematic workflow of the DMS approach is shown in Figure S1A. Briefly, we transfected a diverse library of pgRNAs that include all codon variants at each position in the Core ORF into hepatoma cells and deep-sequenced HBV DNA 2 days post-transfection. In parallel, we deep-sequenced the plasmid library to represent the input. We calculated the fitness of each codon variant by comparing the read counts from the plasmid with that of the newly formed HBV DNA (reverse-transcribed pgRNA).

Next, we assessed the quality of the data. Analysis indicated that C-to-A and G-to-T transversions were overrepresented, likely due to DNA damage introduced during sequencing library preparation (Figure S1B).[14] Codon variants that are one C-to-A or G-to-T mutation from WT were therefore excluded from further analysis. In addition, several codon positions yielded low read counts, likely due to suboptimal mutagenesis during library construction (Figure S1C). To reduce the impact these sites of low-confidence data have on the visual representation of results, we reduced the square size for codons with low read counts in the heatmap in Figure 2B.

The heatmap indicates little selection against mutations in the Core reading frame upstream of Pol. For example, pgRNAs harboring early termination codons in Core are not purged from the population. These results are in concordance with *trans* complementation by functional Core proteins from co-transfected pgRNAs, as we observed in Figure 1C. By contrast, strong signals of negative selection are observed in the region of the Core ORF that overlaps with Pol, which is likely driven by unfavorable amino acid changes in Pol.

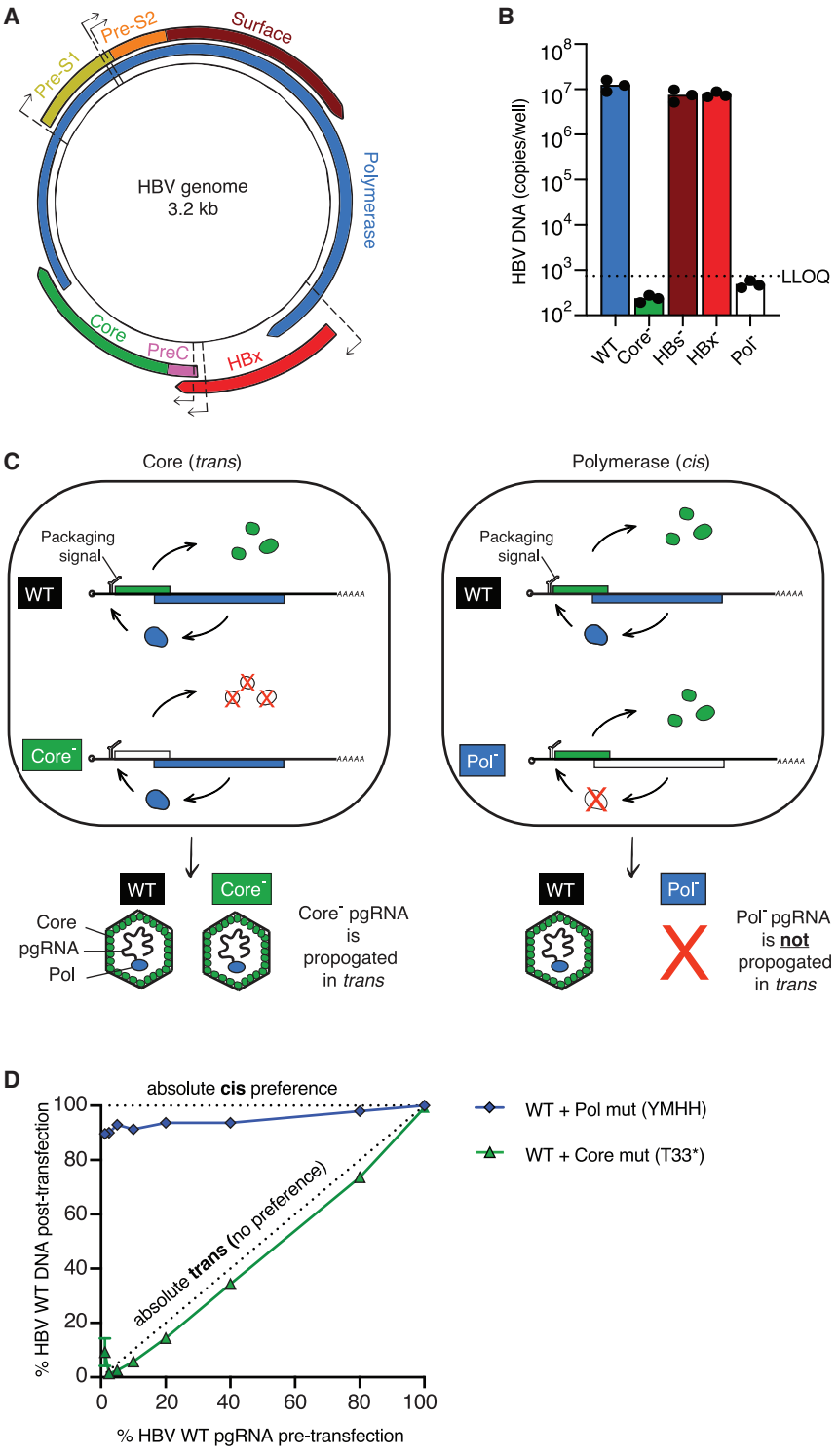


Figure 1. HBV pgRNA transfection can uncouple selection from overlapping reading frames

(A) Schematic of the HBV genome with colored boxes representing ORFs and arrows indicating RNA Pol II transcriptional start sites.

(B) HBV DNA copy number measured by qPCR 2 days after transfecting cells with WT or mutant pgRNAs. Core$^-$ (T33*), HBs$^-$ (C69*), HBx$^-$ (G27*), Pol$^-$ (YMHH active site mutant). Data plotted are mean, with each replicate plotted as a dot. Lower limit of quantification (LLOQ).

(C) Schematic to illustrate *trans* (Core) and *cis* (Pol) activity when multiple pgRNAs are present in a single cell. Left, pgRNA encoding defective Core protein can be encapsidated in *trans* by functional Core from another pgRNA and reverse transcribed to produce HBV DNA. Right, Pol works in *cis*; therefore, pgRNAs encoding defective Pol protein are not reverse transcribed by functional Pol encoded by another pgRNA.

(D) WT and mutant pgRNAs encoding Pol$^-$ or Core$^-$ were co-transfected at the ratios indicated on the x axis. The percentage of WT HBV DNA from total DNA recovered, as determined by amplicon sequencing, is plotted on the y axis. The horizontal dotted line indicates the expected percentage of WT HBV DNA, assuming an absolute *cis* preference. A diagonal dotted line indicates the expected percentage of WT HBV DNA, assuming no *cis* preference. Each symbol represents the mean of three replicates. Error bars are ±SEM.

See also Table S1.

Though there is little selection in the Core-only region, a few notable exceptions exist. Methionine codons are depleted throughout Core, as evidenced by low fitness scores across the Core-only region (see horizontal blue stripe in the heatmap). This depletion is quantified in Figure 2C. This is consistent with the leaky scanning model for Pol translation since ribosome re-initiation in the Core reading frame will reduce Pol translation and, subsequently, the production of HBV DNA. We also observed a slight but statistically significant enrichment of in-frame termination codons upstream of Pol (Figure 2C). These early termination codons could allow ribosomes to re-initiate and resume scanning to the Pol start codon. Also notable in the heatmap are three short, vertical sites of negative selection in the Core-only region. The first major site occurs one codon upstream of an in-frame methionine in Core, known as C1, and two additional selection sites occur at positions flanking the J ORF. The Chen et al. model for leaky scanning suggests that C1 is in a suboptimal context for translation initiation (poor Kozak sequence) (Figure S2A), and indeed, we found that nucleotide variants that improve the C1 Kozak score are underrepresented in the newly formed HBV DNA (Figures 2D and S2B). Conversely, mutations upstream of the J start codon that decrease Kozak consensus or abolish the J start or termination codons are deleterious (Figures 2D, S2C, and S2D). This, too, is in line with the Chen

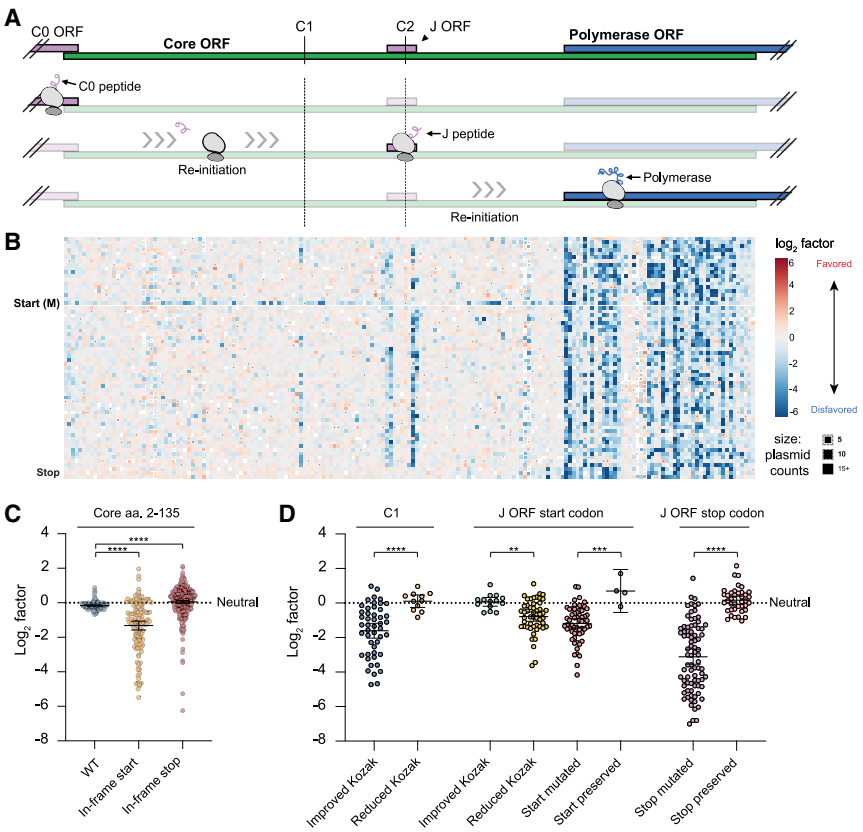


Figure 2. Pol translation model and deep mutational scanning the Core region

(A) A Pol translation model that involves ribosome re-initiation is shown in the context of pgRNA and its open reading frames (ORFs). First, a ribosome translates the C0 ORF, enabling it to bypass the Core start codon. The ribosome then re-initiates scanning downstream of C0, bypasses the C1 methionine codon due to poor Kozak context, and initiates translation at the J ORF, which allows the ribosome to bypass the C2 methionine codon. Lastly, the ribosome re-initiates scanning after translating the J ORF and initiates translation at the Pol start codon.

(B) DMS heatmap of Core, codons 2–186. The 64 codons are depicted on the y axis, with start and termination (stop) codons highlighted. Site positions are portrayed on the x axis. The color of heatmap squares correlates to the $\log_2$ enrichment factor of a given variant. WT codons are symbolized with black dots. White squares are filtered-out variants one G-to-T or C-to-A transversion away from the WT codon and are likely affected by oxidative damage during sequencing library preparation. The size of heatmap squares is reduced if read counts for the plasmid library reference were low.

(C) In-frame start and stop variants in Core are separated into groups for statistical comparison with WT. Data are mean with 95% confidence intervals.

(D) Variants that improve or reduce the Kozak sequence one codon upstream of the C1 or the J ORF start codon are grouped for statistical analysis. Variants that mutate or preserve the J ORF start or stop codon are grouped as well. Data are mean with 95% confidence intervals. $^{**}p < 0.01$. $^{***}p < 0.001$. $^{****}p < 0.0001$ by Mann-Whitney U test. An interactive version of the heatmap in (B) is at https://hbv-dms.github.io/R1/Fig/Fig_2B.html.
See also Figures S1 and S2.

et al. model.[13] Overall, the results from this comprehensive mutagenesis support a leaky scanning model for HBV Pol translation.

DMS uncouples selective pressure on Pol from its overlapping proteins

Next, we examined sequence requirements in Pol. The Pol DMS data were processed like the Core DMS data and displayed as a heatmap (Figures 3A and S3A), which we then compared with a heatmap profile of HBV derived from natural sequences available in the UniRef100 database (Figure S3B).[15] Due to the *cis* preference of Pol and the initiation of HBV replication with pgRNA transfection, our method is uniquely poised to study codon and amino acid preferences of Pol in a way that sequencing of HBV clinical isolates cannot achieve.

Upon examining the heatmap of the Pol ORF in Figure 3A, we observed several noteworthy features that indicate the method is robust. For instance, termination codons were negatively selected throughout the Pol ORF, and prolines, which often disrupt protein structure,[16] were disfavored, except in regions with minimal observed selection. Further, amino acids known to be crucial for catalysis, such as the priming tyrosine in the terminal protein (TP) domain, the two aspartic acid residues in the YMDD motif in the RT domain, and the DEDD motif in the RNase H domain, were intolerant to change.[17]

Next, we used AlphaFold 2 to generate a three-dimensional structure prediction of the Pol protein (Figure 3B) to draw relationships between predicted structural features and the experimental fitness map.[18] The structure prediction we obtained is similar to a recently described prediction of Pol from genotype D HBV[19] Our DMS results agree with both structural predictions and suggest that the predicted partition between the spacer and RT domains should be adjusted. The initial partition was based on a sequence alignment with retroviral RTs,[20] but it was subsequently shown that three cysteine residues in the spacer domain and one cysteine within the predicted HBV RT domain were essential for pgRNA packaging.[21] Based on this, it was hypothesized that these cysteine residues might form a zinc finger.[21] Together, the AlphaFold 2 structure and our DMS results indicate that the spacer domain comprises amino acids (aa) Q179 to S324 in the genotype A HBV sequence used in this study. We observed a requirement for four cysteine residues in an N-terminal extension of the RT domain (aa 325–354) that likely bind zinc. Interestingly, the predicted structure and the DMS fitness map indicate not one but two zinc fingers, both of which require amino acids in a downstream region of RT that we hereafter refer to as the RT insertion (aa 456–508) (Figures 3B and 3C). This N-terminal extension and RT insertion are conserved in members of the

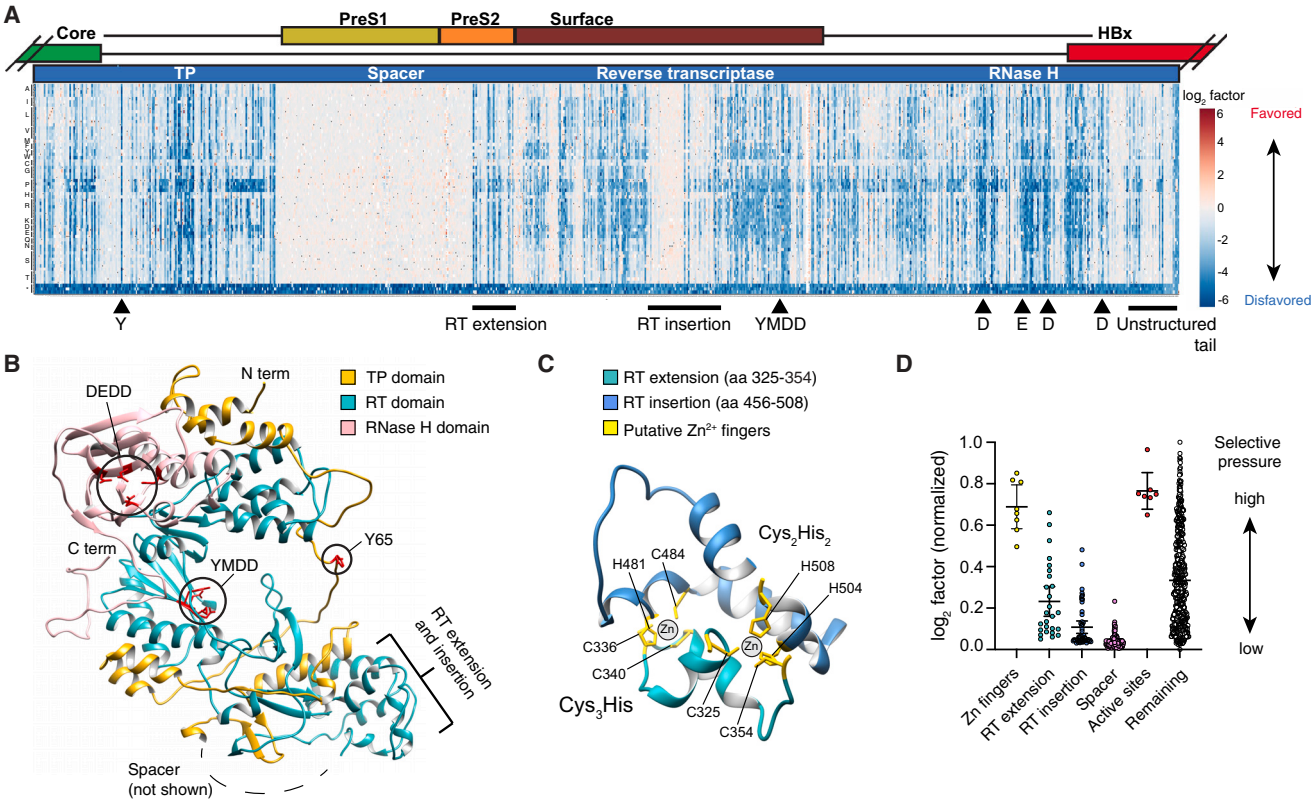


Figure 3. Polymerase deep mutational scanning

(A) Heatmap depicting the fitness of polymerase variants is plotted as in Figure 2B. Pol and its overlapping reading frames are shown as colored boxes (top). Catalytic site amino acids (triangles) and regions of interest (lines) are indicated below the heatmap.

(B) AlphaFold 2 structural predictions of HBV gtA Pol. TP, RT, and RNase H domains are indicated by color. The spacer domain is unstructured and is not shown. N and C termini are labeled. Catalytic site amino acids are circled and labeled. The region of interaction between the RT extension and insertion is labeled.

(C) Close-up view of the predicted interaction between the RT extension and insertion. RT extension and insertion are indicated by color. Amino acids involved in two putative zinc fingers are colored yellow and labeled. The possible locations of zinc ions are shown as gray circles.

(D) Normalized $\log_2$ selective pressure. The absolute value of the $\log_2$ fold change in frequency of all sense codons at each aa position was added to calculate the cumulative selective pressure. All values were normalized to the highest and lowest value in the dataset. Each aa position is represented by a dot. Specific aa were subsetted and plotted as indicated. RT extension (aa 325–354) and insertion (aa 456–508), excluding aa in the putative zinc-finger motifs. Spacer (aa 179–324). Active sites (Y65, D553, D554, D702, E731, D750, and D790). Remaining (all aa excluding those plotted in other categories). Mean and 95% confidence intervals are shown. An interactive version of the heatmap in (A) is at https://hbv-dms.github.io/R1/Fig/Fig_3A.html.

See also Figures S3 and S4, Table S3, and Data S1.

orthohepadnavirus genus but absent in retroviruses and other members of the *hepadnaviridae* family (Figure S4).

The cysteine and histidine residues in the N-terminal RT extension and RT insertion regions are selected to a degree comparable to the seven active site residues, despite little apparent selection of flanking amino acids (Figure 3D). Interestingly, the RT insertion, most of which is tolerant to mutations in the Pol reading frame, corresponds to the nucleotides in the overlapping surface reading frame that encode the major hydrophilic region (Figure S4), corresponding to a loop in the surface protein highly targeted by antibodies.[22]

The C terminus of HBV Pol encodes a ribosome-stalling motif

In the AlphaFold 2 prediction, the C terminus of Pol downstream of the RNase H domain (aa 810–845) is unstructured (Figure 4A). A closer view of the DMS data in this region revealed that selection was generally low, except for a few sites, including two proline residues (aa 844–845) directly preceding the termination codon (Figure 4B). In addition to the overlap with HBx, this unstructured Pol region also contains direct repeat 2 (DR2), which is important for proper plus-strand DNA synthesis; however, since the production of full-length minus-strand DNA is sufficient to produce sequencing libraries,[5,6] we do not capture selection on DR2. Further, the proline codon-specific signature suggests that selection in this region acts on Pol rather than HBx, DR2, or unknown *cis*-acting elements.

Proline is unique among amino acids in that its side chain is covalently bonded to the α-amine of the peptide backbone. This bond results in a rigid structure and reduces the number of hydrogen atoms on the α-amine, which renders proline both a poor donor and acceptor for peptide bond formation.[23] Consequently, proline codons, which lead to the incorporation of proline amino acids during protein synthesis, have been observed

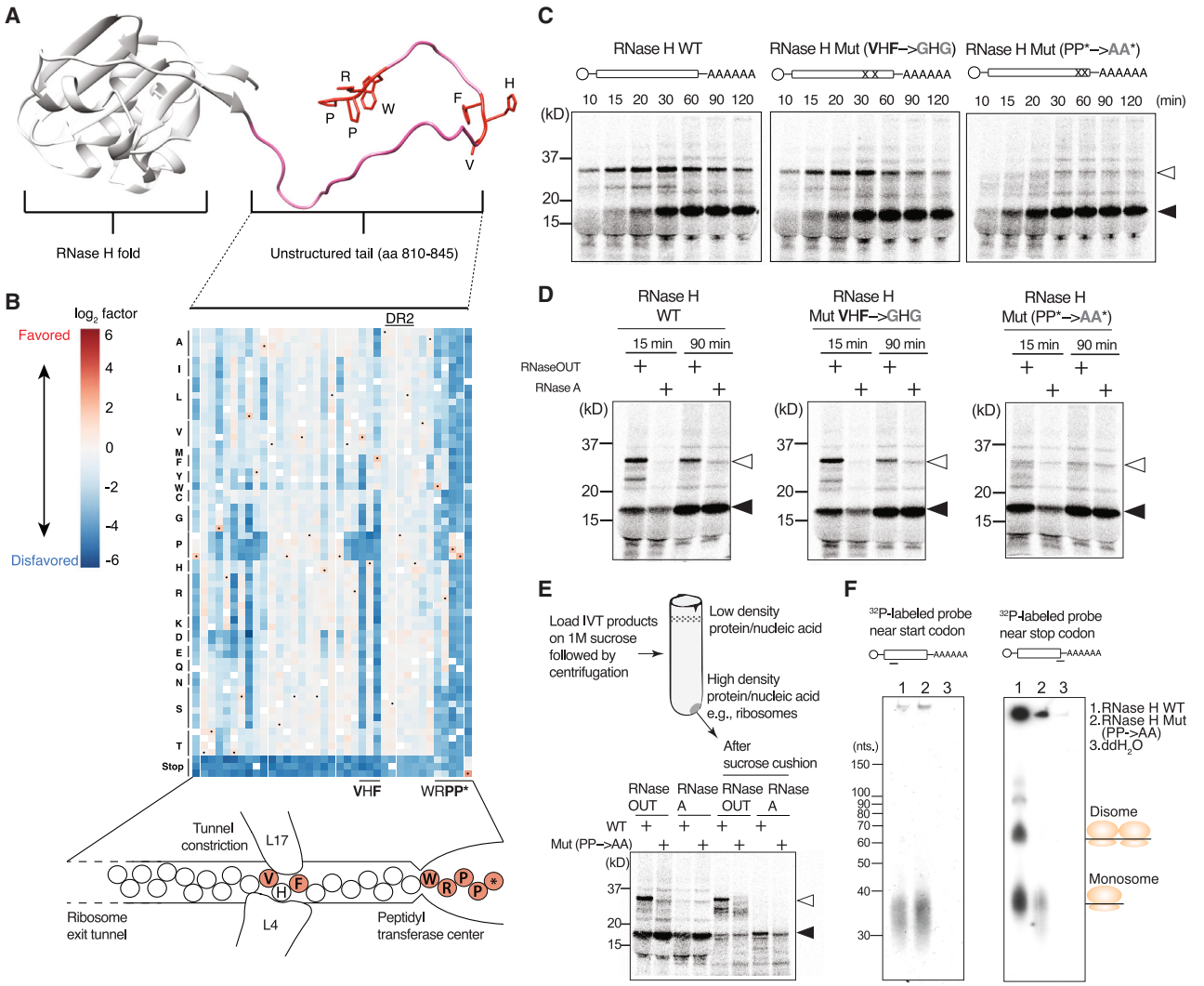


Figure 4. HBV Pol C terminus encodes a ribosome-stalling motif

(A) AlphaFold 2 structure prediction of the HBV RNase H domain. The classical RNase H fold is labeled, and the unstructured tail (aa 810–845) is colored pink and labeled. N and C termini are labeled.

(B) Enlarged heatmap of the unstructured tail (aa 810–845) is plotted as in Figure 2B. The predicted locations of aa in the ribosome exit tunnel are indicated below the heatmap. Amino acids with strong selective pressure are colored and labeled. L4 and L17; ribosomal proteins. Direct repeat 2 (DR2) is labeled.

(C) WT and mutant HBV RNase H constructs (top), produced for the indicated periods *in vitro* in the presence of ^{35}S-methionine, were separated by SDS-PAGE and analyzed by autoradiography. A solid triangle indicates the expected size of the full-length product. An open triangle indicates tRNA-bound intermediates.

(D) WT and mutant HBV RNase H constructs, produced as in (C) for the indicated periods of time, were treated with RNase OUT or RNase A for 15 min to protect or degrade covalently bound tRNA, respectively.

(E) WT and mutant HBV RNase H translation products produced as in (C) (15 min) were separated by centrifugation through a 1 M sucrose cushion, as depicted on the top. Pellets containing high-density protein/nucleic acids were resuspended and then treated with RNase OUT or RNase A for 15 min to protect or degrade covalently bound tRNA, respectively. The ^{35}S-methionine-labeled proteins were analyzed by SDS-PAGE, as shown on the bottom.

(F) Northern blot to detect RNase I-protected fragments from *in vitro* translation assays (15 min) using WT or Mut (PP → AA) templates as in (C). ^{32}P-labeled probes that hybridized either near the start codon (left) or the termination codon (right) were used to detect protected RNA fragments. The samples loaded in each of the three lanes are on the right of the blots. The approximate size of monosome- and disome-protected RNA fragments is indicated on the right.

See also Figure S5, Tables S1 and S3, and Data S1.

to stall elongating ribosomes.[24] In addition to these terminal prolines, we also observed a strong selection for two amino acids (V832 and F834) within the unstructured tail. When the peptidyl transferase center of the ribosome is placed over the two terminal proline codons, the VHF amino acids would be located near a constriction in the ribosome exit tunnel formed by two large ribosomal proteins, L4 and L17 (Figure 4B, bottom).[25,26] Based on the observed selection on the VHF sequence and the terminal prolines, we hypothesized that the proline codons encoding the last two amino acids of HBV Pol might promote ribosome

stalling and that the ~36 aa unstructured tail could serve as a spacer to allow the RNase H domain to emerge from the ribosome exit tunnel and form predicted contacts with the TP and RT domains all while the nascent Pol protein remains tethered to pgRNA by the ribosome. Ribosome stalling at this location could, therefore, facilitate Pol interaction with the RNA packaging signal (epsilon) and provide a mechanism to explain Pol's *cis*-preferential pgRNA packaging and reverse transcription.

If the above hypothesis is correct, the ribosome-stalling mechanism might be conserved among relatives. The VHF and PP motifs are indeed conserved among orthohepadnaviruses but not among more distantly related hepadnaviruses (Figures S5A and S5B). However, we noticed an apparent enrichment of prolines in the C terminus of more distantly related hepadnaviruses, which may confer a similar function. We quantified this observed proline enrichment in Figure S5C.

To test whether the VHF motif and the terminal prolines in HBV Pol stall ribosomes, we produced several HBV RNase H domain expression constructs: WT, VHF mutant (GHG), and diproline mutant (AA*) (Figure 4C, top). Next, we produced 5′ capped and 3′ polyadenylated mRNAs *in vitro* with T7 RNA polymerase, added these mRNAs to rabbit reticulocyte lysate (RRL) in the presence of ^{35}S-methionine, and visualized protein products by SDS-polyacrylamide gel electrophoresis (PAGE) under neutral pH.[27] Under these conditions, peptidyl-tRNA ester bonds formed during translation are preserved,[27,28] and the appearance of a ^{35}S-labeled band migrating ~18 kDa larger than the predicted product is consistent with the accumulation of a tRNA-bound protein intermediate stabilized by a stalled ribosome. Gel-shifted bands of this size have been observed in similar assays with mRNAs encoding SecM and XBP1 ribosome-stalling motifs.[27,28]

As shown in Figure 4C, over a time course of translation, we observed the accumulation of two major bands, one migrating at the expected size of the HBV RNase H domain (16.7 kDa) and a higher molecular weight band (~35 kDa), which is consistent with the size of a tRNA-bound protein intermediate. Notably, the upper band is prominent in both the WT and GHG constructs (left and middle) but not in the diproline mutant construct (right). In both the WT and GHG panels, the upper band is detected at the earliest time points (10, 15, and 20 min), even before the mature protein accumulates. The intensity of the upper band then gradually decreases at later time points (90 and 120 min), supporting the idea that this product is an intermediate formed before ribosomes release the mature RNase H protein product. These results indicate that the C-terminal prolines in HBV Pol stall translation, whereas the VHF motif likely does not.

To further confirm that the upper band was indeed the RNase H domain covalently bound to tRNA, we treated the samples with either RNase A to degrade the tRNA or RNase OUT as a control to protect the tRNA. As expected, RNase A treatment eliminated the upper band (Figure 4D).

Ribosomal subunits may disassociate and release the protein with a covalently bound tRNA. To test whether the tRNA-bound protein remained associated with the ribosome, as would be the case during ribosome stalling, we separated high- and low-density protein/nucleic acids from the *in vitro* translation reactions by centrifugation through a 1 M sucrose cushion. Under these conditions, ribosomes pellet at the bottom of the tube, and lower-density proteins remain at the top. We analyzed the pellets by SDS-PAGE and found that the ^{35}S-labeled tRNA-bound protein remained associated with ribosomes, whereas most of the mature protein did not (Figure 4E). To further confirm that ribosomes stall at the terminal prolines, we used *in vitro* translation reactions as input for an RNase protection assay and visualized protected fragments by northern blot. A ^{32}P-labeled probe complementary to the 5′ end of the mRNA near the initiation codon detected similar amounts of RNA in both WT and proline mutant constructs consistent in size with fragments protected by a single translating ribosome (monosome).[29] By contrast, a ^{32}P-labeled probe complementary to the mRNA near the termination codons detects an increased signal from samples containing the WT template compared with reactions containing the proline mutant template. Further, larger fragments were also detected, consistent with the expected size of fragments protected by multiple adjacent ribosomes (Figure 4F).[30] Altogether, these results support our hypothesis that proline codons near the termination codon of the Pol ORF stall ribosomes.

HBV Pol ribosome-stalling motif ensures *cis*-preferential packaging and reverse transcription

Translation stalling has been reported as a mechanism to facilitate proper protein folding and localization.[31] Next, we aimed to test our hypothesis that ribosome stalling at the end of the Pol ORF promotes the interaction between the nascent Pol protein and its pgRNA template, thereby ensuring *cis*-preferential pgRNA packaging and reverse transcription. We performed co-transfection experiments like those in Figure 1D to test this hypothesis. We used pgRNA mutants shown in Figure 5A, which include a mutant defective in reverse transcription (YMHH), a mutation in one of the putative zinc fingers that leads to a defect in pgRNA packaging (C336A; Figure 3; Kim et al.[21]) as well as the VHF (GHG) and diproline (AA*) mutants described in Figure 4. The bar graph in Figure 5B compares the fitness of each of these mutants as measured by qPCR of HBV DNA. Notably, the GHG and AA* mutants produced approximately 10- to 100-fold less DNA than WT but retained activity. To quantify the *trans*-complementation activity of these mutants, we mixed an equal ratio (1:1) of WT or mutant pgRNAs with YMHH-containing mutants and sequenced the YMHH motif (Figure 5C).

Our results indicated that when WT pgRNA was mixed equally with YMHH pgRNA, only 2% of the recovered HBV DNA contained the YMHH mutation, which is consistent with *cis* preference. However, when the AA* mutation was introduced into these pgRNAs, nearly 20% of the HBV DNA recovered contained the YMHH mutation, indicating that the diprolines at the end of Pol indeed enforce *cis* activity.

Although the YMHH Pol mutant cannot reverse transcribe pgRNA, it can likely still bind pgRNA and potentially block a functional Pol from acting in *trans*. Therefore, we introduced an additional mutant (C336A) to impair pgRNA binding. This further increased *trans* activity, producing over 30% of the HBV DNA containing the YMHH mutation. We next repeated this over a range of ratios and obtained consistent results (Figure S6). And finally, as predicted by the *in vitro* translation results in Figures 4C and 4D, when we co-transfected the triple mutant

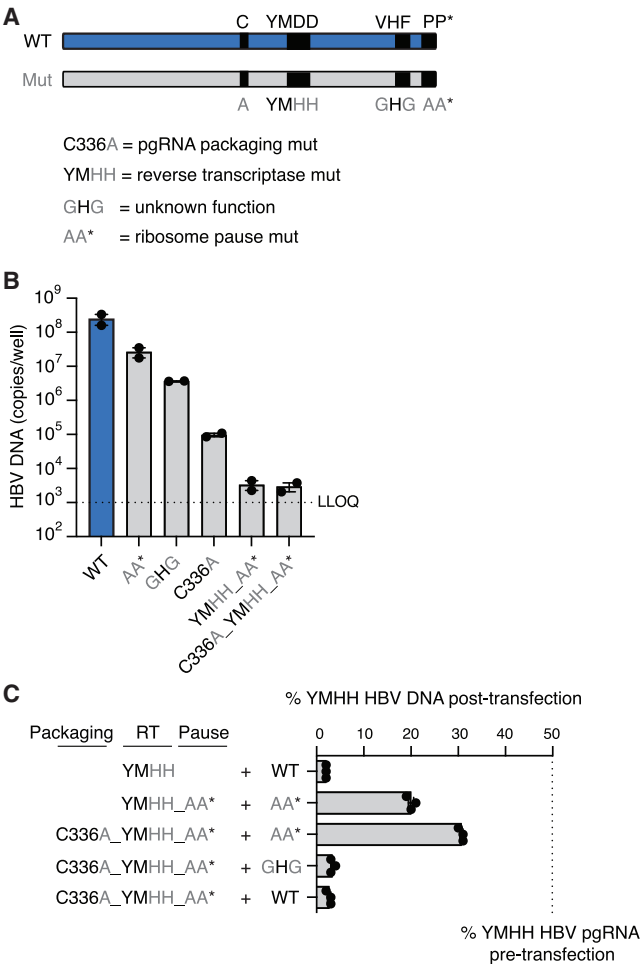


Figure 5. C-terminal prolines contribute to HBV polymerase's *cis* preference

(A) Schematic of HBV Pol mutations.

(B) The effect of mutations on HBV Pol's ability to package and reverse transcribe pgRNA; qPCR quantification of HBV DNA copies per well of a 6-well plate. Replicates are plotted, and column height indicates the mean. Error bars are SEM. LLOQ, lower limit of quantification.

(C) The pgRNAs indicated were co-transfected at equal ratios (50:50), and the percentage of YMHH mutant HBV DNA recovered is plotted. The dotted line indicates the percentage of YMHH pgRNA transfected. Each replicate is plotted, and column height indicates the mean. Error bars are SEM.

See also Figure S6 and Table S1.

(C336A_YMHH_AA*) with WT pgRNA or pgRNA encoding the GHG mutant, almost all the HBV DNA recovered was WT (Figure 5C). These findings demonstrate that the C-terminal prolines in HBV Pol are required for *cis* preference. Based on these results, we propose the model shown in Figure 6.

DISCUSSION

In a previous study, we showed that initiating HBV genome replication with *in vitro*-transcribed pgRNA allowed us to assess the fitness of rare variants in a mixed population.[5] Here, we combined this approach with DMS to assess the fitness of thousands

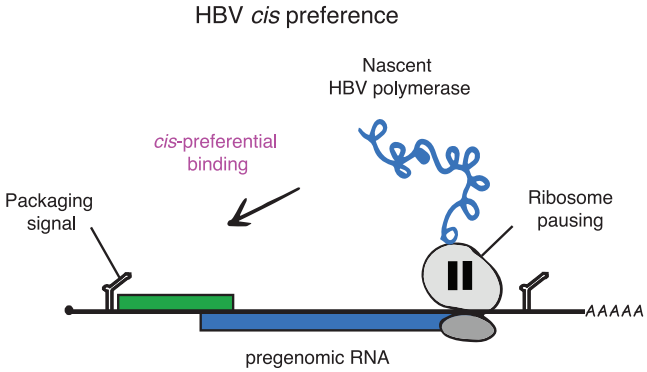


Figure 6. Ribosome stalling leads to *cis*-preferential reverse transcription of HBV pgRNA

A model to explain the molecular basis of HBV Pol's *cis* preference. The model posits that the proline codons encoding the C terminus of Pol stalls ribosomes, thereby tethering Pol to the pgRNA template to facilitate *cis*-preferential binding of the Pol protein to the packaging signal (epsilon), ultimately resulting in *cis*-preferential pgRNA reverse transcription.

of HBV variants in a cell culture assay. This assay measured the frequency that an HBV pgRNA is reverse transcribed to full-length minus-strand HBV DNA. Although the limitations of the assay prevented us from capturing the full spectrum of selective pressures on HBV in natural infections, it afforded us the unique opportunity to selectively visualize several *cis*-acting functions encoded in the HBV Core and Pol ORFs. This proved particularly informative in regions where the genome encodes two proteins in overlapping reading frames. For example, our inability to capture selective pressure on the Core protein enabled us to visualize underlying *cis*-acting elements that control Pol translation. This allowed us to confirm results from past studies and clarify the mechanism of HBV Pol translation, which remains uncertain to this day. For example, Lin and Lo[11] and Fouillet et al.[12] initially proposed that HBV Pol is translated by leaky ribosome scanning. This model was later refined by Chen et al.[13] However, the model was still questioned[32] because it contrasts with a model for duck HBV (DHBV) Pol translation, which is believed to involve ribosome shunting.[33,34] The DMS results overwhelmingly support a leaky ribosome scanning model for HBV Pol translation.

The combination of DMS with pgRNA transfection also enabled us to uncouple Pol amino acid requirements from other viral proteins and acquire a fitness landscape, which, when combined with the Pol AlphaFold 2 structural prediction, provided a powerful tool to uncover HBV biology. One interesting discovery was that indispensable cysteines in the N-terminal extension of the RT domain (formerly thought to be part of the spacer domain) likely form long-distance contacts with cysteine and histidine residues in the RT insertion to form two zinc-finger motifs. These amino acids are conserved in orthohepadnaviruses and, based on previous literature, may be important for pgRNA packaging.[21,35] It is possible that another function of the zinc fingers is to provide structural stability to an otherwise flexible loop in Pol. Such a loop may provide the surface protein in the overlapping reading frame with the mutational freedom needed to escape antibody neutralization.

Our most unexpected discovery was that the unstructured C terminus of HBV Pol encodes a ribosome-stalling motif that enforces Pol's *cis* preference. It has been known for decades that HBV Pol exhibits a *cis* preference for pgRNA binding and reverse transcription,[7] but the mechanism remained a mystery. We showed that mutating the two terminal prolines of this motif reduces ribosome stalling *in vitro* and decreases Pol's *cis* activity in cells. Now, with mechanistic insight into how HBV Pol achieves *cis* preference, it is possible to understand why it is important. One possibility worthy of investigation is whether *cis* preference provides a mechanism to prevent the accumulation of defective viral genomes. A deeper understanding of *cis* preference and the mechanisms by which it is regulated may ultimately highlight weaknesses in HBV to exploit therapeutically.

Lastly, as we found for HBV Pol, it has been proposed that the ORF2 protein of long interspersed nuclear element 1 (LINE-1), an endogenous retroelement, may also reverse transcribe its template mRNA by a mechanism coined "*cis* preference by elongation arrest."[36] Although the LINE-1 mechanism has yet to be illuminated, *cis* preference by ribosome stalling may be a more general feature conserved across biology. With possible roles in the *cis*-preferential replication of additional viral genomes or, more broadly, in the formation of ribonucleoprotein (RNP) complexes, there is much to discover.

Limitations of the study

The DMS results here capture only a fraction of the selective pressures acting on the viral genome throughout the life cycle. As such, the results are limited to the requirements for translating the Pol protein and its ability to convert pgRNA to full-length minus-strand DNA. Consequently, many variants that maintain fitness under the conditions in this study will be unfit in the context of a natural infection. We also acknowledge that despite demonstrating the importance of C-terminal proline codons in the HBV Pol ORF in conferring *cis*-preferential pgRNA packaging and reverse transcription, gaps remain in our mechanistic understanding of this process. Some *cis* preference remains even when these critical proline codons are mutated, suggesting that additional features may contribute to this phenotype. Additional codons or amino acids in the C terminus of HBV Pol may have a role in ribosome stalling. This can be further explored by directly assessing the impact of additional mutations on ribosome stalling *in vitro*. Quantifying *cis* preference in WT and mutant versions of distantly related hepadnaviruses may also illuminate this topic. It is also possible that *cis* preference requires a low level of Pol translation to maintain a low Pol:pgRNA ratio. Lastly, it is unclear how the observed ribosome stalling is resolved and if physical force[37] or additional viral or host factors are required to release the mature protein from the ribosome. Perhaps most interesting will be determining why *cis* preference is important for the virus life cycle, which will require testing these mutants in a more natural setting that includes robust virus amplification and spread.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.04.008.

ACKNOWLEDGMENTS

We thank Dr. Jesse Bloom (Fred Hutchinson Cancer Center) for early advice on implementing DMS. We thank Shira Weingarten-Gabbay, Hans-Heinrich Hoffmann, Paul Bieniasz, and J.T. Poirier for their helpful input and discussion. We also thank the Rockefeller University Genomics, High Performance Computing, and Proteomics Resource Centers. Funding was received from the following: National Institutes of Health grant R01AI091707 (C.M.R.), National Institutes of Health grant R01AI143295 (C.M.R.), National Institutes of Health grant R01AI150275 (C.M.R.), German Center for Infectious Research (DZIF) MD stipend (M.A.K.), the National Center Advancing Translational Sciences, the National Institutes of Health (through Rockefeller University) grant #UL1 TR001866 (L.L.S.), the Robertson Foundation (Y.Y., W.M.S., and C.M.R.), and anonymous donors (C.M.R.). D.S.M. holds a Ben Barres Early Career Award from the Chan Zuckerberg Initiative as part of the Neurodegeneration Challenge Network, CZI2018-191853, and is supported by the Coalition for Epidemic Preparedness Innovations (CEPI).

AUTHOR CONTRIBUTIONS

Conceptualization, Y.Y., W.M.S., and C.M.R.; formal analysis and software, Y.Y., M.A.K., and N.Y.; investigation, Y.Y., M.A.K., M.Z., N.Y., C.A.F., K.P.B., L.C.A., L.L.S., S.V., X.H., A.S., Y.P.d.J., and W.M.S.; supervision, Y.P.d.J., D.S.M., W.M.S., and C.M.R.; visualization, Y.Y., M.A.K., N.Y., and W.M.S.; writing – original draft, Y.Y., M.A.K., and W.M.S.; writing – review & editing, Y.Y., M.A.K., M.Z., N.Y., C.A.F., K.P.B., L.C.A., L.L.S., S.V., X.H., A.S., Y.P.d.J., D.S.M., C.M.R., and W.M.S.

DECLARATION OF INTERESTS

Y.Y., W.M.S., and C.M.R. filed a patent application, US 62/741,032, with Rockefeller University on September 19, 2019, entitled "RNA-Based Methods to Launch Hepatitis B Virus Infection." Patent pending. C.M.R. is a shareholder and member of the scientific advisory board at VIR Biotechnology.

Received: August 2, 2023
Revised: February 12, 2024
Accepted: April 10, 2024
Published: May 8, 2024

REFERENCES

1. WHO (2021). Global Progress Report on HIV, Viral Hepatitis and Sexually Transmitted Infections (WHO).
2. Nassal, M. (2015). HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. Gut *64*, 1972–1984. https://doi.org/10.1136/gutjnl-2015-309809.
3. Martinez, M.G., Boyd, A., Combe, E., Testoni, B., and Zoulim, F. (2021). Covalently closed circular DNA: the ultimate therapeutic target for curing HBV infections. J. Hepatol. *75*, 706–717. https://doi.org/10.1016/j.jhep.2021.05.013.
4. Wright, B.W., Molloy, M.P., and Jaschke, P.R. (2022). Overlapping genes in natural and engineered genomes. Nat. Rev. Genet. *23*, 154–168. https://doi.org/10.1038/s41576-021-00417-w.
5. Yu, Y., Schneider, W.M., Kass, M.A., Michailidis, E., Acevedo, A., Pamplona Mosimann, A.L., Bordignon, J., Koenig, A., Livingston, C.M., van Gijzel, H., et al. (2023). An RNA-based system to study hepatitis B virus replication and evaluate antivirals. Sci. Adv. *9*, eadg6265. https://doi.org/10.1126/sciadv.adg6265.
6. Chang, C., Zhou, S., Ganem, D., and Standring, D.N. (1994). Phenotypic mixing between different hepadnavirus nucleocapsid proteins reveals C protein dimerization to be cis preferential. J. Virol. *68*, 5225–5231. https://doi.org/10.1128/JVI.68.8.5225-5231.1994.
7. Bartenschlager, R., Junker-Niepmann, M., and Schaller, H. (1990). The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. J. Virol. *64*, 5324–5332. https://doi.org/10.1128/JVI.64.11.5324-5332.1990.
8. Leupin, O., Bontron, S., Schaeffer, C., and Strubin, M. (2005). Hepatitis B virus X protein stimulates viral genome replication via a DDB1-dependent pathway distinct from that leading to cell death. J. Virol. *79*, 4238–4245. https://doi.org/10.1128/JVI.79.7.4238-4245.2005.
9. Yeh, C.T., Chien, R.N., Chu, C.M., and Liaw, Y.F. (2000). Clearance of the original hepatitis B virus YMDD-motif mutants with emergence of distinct lamivudine-resistant mutants during prolonged lamivudine therapy. Hepatology *31*, 1318–1326. https://doi.org/10.1053/jhep.2000.7296.
10. Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose, V., and Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell *77*, 943–950. https://doi.org/10.1016/0092-8674(94)90142-2.
11. Lin, C.G., and Lo, S.J. (1992). Evidence for involvement of a ribosomal leaky scanning mechanism in the translation of the hepatitis B virus pol gene from the viral pregenome RNA. Virology *188*, 342–352. https://doi.org/10.1016/0042-6822(92)90763-F.
12. Fouillot, N., Tlouzeau, S., Rossignol, J.M., and Jean-Jean, O. (1993). Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation. J. Virol. *67*, 4886–4895. https://doi.org/10.1128/JVI.67.8.4886-4895.1993.
13. Chen, A., Kao, Y.F., and Brown, C.M. (2005). Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons. Nucleic Acids Res. *33*, 1169–1181. https://doi.org/10.1093/nar/gki251.
14. Hom, N., Gentles, L., Bloom, J.D., and Lee, K.K. (2019). Deep mutational scan of the highly conserved influenza A virus M1 matrix protein reveals substantial intrinsic mutational tolerance. J. Virol. *93*, e00161. e00119. https://doi.org/10.1128/JVI.00161-19.
15. Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C.H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics *23*, 1282–1288. https://doi.org/10.1093/bioinformatics/btm098.
16. MacArthur, M.W., and Thornton, J.M. (1991). Influence of proline residues on protein conformation. J. Mol. Biol. *218*, 397–412. https://doi.org/10.1016/0022-2836(91)90721-H.
17. Clark, D.N., and Hu, J. (2015). Unveiling the roles of HBV polymerase for new antiviral strategies. Future Virol. *10*, 283–295. https://doi.org/10.2217/fvl.14.113.
18. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
19. Tajwar, R., Bradley, D.P., Ponzar, N.L., and Tavis, J.E. (2022). Predicted structure of the hepatitis B virus polymerase reveals an ancient conserved protein fold. Protein Sci. *31*, e4421. https://doi.org/10.1002/pro.4421.
20. Radziwill, G., Tucker, W., and Schaller, H. (1990). Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. J. Virol. *64*, 613–620. https://doi.org/10.1128/JVI.64.2.613-620.1990.
21. Kim, S., Lee, J., and Ryu, W.-S. (2009). Four conserved cysteine residues of the hepatitis B virus polymerase are critical for RNA pregenome encapsidation. J. Virol. *83*, 8032–8040. https://doi.org/10.1128/JVI.00332-09.
22. Carman, W.F. (1997). The clinical significance of surface antigen variants of hepatitis B virus. J. Viral Hepat. *4*, 11–20. https://doi.org/10.1111/j.1365-2893.1997.tb00155.x.
23. Melnikov, S., Mailliot, J., Rigger, L., Neuner, S., Shin, B.-S., Yusupova, G., Dever, T.E., Micura, R., and Yusupov, M. (2016). Molecular insights into protein synthesis with proline residues. EMBO Rep. *17*, 1776–1784. https://doi.org/10.15252/embr.201642943.
24. Hayes, C.S., Bose, B., and Sauer, R.T. (2002). Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. J. Biol. Chem. *277*, 33825–33832. https://doi.org/10.1074/jbc.M205405200.
25. Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science *289*, 920–930. https://doi.org/10.1126/science.289.5481.920.
26. Ito, K., and Chiba, S. (2013). Arrest peptides: cis-acting modulators of translation. Annu. Rev. Biochem. *82*, 171–202. https://doi.org/10.1146/annurev-biochem-080211-105026.
27. Yanagitani, K., Kimata, Y., Kadokura, H., and Kohno, K. (2011). Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. Science *331*, 586–589. https://doi.org/10.1126/science.1197142.
28. Muto, H., Nakatogawa, H., and Ito, K. (2006). Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. Mol. Cell *22*, 545–552. https://doi.org/10.1016/j.molcel.2006.03.033.
29. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223. https://doi.org/10.1126/science.1168978.
30. Han, P., Shichino, Y., Schneider-Poetsch, T., Mito, M., Hashimoto, S., Udagawa, T., Kohno, K., Yoshida, M., Mishima, Y., Inada, T., et al. (2020). Genome-wide survey of ribosome collision. Cell Rep. *31*, 107610. https://doi.org/10.1016/j.celrep.2020.107610.
31. Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. *16*, 589–597. https://doi.org/10.1038/nsmb.1614.
32. Chuang, Y.-C., Tsai, K.-N., and Ou, J.-H.J. (2022). Pathogenicity and virulence of hepatitis B virus. Virulence *13*, 258–296. https://doi.org/10.1080/21505594.2022.2028483.
33. Sen, N., Cao, F., and Tavis, J.E. (2004). Translation of Duck Hepatitis B Virus Reverse Transcriptase by Ribosomal Shunting. J. Virol. *78*, 11751–11757. https://doi.org/10.1128/JVI.78.21.11751-11757.2004.

34. Cao, F., and Tavis, J.E. (2011). RNA elements directing translation of the duck hepatitis B virus polymerase via ribosomal shunting. J. Virol. *85*, 6343–6352. https://doi.org/10.1128/JVI.00101-11.

35. Jones, S.A., Clark, D.N., Cao, F., Tavis, J.E., and Hu, J. (2014). Comparative analysis of hepatitis B virus polymerase sequences required for viral RNA binding, RNA packaging, and protein priming. J. Virol. *88*, 1564–1572. https://doi.org/10.1128/JVI.02852-13.

36. Ahl, V., Keller, H., Schmidt, S., and Weichenrieder, O. (2015). Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. Mol. Cell *60*, 715–727. https://doi.org/10.1016/j.molcel.2015.10.003.

37. Nilsson, O.B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Müller-Lucks, A., Trovato, F., Puglisi, J.D., O'Brien, E.P., et al. (2015). Cotranslational protein folding inside the ribosome exit tunnel. Cell Rep. *12*, 1533–1540. https://doi.org/10.1016/j.celrep.2015.07.065.

38. Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. *27*, 221–224. https://doi.org/10.1093/molbev/msp259.

39. Bloom, J.D. (2015). Software for the analysis and visualization of deep mutational scanning data. BMC Bioinformatics *16*, 168. https://doi.org/10.1186/s12859-015-0590-4.

40. Gleason, A.C., Ghadge, G., Sonobe, Y., and Roos, R.P. (2022). Kozak similarity score algorithm identifies alternative translation initiation codons implicated in cancers. Int. J. Mol. Sci. *23*, 10564. https://doi.org/10.3390/ijms231810564.

41. Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Schärfe, C.P.I., Springer, M., Sander, C., and Marks, D.S. (2017). Mutation effects predicted from sequence co-variation. Nat. Biotechnol. *35*, 128–135. https://doi.org/10.1038/nbt.3769.

42. Bloom, J.D. (2014). An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol. Biol. Evol. *31*, 1956–1978. https://doi.org/10.1093/molbev/msu173.

43. Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. Nat. Protoc. *3*, 1077–1084. https://doi.org/10.1038/nprot.2008.67.

44. Lauber, C., Seitz, S., Mattei, S., Suh, A., Beck, J., Herstein, J., Börold, J., Salzburger, W., Kaderali, L., Briggs, J.A.G., et al. (2017). Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. Cell Host Microbe *22*, 387–399.e6. https://doi.org/10.1016/j.chom.2017.07.019.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
XL1-Blue electrocompetent bacteria	Agilent	Cat#200228
Chemicals, peptides, and recombinant proteins		
Dulbecco's Modified Eagle Medium (DMEM)	ThermoFisher Scientific	Cat#11995065
Nonessential amino acids (NEAA)	ThermoFisher Scientific	Cat#11140076
Hyclone fetal bovine serum (FBS)	HyClone Laboratories	Lot#AUJ35777
Lipofectamine™ 2000	ThermoFisher Scientific	Cat#11668019
Opti-MEM Reduced Serum Medium	ThermoFisher Scientific	Cat#51985034
ADB (Agarose Dissolving Buffer)	Zymo Research	Cat#D4001-1-50
EasyTag L-[35S]-Methionine, 500μCi (18.5MBq), Stabilized Aqueous Solution	PerkinElmer	Cat#NEG709A500UC
RNaseOUT™ Recombinant Ribonuclease Inhibitor	ThermoFisher Scientific	Cat#10777019
RNase A, DNase and protease-free (10 mg/mL)	ThermoFisher Scientific	Cat#EN0531
NuPAGE™ LDS Sample Buffer (4X)	ThermoFisher Scientific	Cat#NP0007
NuPAGE™ 4 to 12%, Bis-Tris, 1.0–1.5 mm, Mini Protein Gels	ThermoFisher Scientific	Cat#NP0321BOX
NuPAGE™ MES SDS running buffer	ThermoFisher Scientific	Cat#NP0060
Phosphorimager Exposure Cassette 35 × 43	Cytiva	Cat#29-1755-24
Cycloheximide	Sigma-Aldrich	Cat#C1988-1G
Tube, thickwall, Polypropylene, 3.5mL, 13 x 51 mm	ThermoFisher Scientific	Cat#NC9529688
SUPERase• In™ RNase Inhibitor (20 U/μL)	ThermoFisher Scientific	Cat#AM2696
cOmplete Proteinase Inhibitor, Mini, EDTA-free	Roche Applied Science	Cat#11836170001
RNase I	Biosearch Technologies	Cat#N6901K
Trizol	ThermoFisher Scientific	Cat#15596026
Chloroform – isoamyl alcohol mixture	Sigma-Aldrich	Cat#25668
MaXtract High Density	Qiagen	Cat#129056
GlycoBlue	ThermoFisher Scientific	Cat#AM9515
32P-labeled Decade marker	ThermoFisher Scientific	Cat#AM7778
BrightStar™-Plus Positively Charged Nylon Membrane	ThermoFisher Scientific	Cat#AM10104
20 × SSC	ThermoFisher Scientific	Cat#AM9763
All enzymes used to clone libraries are in Table S2	New England Biolabs	N/A
Critical commercial assays		
QIAamp DNA blood mini kit	Qiagen	Cat#51106
PowerUp™ SYBR™ Green Master Mix for qPCR	ThermoFisher Scientific	Cat#A25742
Zymo Clean and Concentrator-5	Zymo Research	Cat#D4004
HiSpeed Plasmid Maxi Kit	Qiagen	Cat#12663
MilliporeSigma™ Novagen™ KOD Xtreme™ Hot Start DNA Polymerase	ThermoFisher Scientific	Cat#71-975-3
2x KOD Hot Start Master Mix	ThermoFisher Scientific	Cat#71-842-3
Qubit™ dsDNA HS Assay Kit - 500 assays	ThermoFisher Scientific	Cat#Q32854
MinElute PCR Purification Kit (50)	Qiagen	Cat#28004
T7 RiboMAX™ Express Large Scale RNA Production System	Promega	Cat#P1320
T7 mScript™ Standard mRNA Production System	Cellscript	Cat#C-MSC10062525

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNase-Free DNase Set (50)	Qiagen	Cat#79254
RNeasy mini column	Qiagen	Cat#81900
Lexi® Rabbit Reticulocyte Lysate System	Promega	Cat#L4540
Deposited data		
BioProject number for the raw NGS reads	This study	BioProject: PRJNA1002397
Data files	This study	https://doi.org/10.5061/dryad.x3ffbg7qx
Experimental models: Cell lines		
Human: Huh-7.5-NTCP (hepatocyte)	Laboratory of Charles M. Rice (Yu et al.[5])	N/A
Oligonucleotides		
See Table S2	N/A	N/A
Recombinant DNA		
pGEM-3Z-HBV gtA	Yu et al.[5]	GenBank: MN172185
See Table S1 for mutant versions of the above plasmid	N/A	N/A
See Table S2 for construction of DMS libraries of the above plasmid	N/A	N/A
Software and algorithms		
GraphPad Prism 9	GraphPad Software	https://www.graphpad.com/; RRID: SCR_000306
PyMOL	Schrodinger	https://pymol.org/2/; RRID: SCR_000305
SeaView2	Gouy et al.[38]	https://doua.prabi.fr/software/seaview
Alphafold2	Jumper et al.[18]	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
Altair: Python package used to generate heatmaps	Vega-Altair developers	https://altair-viz.github.io/; RRID: SCR_024752
dms_tools2 used to analyze FastQ files	Bloom[39]	https://github.com/jbloomlab/dms_tools2
Kozak similarity score algorithm	Gleason et al.[40]	https://doi.org/10.5281/zenodo.6987364
Estimating fitness effects of mutations on HBV Pol	Hopf et al.[41]	N/A
Sliding window proline analysis	This study	https://doi.org/10.5061/dryad.x3ffbg7qx

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Charles M. Rice (ricec@rockefeller.edu).

Materials availability

Cell lines and plasmids used in this study can be provided by C.M.R. pending scientific review and a completed material transfer agreement.

Data and code availability

- Deep mutational scanning data have been deposited at SRA and are publicly available as of publication. Accession numbers are listed in the key resources table.
- All code required to analyze the data is available on GitHub and has been deposited at Dryad and is publicly available as of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cells

Huh-7.5-NTCP hepatoma cells (human; sex: male, liver epithelial) were maintained at 37°C, 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Fisher Scientific, cat. #11995065) supplemented with 0.1 mM nonessential amino acids (NEAA, Fisher Scientific, cat. #11140076) and 10% hyclone fetal bovine serum (FBS, HyClone Laboratories, Lot. #AUJ35777) as previously described.[5]

Plasmids

All pgRNA mutants and DMS libraries used in this study were constructed on a genotype A plasmid backbone (GenBank: MN172185) as described previously.[5] The nucleotide sequences for the Core- (T33*), HBs- (C69*), HBx- (G27*), Pol- (active site mutant, YMDD551YMHH), C336A, VHF832GHG, AA* (PP844AA), YMHH_AA*, and C336A_YMHH_AA* mutants are listed in Table S1.

METHOD DETAILS

HBV pgRNA synthesis, transfection, DNA extraction, and quantification

5′ capped and 3′ polyadenylated HBV pgRNAs for transfection into cells were generated as previously described.[5] HBV pgRNA transfection was performed using 0.5 μg of pgRNAs, 2.5 μl of Lipofectamine 2000 (Fisher Scientific, cat. #11668019), and 250 μl of Opti-MEM Reduced Serum Medium (Fisher Scientific, cat. #51985034) in a 6-well plate containing Huh7.5-NTCPs cells seeded at a density of $2.5X10^5$ cells per well 2 days prior to transfection. Cellular DNA was extracted 2 days after transfection using QIAamp DNA blood mini kit (Qiagen cat. #51106), and realtime PCR was performed using SYBR green with primers in the Core region as previously described.[5]

DMS library construction

Mutagenesis to construct the Core and Pol DMS libraries was performed using PCR as described by Bloom.[42] Briefly, we generated a list of codon tiling primers using a Python script from the Bloom laboratory (https://github.com/jbloomlab/CodonTilingPrimers). This protocol mutates each codon to all possible codons (NNN). We ordered all mutagenic forward and reverse primers from IDT in deep-well 96-well plates. All mutagenic reverse primers were combined at an equimolar ratio and used with a common forward primer for 10 cycles using the enzyme and cycling conditions described by Bloom.[42] Similarly, all mutagenic forward primers were combined at an equimolar ratio and used with a common reverse primer. Due to the large size of the polymerase open reading frame, we constructed four libraries representing the subdomains (TP, spacer, RT, and RNase H domains) based on previously defined boundaries, and as indicated in Table S2. This ensured we could obtain sufficient library coverage for each codon. Each library was experimentally tested separately to ensure sufficient sampling in cell culture. The common forward and reverse primers used for constructing each library are listed in Table S2. The products of these two PCRs were used as input for 20 cycles of the joining reaction. The resulting PCR product was purified using Zymo Clean and Concentrator-5 (Zymo, cat. #D4004), digested with the restriction enzymes listed in Table S2, gel-extracted with agarose dissolving buffer (ADB) (Zymo, cat. #D4001-1-50), and again purified with a DNA Clean & Concentrator-5 column. This digested product was ligated into the digested genotype A backbone with T4 DNA ligase and transformed into XL1-Blue electrocompetent bacteria (Agilent, cat. #200228). For each library, bacterial colonies were grown overnight at 37°C on three 245 mm × 245 mm LB-carbenicillin plates. The following day, colonies were collected, and DNA was extracted using the Qiagen HiSpeed Plasmid Maxi Kit (cat. #12663). Each library contained more than one million transformants equal to or greater than roughly 100-fold coverage for each possible codon variant.

DMS library sequencing

HBV DNA from cell lysates was first amplified by PCR0 to obtain a near full-length HBV amplicon using primers listed in Table S2 using KOD Xtreme Hot Start DNA Polymerase (Sigma, cat. #71975-3) and the following PCR conditions: 15 ul 2x buffer, 1 ul each primer (10 uM stock), 5 ul 2mM dNTPs, 1 ul KOD Xtreme polymerase, 5 ul HBV DNA, 2 ul H2O. The PCR was performed by (i) denaturation at 94 °C for 2 min (one cycle); (ii) PCR at 98 °C for 10 s, 60°C for 30 s, and 68 °C for 3 min (iii) 68 °C for 2 min and 10 °C to hold. The number of cycles was determined based on copy number from SYBR core amplicon qPCR results using the following equation: $y = -1.864 * \ln(x) + 43.336$, where x = copy number/5 ul. The PCR products were separated in a 1% TAE agarose gel and purified with ADB (Zymo, cat. #D4001-1-50) and a Zymo Clean and Concentrator-5 column. The PCR product was eluted in 10 ul of H2O and quantified by Qubit. The purified PCR0 product was used as a template for PCR1 of the subamplicon sequencing method described in Bloom.[42] PCR1 contained 12 ul of 2x KOD Hot Start Master Mix (MilliporeSigma cat. # 718423), 2 ul fwd primer (5 uM stock), 2 ul rev primer (5 uM stock), and template DNA. For sequencing input (plasmid) libraries, we used 16 ng of library plasmid as template for PCR1. For sequencing HBV DNA, we used 4 ng of purified PCR0 product. The primers used for each subamplicon are provided in Table S2. The PCR was performed by (i) denaturation at 95°C for 2 min (one cycle); (ii) 10 cycles of PCR at 95°C for 20 s, 70°C for 1 s, 54°C for 20 s, and 70°C for 20 s (iii) one cycle 95°C for 1 min and 10°C to hold. PCR1 products were purified on Zymo Clean and Concentrator-5 column as above, quantified by Qubit and diluted to be used as a template for the barcoding and indexing PCR (PCR2). Templates were bottlenecked at 350,000 double-stranded DNA molecules for each PCR2 reaction. PCR2 contained 12 ul of 2x KOD Hot Start Master Mix (MilliporeSigma cat. # 718423), 2 ul fwd primer (5 uM stock), 2 ul rev primer (5 uM stock), and

template DNA. The PCR2 primers are provided in Table S2. The PCR was performed by (i) denaturation at 95°C for 2 min (one cycle); (ii) 24 cycles of PCR at 95°C for 20 s, 70°C for 1 s, 55°C for 20 s, and 70°C for 20 s (iii) 10°C to hold. PCR2 products were gel purified and sequenced at The Rockefeller University Genomics Resource Center using NovaSeq SP PE250.

HBV pgRNA cotransfection experiments to quantify cis vs trans activity

To quantify test *cis* vs. *trans* activity, HBV pgRNAs were mixed with mutants at ratios described in figure panels. HBV pgRNA transfection, DNA extraction, real-time PCR, and PCR0 were performed as described above. PCR0 was used as a template to amplify the region encoding the WT or Core T33* mutation or the WT or YMHH reverse transcriptase mutation, followed by PCR1 and PCR2 as described above with the following exception: we used 10 ng of template DNA for PCR1. The PCR1 product was not bottlenecked before PCR2. Instead, 10 ng of PCR1 was used as a template for PCR2. The primers used for each amplicon are provided in Table S2. The percentage of WT or mutant DNA was quantified by MiSeq Nano sequencing at the Rockefeller University Genomics Resource Center.

In vitro translation assays

DNA templates that contain the RNase H subdomains of either WT or PP*->AA* mutant were made by two rounds of PCR using genotype A HBV (GenBank: MN172185) plasmid as template. The first round of the PCR was to add a 5' UTR and initiation methionine codon to the WT or mutant RNase H domains and additional termination codons at the 3' end. The second round of PCR added a T7 promoter at the 5' end for in vitro transcription. The VHF->GHG mutant was made with two rounds of PCR using the VHF->GHG mutant plasmid listed in Table S1. The primers used to make the above constructs are listed in Table S2.

RNAs for in vitro translation assays were transcribed, 5' capped, 3' polyadenylated, and purified as above for HBV pgRNA, with the following exception: 1 μg of gel-isolated PCR product was used as a template transcription. Wildtype or mutant RNAs (0.1 μg) were translated using the Lexi® Rabbit Reticulocyte Lysate System (Promega, cat. # L4540) in a 20 μl reaction volume supplemented with 25 mM KCl, 1.25 mM MgOAc, and 0.2 mCi/ml [^{35}S]-methionine (PerkinElmer, cat. #NEG709A500UC). The RNAs were preheated at 65°C for 3 min, chilled on ice, then incubated in the lysate at 30°C for the specific time intervals described in the figure panels. For RNase A treatment, the 20 μl reactions were split into two tubes and treated with either RNaseOUT (2 U/μl) (Fisher Scientific, cat. #10777019) or RNase A (0.5 μg/μl) (Fisher Scientific, cat. #EN0531) for 5 min at 30°C. The reactions were stopped by adding NuPAGE LDS sample buffer (Fisher Scientific, cat. #NP007) supplemented with 50 mM DTT to each reaction. Samples were heated at 70°C for 10 min, and 10 μl of the reaction was analyzed using a NuPAGE 4-12% gel (Fisher Scientific, cat. #NP0321BOX) with NuPAGE™ MES SDS running buffer (Fisher Scientific, cat. #NP0060). Electrophoresis was stopped when the bromophenol blue dye ran out of gel. The gel was then fixed by adding a solution containing 30% methanol and 10% acetic acid, dried, and visualized using a phosphorimager cassette (VWR, cat. #29-1755-24) and scanned on a Typhoon instrument.

Ribosome purification by sucrose cushion

A 50 μl volume of in vitro-translated products was loaded on top of a 3.2 ml sucrose cushion (1M sucrose, 20 mM Tris (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 μg/ml cycloheximide, and 20 U/ml Superase.In) in a 13x51 mm polycarbonate tube, and ribosomal complexes were pelleted by centrifugation in a TLA100.3 rotor at 70,000 rpm (265,000 x g) at 4°C for 4 h. To prevent contaminating the pellet with ribonucleoprotein (RNP) from the top of the cushion, the top 500 μl solution was immediately removed after centrifugation, and 500 μl of Tris (10 mM, pH 8.0) was added. This procedure was repeated a second time before removing all the supernatant. Pellets were dissolved in 50 μl of Buffer A (20 mM Tris (pH 7.5), 100 mM NaCl, 2.5 mM $MgCl_2$) and processed the same as samples without a sucrose cushion.

RNase I digestion followed by northern blot

A 60 μl volume of in vitro-translated products was diluted with 140 μl Buffer B (20 mM Tris, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 100 μg/ml cycloheximide, 1X proteinase.In). Additionally, 15 U of RNase I (Biosearch Technologies, cat. #N6901K) were added, and the mixture was incubated at 25°C for 45 mins with shaking at 750 rpm. Reactions were stopped by adding 800 μl of Trizol (Fisher Scientific, cat. # 15596026), mixed with 200 μl of chloroform/isoamyl alcohol (Sigma, cat. #25668), and phase separation was enhanced with a phase lock xtract (Qiagen, cat. # 129046). After spinning at 13,500 x g at 4°C for 10 min, approximately 600 μl of supernatant was transferred to a new tube and mixed with 1.5 μl GlycoBlue (Fisher Scientific, cat. # AM9515), 60 μl NaAc (3 M), 750 μl isopropanol. The mixture was stored at -80°C for at least 1 h or at -20°C overnight. The pellet was obtained by centrifugation at 18,500 x g at 4°C for 20 min, washed twice with 70% ethanol, and dissolved in 8 μl 1x RNA loading dye (5mM EDTA, 49% formamide, 0.0125% Bromophenol Blue, 0.0125% Xylene Cyanol).

Northern blots were performed as previously described.[43] Briefly, 50 ng of RNA from each digestion was mixed at equal volume with 2x loading dye (10 mM EDTA, 98% formamide, 0.025% Bromophenol Blue, 0.025% Xylene Cyanol) before being run in a denaturing polyacrylamide gel containing 15% acrylamide and 7M urea. Samples were run alongside a ^{32}P-labeled Decade marker (Fisher Scientific, cat# AM7778) prepared according to the manufacturer's instructions. RNA was transferred to a BrightStar™-Plus Positively Charged Nylon Membrane (Fisher Scientific, cat# AM10104) and then UV cross-linked twice at 150 mJoule.

Membranes were incubated in blocking buffer (6× SSC (diluted from 20 × SSC (Fisher Scientific, cat# AM9763)) and 7% SDS) for at least 30 min at 65°C. Hybridization of a ^{32}P end-labeled DNA oligonucleotide probe was performed overnight at 42°C in blocking

buffer. Blots were washed three times for 15 min in 3× SSC and 0.1% SDS and exposed to film at -80°C. Blots were stripped at 65°C overnight in blocking buffer, then washed 3 times for 15 min each in 3× SSC and 0.1% SDS before proceeding with an additional probe. Probe sequences are listed in Table S2.

AlphaFold 2 structure prediction

Pol structures of HBV and WSHBV were generated using full-length sequences of a genotype A HBV Pol (GenBank: MN172185) and a WSHBV Pol (GenBank: KR229754), respectively, utilizing AlphaFold2).[18] Models with the best pLDD scores were used for further analysis. Structures of DHBV, RNDV, SkHBV, and TMDV were downloaded from previous work.[19] The structure of HIV (PDB: 6BSG) was downloaded from the PDB website (https://www.rcsb.org). All PDB files used in this study are included in the supplement. The RNase H domains from the above structures were aligned using PyMOL with its built-in alignment module.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

The raw sequencing data are available through the NCBI sequence read archive (SRA) under BioProject: PRJNA1002397. Data files and a Jupyter Notebook (JNote.ipynb) containing detailed code to generate some of the graphs in the paper are provided on GitHub (https://github.com/HBV-DMS/R1/). CSV files with numeric values for the figures are at https://github.com/HBV-DMS/R1/tree/main/Fig/CSV. Additionally, these files are archived in Dryad (https://doi.org/10.5061/dryad.x3ffbg7qx). Fastq files were analyzed using dms2tools developed by the Bloom laboratory. See here for a general introduction: https://github.com/jbloomlab/dms_tools2. The bcsubamp tool was used to correct sequencing errors and quantify variants. The resulting variant counts (https://github.com/HBV-DMS/repo/tree/main/data/codoncounts) were inputted in JNote to calculate the enrichment ratio (ER).[39] To account for DNA damage in the library preparation, we applied an oxidative damage filter to eliminate data from codons one G-to-T or C-to-A transversion away from WT from the codon-counts files. The Python package Altair was used to generate heatmaps, which translated $\log_2$ ER into a color scale for the heatmap squares. In the heatmaps, the square size was reduced to account for low mutagenesis quality based on the variant read count of the plasmid samples per 1E6 reads. The square size was reduced for codons with 15 or fewer read counts per 1E6 reads. To determine the Kozak scores, we used a Kozak similarity score algorithm described by Gleason et al.[40] Analysis of certain groups of variants to determine statistical significance was done using GraphPad Prism 9 and applying the Mann-Whitney U test. These results can be found in the corresponding figure legends.

Fitness effect prediction from natural sequences

We used methods outlined by Hopf et al.[41] to estimate the fitness effect of mutations on the HBV polymerase. We used the HBV polymerase from GenBank: MN172185 as the query sequence and searched against the Uniref100 dataset[15] downloaded on March 16, 2022 with bitscore 0.9. Alignments were processed to remove sequences that aligned to less than 50% of our query sequence and positions with more than 30% gap characters. All parameters needed to reproduce the model are included in the files, ''hbv_pol_model_parameter_details.txt'' and ''proline_sliding_window_final.ipynb'' located on GitHub (https://github.com/HBV-DMS/R1/) and archived in Dryad (https://doi.org/10.5061/dryad.x3ffbg7qx).

RT and RNase H domain alignments

We extracted the RT and RNase H domain sequences from all sequences used in Lauber et al.[44] and included RT and RNase H domains from additional retroviruses (Table S3). We aligned amino acid sequences using SeaView2[38] using Muscle with default parameters. The resulting alignment is in Data S1. A phylogenic tree was created using the SeaView2 parsimony analysis with default parameters.

Sliding window proline analysis

To investigate the frequency of proline across the protein, we performed a sliding window analysis using the fasta file from Data S1. The window size was fixed at 25 amino acids, and the proline counts for each RNase H domain were grouped according to the categories listed in Figure S5C.

Supplemental figures

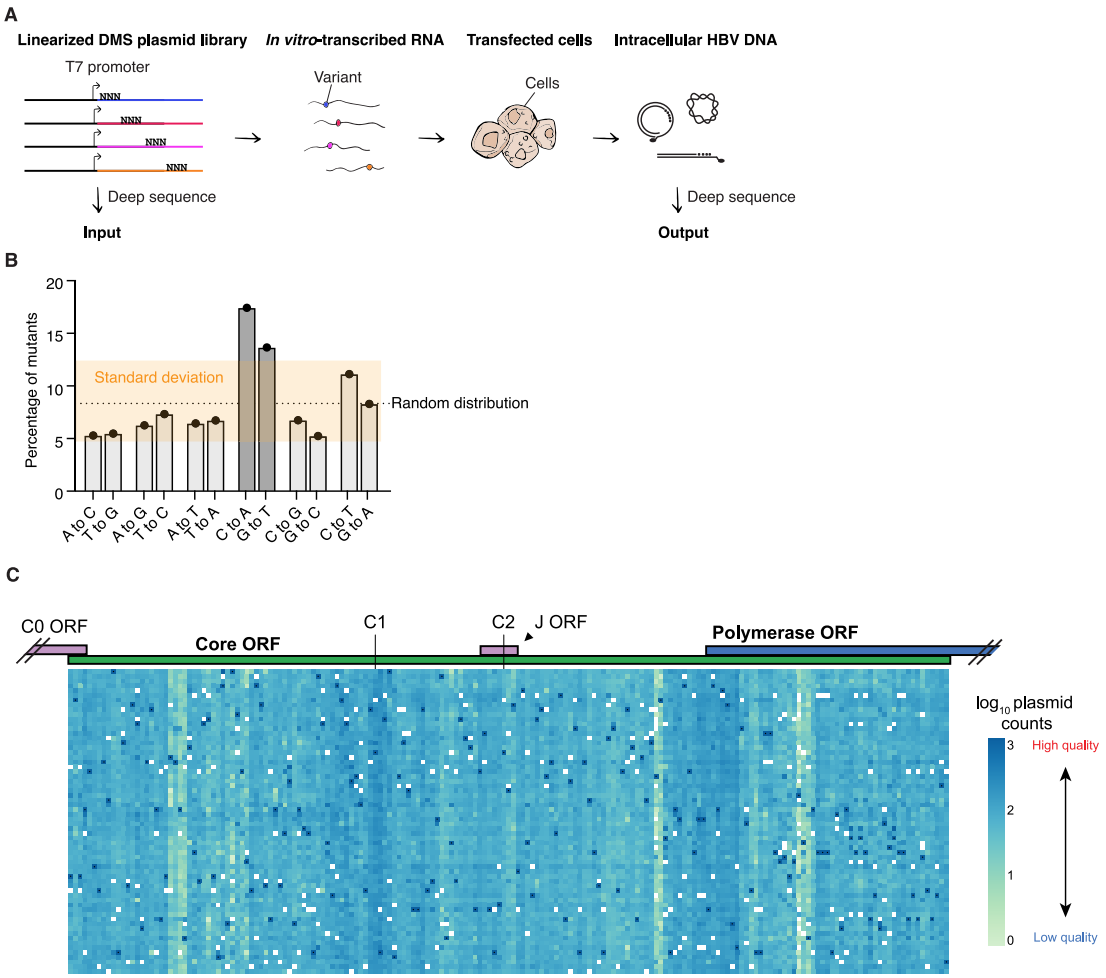


Figure S1. Deep mutational scanning approach, data filtering, and data quality, related to Figure 2

(A) Schematic workflow of the DMS approach. Plasmid libraries containing codon variants were constructed by PCR following a previously published method[42] and as indicated in the STAR Methods section. A plasmid library containing randomized codons is then linearized and used as a template for *in vitro* transcription with T7 RNA polymerase. The resulting library of pgRNA variants is transfected into Huh-7.5-NTCP cells, and intracellular DNA is purified 2 days later. Various forms of HBV DNA, including relaxed circular, covalently closed circular, and double-strand linear DNA are represented. The plasmid and purified intracellular DNA are used as templates for PCR to prepare sequencing libraries, as described in the STAR Methods section.

(B) Percentage of all single-nucleotide mutations recovered from transfected cells. C to A and G to T mutants were overrepresented and identified as outliers using the robust regression and outlier removal (ROUT) method (Q = 1%) in GraphPad Prism 9. The standard deviation is indicated by orange shading.

(C) The color of heatmap squares correlates to the $\log_{10}$ counts of the input plasmid library per 1E6 reads, indicating data quality. Codon substitutions are depicted on the y axis. Site positions are portrayed on the x axis. White squares are variants one G-to-T or C-to-A transversion away from WT codons likely affected by oxidative damage during sequencing library preparation. An interactive version of the heatmap in (C) is at https://hbv-dms.github.io/R1/Fig/Fig_S1C.html.

See also Figure 2 and Table S2.

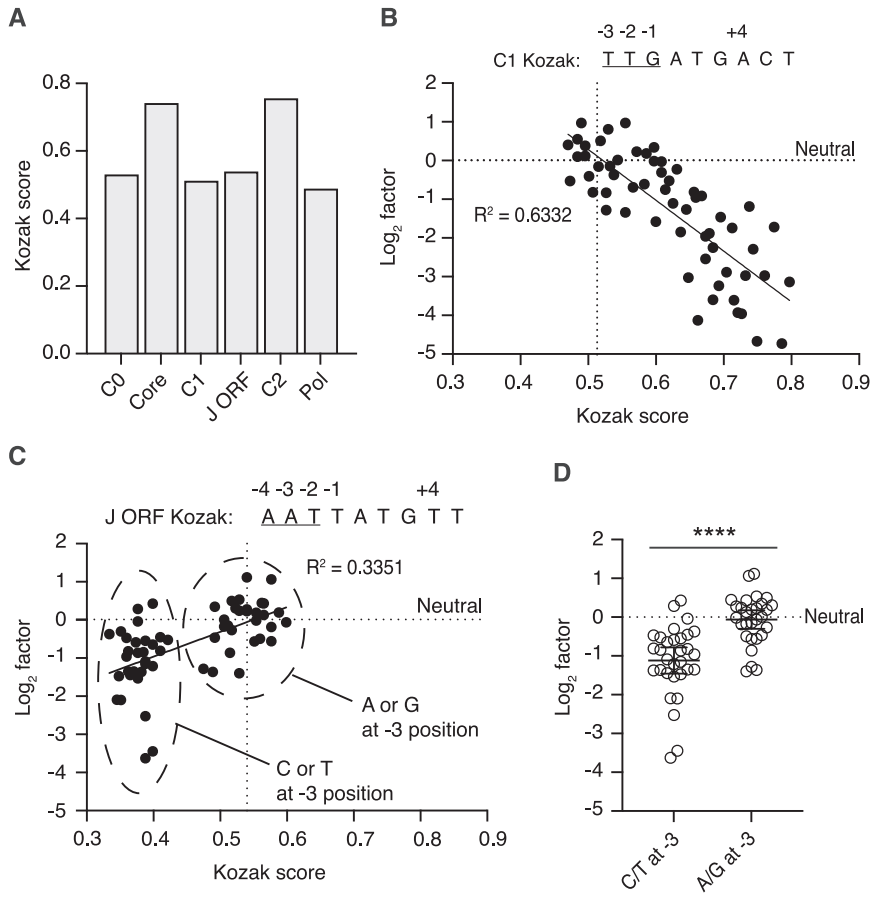


Figure S2. Kozak analysis, related to Figure 2

(A) Predicted Kozak scores for all ATG codons in pgRNA up to the Pol start codon.

(B) Predicted Kozak score for all codon variants upstream of C1 plotted with their $\log_2$ fold change in abundance relative to input. WT C1 Kozak sequence (top). The R^2 value for the line is indicated. The vertical dotted line denotes the WT C1 Kozak score. The horizontal dotted line denotes no change in variant abundance (neutral).

(C) Similar to (B), but for the J ORF start codon. The position of the data on the graph tracked with the nucleotide at the −3 position, as indicated by dashed circles.

(D) Variants encircled in (C) were grouped and plotted for statistical comparison. Data are mean with 95% confidence intervals. ****$p < 0.0001$ by Mann-Whitney U test.

See also Figure 2.

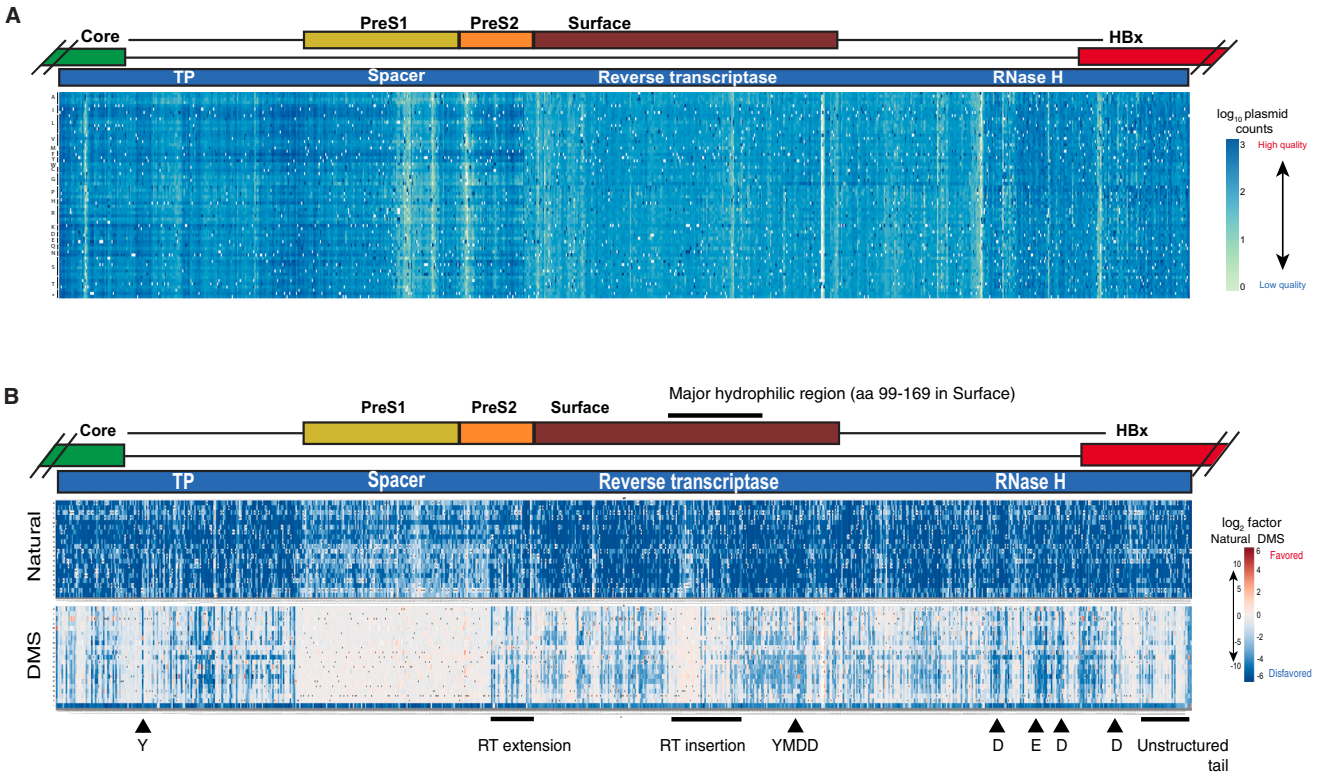
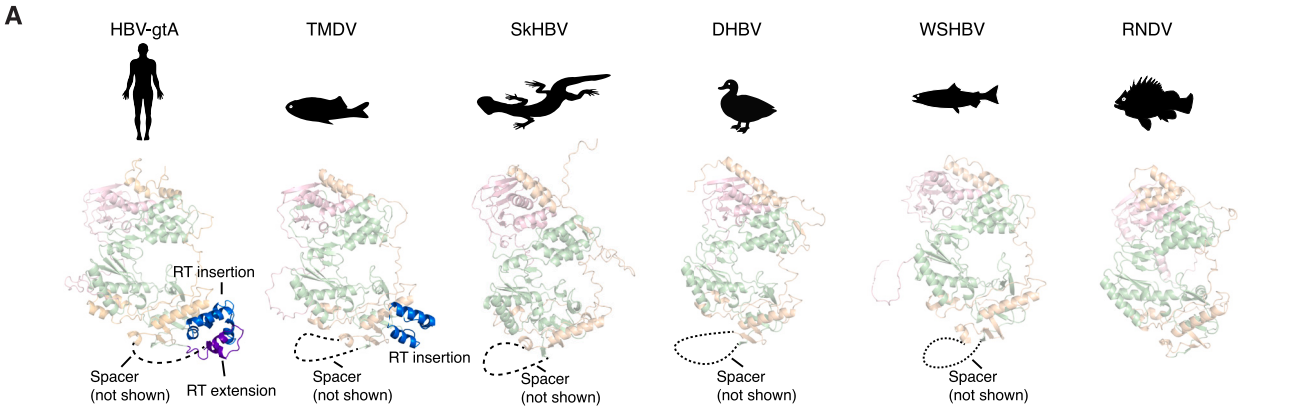
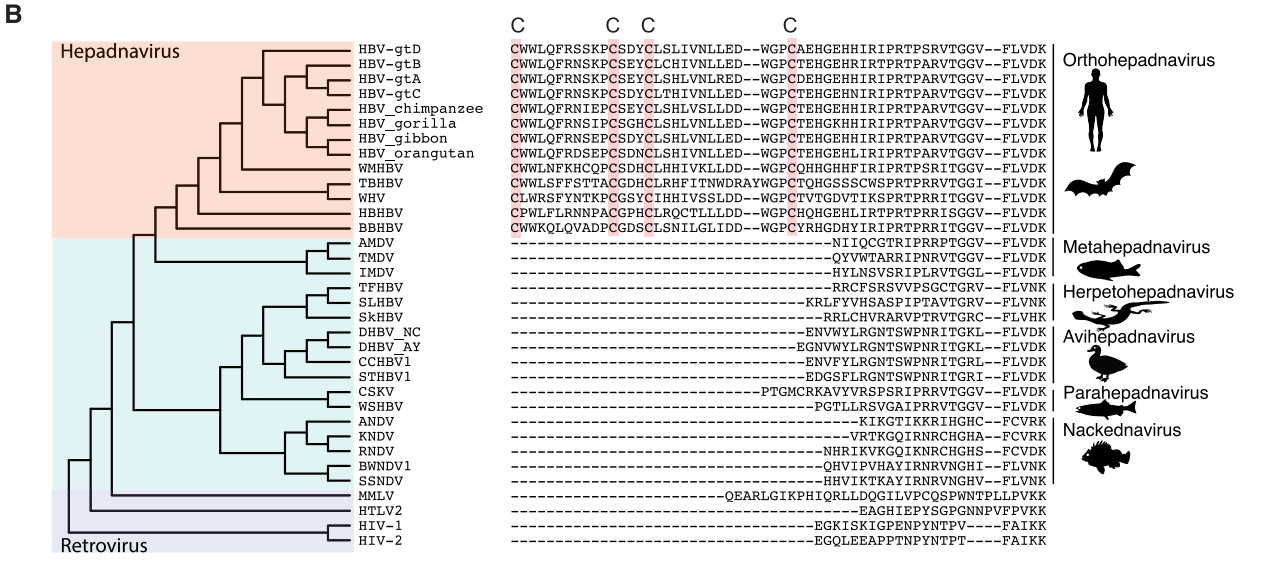


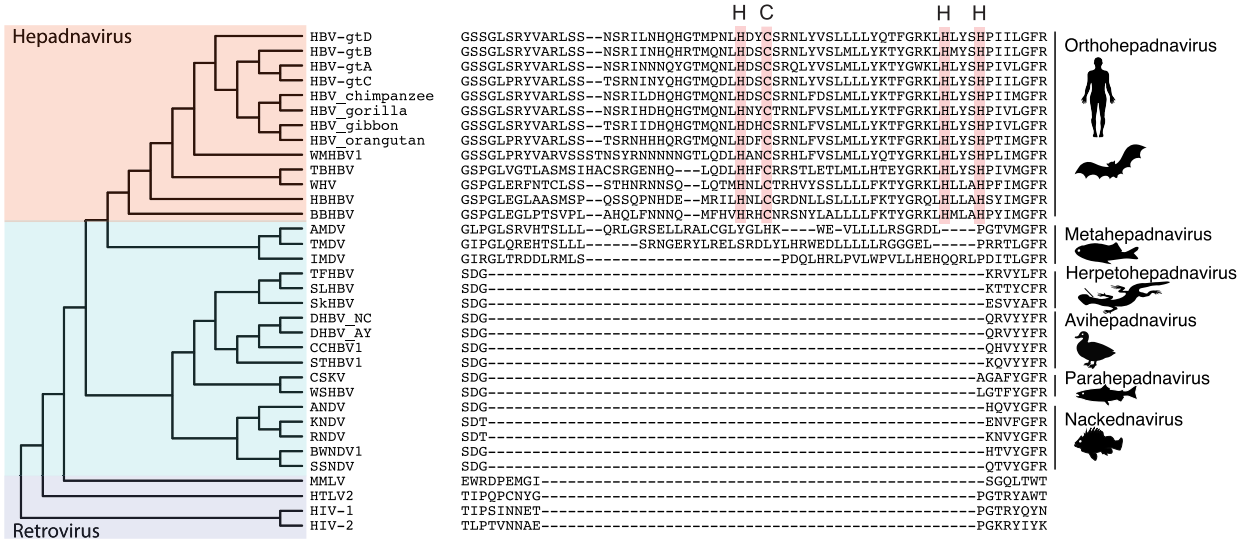
Figure S3. Data quality heatmap of Pol DMS and fitness of polymerase variants compared with natural sequences, related to Figure 3

(A) The heatmap squares correlate to the $\log_{10}$ counts of the input plasmid library per 1E6 reads, indicating data quality. Substitutions are depicted on the y axis. Site positions are portrayed on the x axis. White squares are variants one G-to-T or C-to-A transversion away from the WT codon likely affected by oxidative damage during sequencing library preparation.

(B) Heatmaps depicting the fitness of polymerase variants (amino acids) obtained from Uniref100 (natural, top) and experimental results (DMS, bottom). Overlapping reading frames are shown as colored boxes (top). Catalytic site amino acids (triangles) and regions of interest (lines) are indicated above and below the heatmap. Interactive versions of the heatmaps are at https://hbv-dms.github.io/R1/Fig/Fig_S3A.html and https://hbv-dms.github.io/R1/Fig/Fig_S3B.html. See also Figure 3.







(legend on next page)

Figure S4. Conservation of RT extension and insertion among the *hepadnaviridae*, related to Figure 3

(A) AlphaFold 2 predictions of the polymerase proteins for representative hepadnaviruses. TP, RT, and RNase H domains are colored similarly across the different proteins. The RT extension and insertion, if present, are indicated. The unstructured spacer domain is depicted as a dashed line.

(B) Amino acid sequence of the RT N-terminal extension for representative hepadnavirus and retrovirus polymerases. Domain boundaries were defined by the AlphaFold 2 predictions. The evolutionary relationship for the representative viruses is depicted as a dendrogram (left). Cysteine residues in the predicted zinc fingers are highlighted (right).

(C) Amino acid sequence of the RT insertion for representative hepadnavirus and retrovirus polymerases. The evolutionary relationship for the representative viruses is depicted as a dendrogram (left). Cysteine and histidine residues in the predicted zinc fingers are highlighted (right).

See also Figure 3 and Table S3.

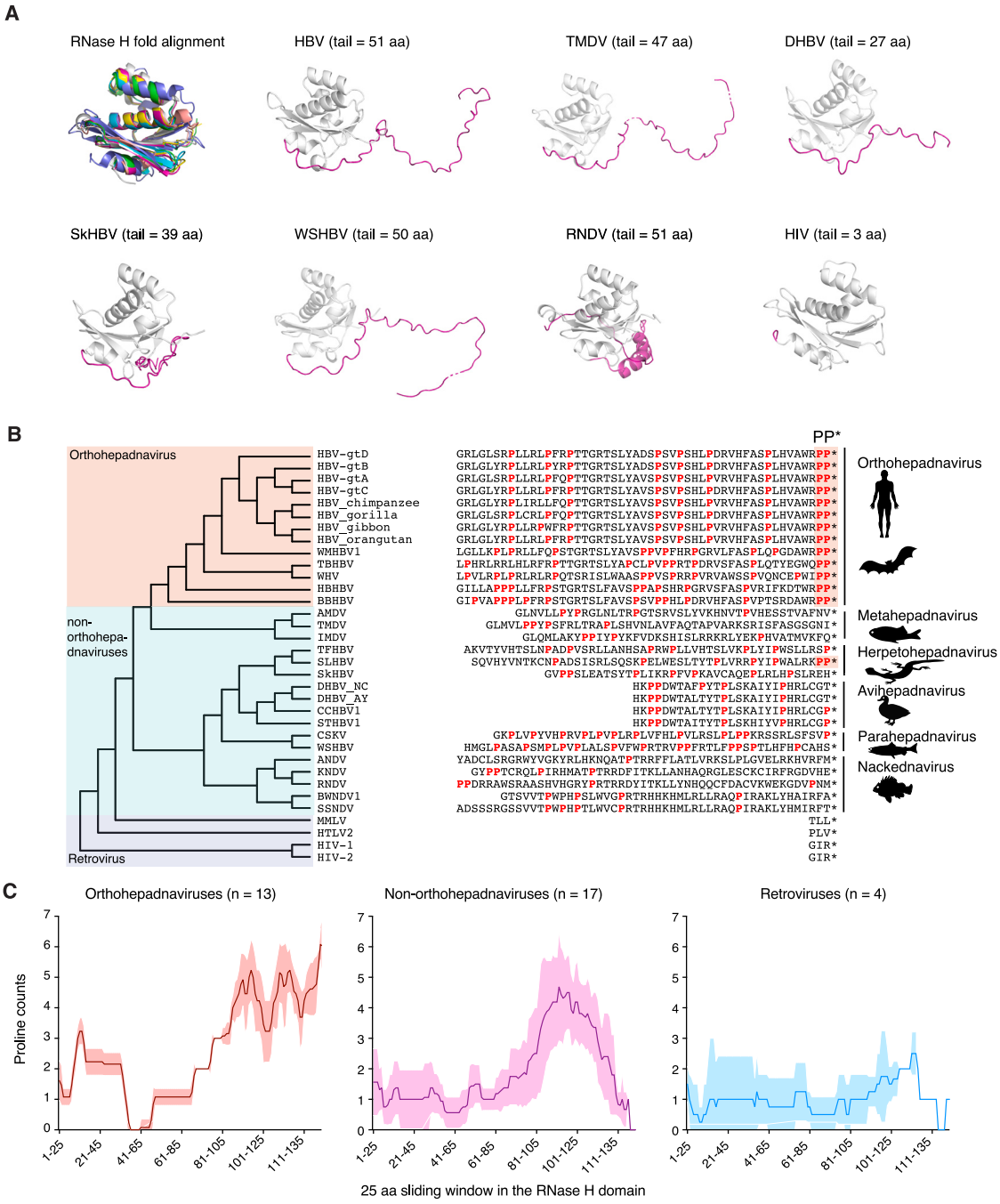


Figure S5. Ribosome-pausing motif conservation, related to Figure 4

(A) AlphaFold 2 prediction of RNase H domains for representative hepadnaviruses and the retrovirus, HIV. The classical RNase H fold for each structure overlays well (top left), whereas the downstream residues (colored pink) are predicted to be unstructured.

(B) Amino acid sequences of the unstructured C-terminal tails for representative hepadnavirus and retrovirus polymerases. The evolutionary relationship for the representative viruses adapted from Lauber et al.[44] is depicted as a dendrogram (left). Unstructured tail sequences are aligned to the termination codon, and proline residues are highlighted in red (right).

(C) Twenty-five amino acid sliding window of proline density is depicted for the RNase H domains of the viruses from (B). Solid line represents the mean, and shading represents standard deviation.

See also Figure 4 and Table S3.

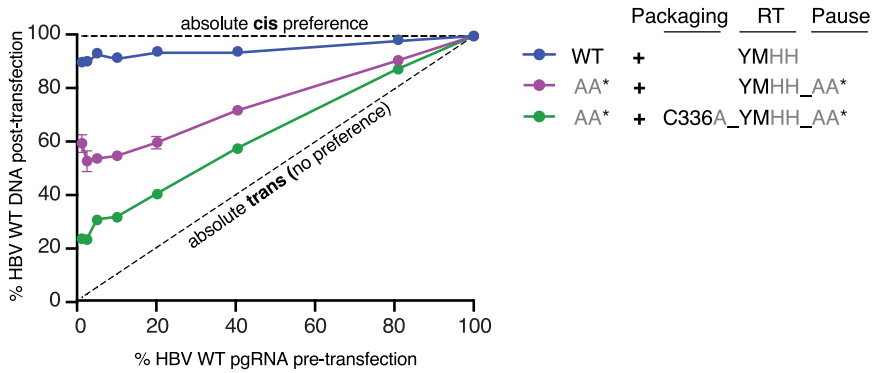


Figure S6. HBV polymerase *cis*-preference profile, related to Figure 5

WT and mutant pgRNAs were co-transfected at the ratios indicated on the x axis, and the percentage of WT HBV DNA recovered is plotted on the y axis. Each point represents the mean of three replicates ± SEM. The data for WT + YMHH (blue line) is reproduced from Figure 1D for reference.
See also Figure 5.