Encapsidated hepatitis B virus reverse transcriptase is poised on an ordered RNA lattice

Joseph Che-Yen Wanga, David G. Nickensa, Thomas B. Lentzb, Daniel D. Loebb, and Adam Zlotnicka,1

*Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, IN 47405; and bMcArdle Laboratory for Cancer Research, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706

Assembly of a hepatitis B virus (HBV) virion begins with the formation of an RNA-filled core composed of a symmetrical capsid (built of core protein), viral pregenomic RNA, and viral reverse transcriptase. To generate the circular dsDNA genome of HBV, reverse transcription requires multiple template switches within the confines of the capsid. To date, most anti-HBV therapeutics target this reverse transcription process. The detailed molecular mechanisms of this crucial process are poorly understood because of the lack of structural information. We hypothesized that capsid, RNA, and viral reverse transcriptase would need a precise geometric organization to accomplish reverse transcription. Here we present the asymmetric structure of authentic RNA-filled cores, determined to 14.5-Å resolution from cryo-EM data. Capsid and RNA are concentric. On the interior of the RNA, we see a distinct donut-like density, assigned to viral reverse transcriptase, which pins the viral pregenomic RNA to the capsid inner surface. The observation of a unique ordered structure inside the core suggests that assembly and the first steps of reverse transcription follow a single, determinate pathway and strongly suggests that all subsequent steps in DNA synthesis do as well.

Significance

Hepatitis B virus (HBV) is one of the most serious human pathogens: 350 million people suffer from chronic HBV, and 600,000 die from it annually. Its 3,200-base pair genome marks it as one of the smallest pathogens (1). HBV is an enveloped double-stranded DNA virus that is reverse transcribed from a RNA pregenome (pgRNA). To date, most anti-HBV therapeutics target polymerase (P) (2). However, the structural interplay between the components of the reverse transcription complex has not been described.

In an infected cell, HBV cores are the metabolic compartments for reverse transcription; DNA-filled cores go on to become virions. The basic building block of the capsid, the protein shell of the core, is a dimer of the 183-residue core protein (Cp). Cp is highly conserved among different genotypes (3). The N-terminal 149 amino acids of the Cp form the assembly domain, which can self-assemble into an icosahedral capsid (4). In vitro and in vivo, Cp dimers self-assemble mainly into \( T = 4 \) particles (120 dimers), with a small fraction of smaller \( T = 3 \) particles (5, 6). Capsids are dynamic: they are held together by very weak Cp-Cp interactions and are subject to breathing modes involving partial unfolding of Cp (7, 8). The 34-amino acid C-terminal domain (CTD), rich in arginine residues, is essential for packaging of the pgRNA (9). The CTD is subject to phosphorylation, which is required for RNA packaging and plays a role in reorganizing packaged nucleic acid and the solution behavior of the core (10–16).

Synthesis of the gapped, circular DNA genome of mature HBV from the linear pgRNA template takes place inside the core, as it resides in the host cytoplasm (17) (Fig. 1). Core formation begins with P protein binding to a stem loop, s, near the 5′ end of the pgRNA (18). This complex is packaged by phosphorylated Cp to yield the immature, RNA-filled core (15, 19–21). Complementarity between the 5′ and 3′ ends of the RNA required for subsequent reverse transcription suggests that P and both ends of the RNA form a compact complex (22). Within the immature core, tyrosine 63 of P protein’s terminal domain (TP) primes reverse transcription (17, 23). P protein remains covalently bound to the 5′ end of the nascent minus strand, although P will switch templates three times to complete dsDNA synthesis (Fig. 1) (24, 25). Mature DNA-filled cores contain a complete minus strand with a covalently attached P protein and a partial plus strand. This DNA product is the relaxed circular DNA genome of infectious virions. DNA-filled cores are either transported to the nucleus or acquire an envelope and leave the cell (26).

We hypothesized that for successful reverse transcription, the 5′ and 3′ ends of the pgRNA, the P protein, and the capsid must form a uniform quaternary structure. To test our hypothesis, we have determined a single particle image reconstruction of authentic RNA-filled cores. We have introduced asymmetric reconstruction to obtain the unique structure of P inside the icosahedral capsid and identified a putative P protein by its donut-like density, which was consistent with structures of polymerase superfamily members, including HIV RT and HCV polymerase. The P protein was located at a unique site within the core, touching pgRNA where the pgRNA also contacted the inner surface of the core, near an icosahedral threefold axis. These ordered, asymmetric features suggest that assembly and the first steps of reverse transcription follow a single, determinate pathway.

Results and Discussion

We reengineered an HBV expression system to isolate the uniformity of RNA-filled cores (27). The modified system has three key features: P protein was mutated (Y63F) so that it packages asymmetric reconstruction | H BV intermediate

Hepatitis B virus (HBV) is a double-stranded DNA virus that packages a single-stranded RNA pregenome (pgRNA). The linear pgRNA is reverse transcribed to a gapped circular dsDNA within the confines of the virus capsid. We hypothesized that a specific capsid-RNA-reverse transcriptase structure would be required to accomplish this task. In this article, we report the structure of the authentic pgRNA-filled HBV core as determined by cryo-EM and asymmetric 3D reconstruction. The observed ordered structure suggests the assembly process and the first steps of reverse transcription follow a single, determinate pathway.


Conflict of interest statement: A.Z. has an interest in a biotechnology startup.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321424111/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1321424111

PNAS | August 5, 2014 | vol. 111 | no. 31 | 11329–11334
pgRNA but cannot prime DNA synthesis because of the replacement of the reactive tyrosine (28); a 4-nucleotide substitution eliminated the primary splice acceptor site in the HBV pgRNA (the majority of pgRNA produced from this mutant is full length) (29); and translation of HBV surface proteins was inactivated, eliminating a potentially complicating set of paths for immature virions. We validated that particles purified from our expression system had hallmarks of authentic HBV cores. SDS/PAGE showed a single dominant protein (Fig. 2 and Fig. S1), identified as Cp by Western blot; proteins other than Cp were detectable by silver stain. One expects 240 copies of Cp and only one copy of P per capsid. In this fraction, the pgRNA:capsid molar ratio was 1.00 ± 0.14.

Cryo-EM of the highly purified samples showed the expected 36-nm-diameter T = 4 HBV capsids (Fig. 2D). The small number (∼5%) of 32-nm T = 3 particles were not examined further. When images were translationally averaged, an inner ring of RNA was evident that was stronger than the outer protein density (Fig. 2D, Inset), similar to pgRNA-filled capsids assembled in vitro (13). A de novo-built icosahedrally symmetrized image reconstruction of 13,575 capsids, calculated to 10.2 Å resolution, showed a typical T = 4 configuration of the HBV particle (Fig. 3A). A similar result was also obtained when using low-pass filtered atomic structure (PDB ID code 1QGT) as an initial model (Fig. S2). The inner layer of density had characteristics of RNA seen in capsids of unphosphorylated and phosphomimic Cp (13), and RNA was particularly strong underneath fivefold and quasi-sixfold arrays of Cp. However, the lumen of the symmetrized capsid had weak density not found in in vitro assembled particles (Fig. 3B and C) (13).

The major challenge to identifying unique features in an icosahedral complex is that the symmetrical protein shell leads to misalignments that average and obscure asymmetric density. A prediction of our previously stated hypothesis is that P protein, probably present as a single copy (30), should be evident at the lower resolution, even in an averaged map, as a weak but discrete density. The key to identifying the correct orientation for

---

**Fig. 1.** Reverse transcription in HBV. (A) Plus-sense pgRNA (green) is transcribed from covalently closed circular DNA (cccDNA) (red: plus-strand; black: minus-strand). Direct repeats 1 and 2 (DR1, DR2) ε, ϕ, and ω are cis-acting sequences involved in genome replication. (B) Initiation of minus-strand DNA synthesis and the first template switch. Cp packages a complex of P protein (blue circle) bound to the encapsidation signal, ε. P primes and synthesizes the first 3 nucleotides of minus-strand DNA, using ε as template. The P-minus strand complex switches template to the 3′ DR1 acceptor site. Interaction of ε with ϕ and ω facilitate switching. (C and D) Minus-strand DNA elongation proceeds with concomitant pgRNA degradation. (E) Minus-strand elongation proceeds to the 5′ end of pgRNA. RNase H leaves a 17-nt RNA from the 5′-end of pgRNA. This nascent primer dissociates from the 3′ end of the minus strand and anneals to the complementary DR2c site. (F) Plus-strand synthesis proceeds to the 5′ end of the minus-strand template. (G) Circularization and the third template switch are accomplished when the 3′ end of plus-strand switches to the 3′ end of minus-strand DNA. (H and I) Elongation and completion of plus-strand DNA synthesis yields a relaxed-circular DNA genome. Cores containing relaxed-circular DNA may be transported to the nucleus or acquire an envelope and leave the cell. Steps B–I occur in the confines of the capsid. We have determined the structure of the intermediate in B.

**Fig. 2.** Purified cores are uniform. (A) Size exclusion chromatography, the last step in purification, shows a single major peak. (B) The peak is dominated by Cp, confirmed by immunoblot (Fig. S1). (C) The ratios of 260 nm to 280 nm absorbance and pgRNA to capsid indicate cores have uniform RNA content. These are shown as squares and triangles, respectively, in A. (D) Cryo-EM shows mainly 36 nm T = 4 particles, with a small number of 32 nm T = 3 particles. A translational average of T = 4 particles (inset) confirmed the presence of an inner concentric layer, assigned as RNA.
asymmetric reconstruction was to choose a feature to use as a fiducial marker, which was a trial-and-error process. We finally proceeded with a lower-resolution search model that started with a clump of density located near the icosahedral threefold axis (see Fig. S3 and Materials and Methods for details). This volume was set to the density level of the protein shell, and the rest of equivalent copies were masked out. The resulting asymmetric model was compared with each image to identify the best of the 60 icosahedrally equivalent orientations previously determined for each image during calculation of the symmetrical structure. The best orientation for all particles had an average cross-correlation of 2.1 σ above the mean calculated for all icosahedrally equivalent orientations. When the internal clump of density was moved to different locations to simulate alternative initial asymmetric models as control experiments, the average cross-correlation of the “best orientation” in all cases dropped to ~1.9 σ (Fig. S4). In these structures, internal density was noisy and recapitulated input features, suggesting the wrong starting models or orientations. After 50 iterations of refinement, the resulting 14.5-Å resolution map had a symmetric exterior and internal density that was clearly asymmetric (Fig. 3D and E and Fig. S5). For our final model of 11,727 images, the average cross-correlation of the “best orientation” was improved to 2.3 σ (Fig. 3F), a correlation between the asymmetric model and image significantly above the level expected for 60 random orientations. The model we identified as the basis of our reconstructions led to a structure with unique features and was significantly better than the alternatives.

The exterior of the asymmetric reconstruction resembled symmetrized T = 4 structures (Figs. 3D and 4A and Fig. S4A). There were no systematic differences between a molecular capsid model [IQGT (31)] and our asymmetric density (Fig. S6B and Movie S1). Diameters measured from equivalent points confirm the visually apparent symmetry. Thus, small differences in spike shape reflect the low noise level of this asymmetric structure.

In the asymmetric reconstruction, the pgRNA layer of the core has symmetric and asymmetric character (Fig. 4B and C and Fig. S6). RNA density around fivelfold and sixfold vertices correlates with the position of the RNA-binding CTD of Cp (Fig. S7), a stretch of 34 amino acids that includes 17 arginines (13). However, the connectivity between vertices is not regular (Fig. 4C), suggesting a preferred organization for the RNA.

The innermost layer of density in this reconstruction was unambiguously asymmetric (Fig. 3E and 4B) and was evident in the reference-free 2D class averages (Fig. S8A). A number of distinct masses were observed, all clustered in one hemisphere. In a control experiment, the internal density disappeared when the model was used to refine orientations of in vitro-reassembled pgRNA-filled C183 capsids (lacking P and host factors; Fig. S9). The strength of their electron density was similar to that of the RNA and of the capsid; therefore, they were not likely to represent low-occupancy proteins. The largest mass was an elliptical donut that was ~75 Å in diameter at its widest. The donut touched the RNA layer, where the RNA layer also contacted a Cp dimer near an icosahedral threefold axis of symmetry. Although colored red, gold, and gray in the figure, at this resolution it is not possible to specifically state exactly where one component ends and another begins. Nevertheless, the Fourier shell correlation at the radii corresponding to the donut density (60–90 Å) indicated a resolution of 15.4 Å, which is only slightly worse than the resolution estimated for the whole structure (Fig. S8B).

The well-resolved portion of the donut was consistent with the curved shape of polymerases: the polymerase domains of HIV reverse transcriptase and HCV RNA-dependent RNA polymerase fit neatly into density (32, 33). HBV P protein is monomeric but larger, at 95 kDa, than most reverse transcriptases. The extra mass is localized to the N-terminal TP and spacer domains. These and the C-terminal RNaseH would occupy a larger fraction of the donut. We caution that this identification is only tentative. Interestingly, when the polymerase domain of the p66 subunit of HIV RT is fit to the donut, the RNaseH domain hangs out of density. Efforts to work with P from HBV and from the related duck virus have shown that the protein in solution requires chaperones for solubility. However, the
The most important conclusion we draw from this study is that the asymmetric interior of the pgRNA-containing HBV core is uniform from particle to particle. RNA and non-RNA density was well-ordered. Therefore, we suggest that 5' and 3' ends of the RNA, critical for reverse transcription, are structurally identical from capsid to capsid. This may be accomplished during assembly if the P-pgRNA complex acts to nucleate assembly, accumulating the first few subunits of the growing capsid, which direct subsequent growth and symmetry. It is not possible to unambiguously model a single path for the whole RNA, consistent with observation that genomes with large deletions, which retained 5' and 3' ends, were still able to support reverse transcription (43). The density internal to the RNA shows more order than the RNA itself.

Reverse transcription (Fig. 1) would be impossibly complex if the ends of the nucleic acid had random locations; if the ends can conform to one structure, as if the capsid were a molecular jig, the template switches are much easier to envision. Because of the organization of the RNA and the presence of the P complex inside the RNA shell, maximally interacting with the capsid, we propose that the P complex is mobile, progressively traveling along the viral nucleic acid, similar to a train on a track. [With current data, we cannot preclude a reovirus-like (44) model of P function, where P is static. However, the train track model allows the ends of the genome to be organized and does not require retaining large sequences of nucleic acid from the very basic capsid surface so they can be threaded through a replication machine.] In its travels, the P complex would also interact with Cp, so that bound kinases and phosphatases could sequentially modify Cp phosphorylation, a progressive change observed during duck HBV replication (11, 12). Indeed, the capsid has been implicated as an active participant in reverse transcription (43, 45). Following the nucleic acid track, the P complex would return to the same location to perform the template switches required for dsDNA synthesis. This “train on a track” model requires a nonoverlapping Hamiltonian path proposed to be general to ssRNA viruses (46, 47); such paths also imply a deterministic capsid assembly path.

Materials and Methods

Molecular Clones and Plasmids. Molecular clones of HBV used in this work have been deposited in the GenBank database (accession no. V01460, subtype ayw, genotype D). The plasmid 1159 is a derivative of the HBV expression plasmid TL7 (27), which contains the 7xTetO-CMVp-HBV sequence and the oriP sequence of the Epstein Barr virus. The plasmid TL7 virus. The plasmid TL7 includes a series of base substitutions that result in stop codons that prevent expression of viral envelope proteins without altering the overlapping polymerase gene. These mutations block the major route of virus excretion from the cell, leading to an intracellular accumulation of pgRNA-filled particles. To construct 1159 from TL7, mutations were introduced that create a Y63F change in HBV Pol eliminating polymerase priming activity and that allow accumulation of pgRNA-filled cores. A 4-nucleotide substitution within the splice acceptor site at nt 485 in the HBV sequence (cagg→gtcc at positions 484–487) was introduced. This mutation has been shown to eliminate the majority of pgRNA-filled particles, increasing the percentage of full-length pgRNA filled cores (29).

Cell Cultures and Transfections. Human hepatoma cell line HuH7-H1 was used for production of pgRNA-filled cores and has been demonstrated to support high levels of HBV replication (27). Cells were maintained in DMEM/F12 media supplemented with 5% FBS and 600 μg/mL G418. Cell cultures were transfected using the calcium phosphate precipitation method described previously (48). For production of pgRNA-filled cores, 105 cells were seeded into 100-mm plates and grown to 70–80% confluence. Media was removed and replaced by DMEM/F12 supplemented with 5% (vol/vol) FBS without G418 0–4 h before transfection. Cells were transfected with 32 μg plasmid 1159 and 1 μg plasmid 1929. Medium [DMEM/F12 plus 5% (vol/vol) FBS] was replaced at 16–20 h after transfection and was changed every 24 h for 4 d.
Purification of HBV Virions. Four days after transfection cells were washed with 10 mM ice-cold Hepes-buffered saline (HBS) plus EGTA buffer (2 mM Hepes at pH 7.5, 150 mM NaCl, and 0.5 mM EGTA) and frozen at −80 °C. Plates were thawed and cells were treated with 1 mL lysis buffer [50 mM Mepes at pH 7.5, 100 mM NaCl, 0.25% Nonidet P-40 (wt/vol), 0.5 mM EDTA, and one Complete Mini EDTA-free protease inhibitor tablet (Roche) per 10 mL lysis solution]. Cells were scraped, collected, and vortexed several times. Lysate was clarified by centrifugation at 10,000 × g for 1 h to pellet cell debris.

Clariﬁed lysate was loaded onto a 40%/50%/60% (wt/vol) sucrose step gradient in 25 × 89 mm Ultra-Clear centrifuge tubes (Beckman Coulter). Each sucrose solution was prepared in CLPM buffer [50 mM Mepes at pH 7.5, 100 mM NaCl, 50 mM l-arginine, 50 mM l-glutamate, 1 mM EDTA, 2 mM DTT, 0.05% Nonidet P-40 (wt/vol), and 1% Trehalose and one Complete EDTA-free protease inhibitor tablet (Roche) per 50 mL sucrose solution]. Sucrose gradients were centrifuged for 16 h at 110,000 × g for 1 h to pellet cell debris.

Sucrose gradient fractions with cores were identiﬁed by a native-particle blot Western assay. Samples were resolved using a 1% agarose gel run at 80 V and 300 mA in 1x Tris-acetate-EDTA buffer for 75 min. Gels and ﬁlter papers were equilibrated in 1x Tris-NaCl-EDTA (TNE) buffer (10 mM Tris at pH 8.0, 150 mM NaCl, 1 mM EDTA) for at least 15 min. Immobilon-P 0.45-μm pore size transfer membranes (Millipore) were treated with methanol and then equilibrated in TNE. Capsids were transferred to this membrane by capillary action, and then membranes were air dried and blocked by incubation at room temperature for 1 h in 1x phosphate-buffered saline with 0.05% Tween 20 (PBST) buffer supplemented with 5% (wt/vol) skim milk powder. Capsid particles were detected by incubation with a polyclonal rabbit HBV core antibody (Dako) at a 1:2,000 dilution for 1 h in 1x PBST buffer. Membranes were washed three times for 15 min with 1x PBST and then incubated for 1 h at room temperature with a Protein A-HPR conjugate (Thermo/Pierce) at a 1:5,000 dilution in 1x PBST buffer. Membranes were washed three times for 15 min with 1x PBST and then for 5 min in PBS to remove Tween. Washed membranes were treated with Pierce ECL Western Blotting Substrate according to the manufacturer’s recommendations (Thermo/Pierce). Selected sucrose gradient fractions were pooled and loaded onto a 200-ml Sephacryl 300-high resolution column equilibrated at 4 °C with CLPM buffer without protease inhibitor. Fractions with high A260/A280 ratios were shown by Western blotting to contain the highest concentration of cores. These were loaded onto a Superose6 column (GE Healthcare) equilibrated with 1x CLPM buffer. Samples were run at 0.5 mL/min, and fractions with high A260/A280 ratios were collected. Superose6 fractions with the highest core concentration samples were pooled and concentrated with a Nanosep 100K Omega spin concentrators at 1500 × g (Pall Life Sciences). Concentrated samples were stored at 4 °C.

pgRNA/Capsid Ratio Measurement. To determine the ratio of pgRNA/Capsid, pgRNA was isolated from pgRNA-ﬁilled core preparations at each stage of purification. To each sample, an equal volume of 2x Prok digestion buffer (50 mM Hepes, 80.0 mM NaCl, 4 mM EDTA, 1% (wt/vol) SDS, Protease K 22 ng/mL) was added, and samples were incubated for 1 h at 37 °C. Samples were then extracted with an equal volume of TRizol and precipitated ions according to the manufacturer’s protocol. Samples were resuspended in 10 μL nucleic-acid-free water and stored at −20 °C.

pgRNA for use as a standard was prepared by in vitro transcription from plasmid 113S5 (9). Standard pgRNA concentrations were determined by absorbance to prepare standards from 3–0.003 fmol per reverse transcription quantitative PCR (RT-qPCR) reaction.

The concentration of pgRNA was determined by RT-qPCR. To copy pgRNA to a cDNA, 2.5 μL RNA sample was annealed to 2 pmol oligonucleotide RevC1 (5'-CCGGACATGAGACGACGACGAC-3') complementary to the 3' region of the pgRNA. Annealing was performed in 6 μL at 65 °C for 5 min followed by 5 min on ice. Reverse transcription was performed in SuperScript III (Invitrogen) reaction buffer (6 μL of annealing reaction, 1X reaction buffer, 5 mM DTT, 750 μM dNTPs, 1 mM MgCl2, 60 U SuperScript III reverse transcriptase) in 20-μL reaction volumes at 52 °C for 4 h. A 2.5-μL aliquot of each reverse transcription reaction was added to 12.5-μL qPCR reaction mix.


