Virus self-assembly proceeds through contact-rich energy minima

Buzón, Pedro; Maity, Sourav; Christodoulis, Panagiotis; Wiertsema, Monique J; Dunkelbarger, Steven; Kim, Christine; Wuite, Gijs J L; Zlotnick, Adam; Roos, Wouter H

Published in:
Science Advances

DOI:
10.1126/sciadv.abg0811

Publication date:
2021

LIFE SCIENCES

Virus self-assembly proceeds through contact-rich energy minima

Pedro Buzón[1], Sourav Maity[1], Panagiotis Christodoulis[1], Monique J. Wiertsema[1], Steven Dunkelbarger[2], Christine Kim[2], Gijs J.L. Wuite[3], Adam Zlotnick[2], Wouter H. Roos[1]*

Self-assembly of supramolecular complexes such as viral capsids occurs prominently in nature. Nonetheless, the mechanisms underlying these processes remain poorly understood. Here, we uncover the assembly pathway of hepatitis B virus (HBV), applying fluorescence optical tweezers and high-speed atomic force microscopy. This allows tracking the assembly process in real time with single-molecule resolution. Our results identify a specific, contact-rich pentameric arrangement of HBV capsid proteins as a key on-path assembly intermediate and reveal the energy balance of the self-assembly process. Real-time nucleic acid packaging experiments show that a free energy change of ~1.4 k_BT per condensed nucleotide is used to drive protein oligomerization. The finding that HBV assembly occurs via contact-rich energy minima has implications for our understanding of the assembly of HBV and other viruses and also for the development of new antiviral strategies and the rational design of self-assembling nanomaterials.

INTRODUCTION

A pronounced feature of some of the most fundamental cellular processes is the self-assembly of supramolecular complexes (*1*). Viruses use this strategy to spread through their hosts, i.e., by forming new viral particles. Virus self-assembly is an orchestrated process, in which the formation of the protective protein shell (the capsid) and the encapsidation of the viral genome can occur concomitantly. The self-assembly reaction is driven by interactions between capsid proteins and often with the viral genome. The energetics of these interactions must be finely tuned to avoid trapping defects. Although capsid formation is a crucial step to ensure infectivity and viral propagation, the mechanisms underlying this process still remain poorly understood, particularly in the presence of nucleic acid and at the molecular level (*2*–*5*). The experimental challenges associated with the characterization of the viral assembly process reside in the heterogeneity generated in the system, including the conformational plasticity that viral proteins exhibit during oligomerization (*6*) and their interaction with the genome (*7*). In addition to the clear implications in the development of new antiviral strategies, the molecular mechanisms of assembly are of growing interest in biomedicine and (bio)nanotechnology (*8*, *9*).

In this context, hepatitis B virus (HBV) has emerged as a model system to study capsid assembly. HBV forms $T = 4$ icosahedral capsids as the major product of in vivo assembly (*10*, *11*), under conditions that can be reproduced in vitro (*12*, *13*). HBV $T = 4$ capsids are formed by 120 core protein (Cp) homodimers; each Cp consists of an assembly domain (residues 1 to 149) and a C-terminal domain (CTD; residues 150 to 183), an intrinsically disordered region of 34 amino acids (Fig. 1, A and B). The assembly domain is necessary and sufficient for the assembly of empty capsids, which are indistinguishable from the nucleic acid–filled ones (*14*). The CTD is not necessary for capsid assembly, but its basic nature (provided by 16 arginines) facilitates the interaction of the protein with nucleic acid and is essential for the formation of genome-containing capsids (*15*). In vivo, HBV Cp either forms empty capsids or encapsidates an mRNA transcript of the viral genome [pregenomic RNA (pgRNA)] and a copy of viral reverse transcriptase; the pgRNA is reverse transcribed to a circular double-stranded DNA (dsDNA) within the capsid (*16*, *17*).

HBV capsid assembly is assumed to start with the formation of a nucleus (*18*), which then evolves into a complete capsid (Fig. 1C) by the addition of Cp dimers (*13*). The formation of assembly nuclei is expected to be a relatively rare event. In addition, small capsid intermediates are thought to be transient (*19*, *20*). Thus, while bulk methods have been instrumental to study the thermodynamics of assembly, its kinetics has largely remained elusive. Only recently, it has become possible to study the dynamics of viral self-assembly at the single-particle level (*21*–*25*). Here, we use a combination of optical tweezers (OT) with confocal fluorescence microscopy (CFM) and high-speed atomic force microscopy (HS-AFM) to follow the assembly of HBV Cp in real time. Our approach allows tracking the elusive first steps of assembly with high temporal resolution and single-molecule sensitivity, thereby revealing how on-path intermediates follow contact-rich energy minima during nucleation and growth in the HBV assembly pathway.

RESULTS

Capsid assembly occurs through the active packaging of nucleic acids

The assembly of Cp around single-stranded DNA (ssDNA) was followed in real time by OT isotensional experiments. Although in vivo Cp assembly occurs around the ssRNA HBV pregenome, it has been shown that in vitro Cp assembles on ssDNA, as well as ssRNA of heterologous sequences, with similar levels of cooperativity and producing viral particles that are indistinguishable from the pregenome-filled ones (*13*, *26*). Although both ssDNA and ssRNA have comparable physical chemical properties, including a similar persistence length, ssDNA presents higher chemical stability. In addition to that, we have chosen to work with heterologous sequences of ssDNA to avoid the high degree of branching found in viral ssRNA

[1]Moleculaire Biofysica, Zernike Instituut, Rijksuniversiteit Groningen, Groningen, Netherlands. [2]Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, IN, USA. [3]Physics of Living Systems, Vrije Universiteit, Amsterdam, Netherlands. *Corresponding author. Email: w.h.roos@rug.nl

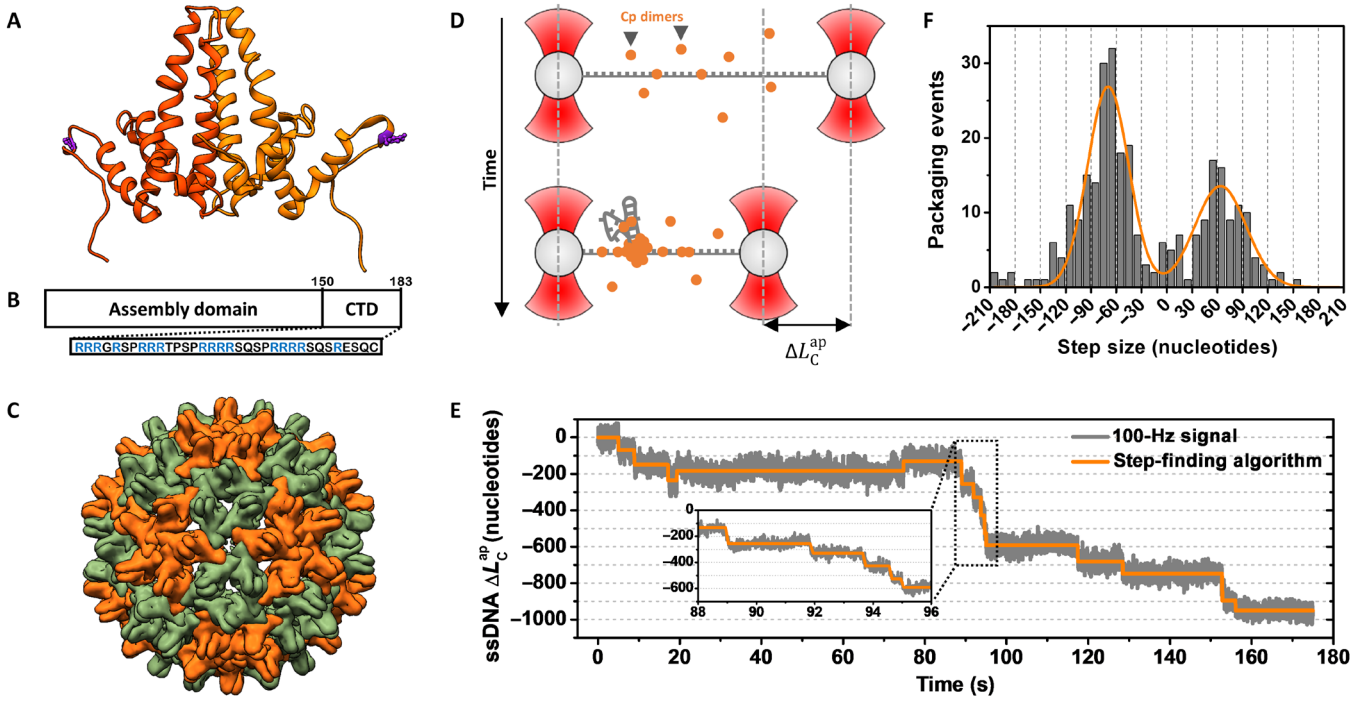

Fig. 1. Real-time nucleic acid packaging by the HBV capsid protein (Cp). (**A**) Structure of the assembly domain of a Cp dimer (orange) in the context of a capsid with the residue Y132 highlighted in purple [Protein Data Bank (PDB) 2G33 (*39*)]. (**B**) Scheme of the primary structure of Cp, including the sequence of the C-terminal domain (CTD). (**C**) Surface representation of the $T = 4$ HBV capsid [PDB 2G33 (*39*)], where A-B dimers conform the fivefold symmetry axes (orange) and C-D dimers localize around the threefold symmetry axes (green). (**D**) Illustrative cartoon of an OT isotensional experiment showing ssDNA packaging by Cp dimers (orange) upon binding. (**E**) Representative real-time assembly trace obtained by an OT isotensional experiment. A single molecule of ssDNA (pKYB1) is held at constant force of 11 pN in the presence of Cp dimers, kept in solution, while monitoring the overall change in the apparent contour length (ΔL_C^{ap}) over time (gray). The fit to the step-finding algorithm is shown in solid orange line. The inset highlights a period of 8 s, at the center of the trace, showing fast Cp packaging. (**F**) Step size distribution in number of nucleotides obtained from the analysis of the real-time assembly traces by the step-finding algorithm ($N = 287$ steps). Orange line represents the fit to a double Gaussian ($xc_1 = -70 \pm 2$ nt; $xc_2 = 64 \pm 4$ nt).

genomes (*27*), under conditions that successfully produce virus-like particles in vitro (*26*).

Molecules of ssDNA were held at constant tension in the presence of purified Cp dimers under assembly conditions (see Fig. 1D and Materials and Methods). Under such conditions, capsid assembly is followed by monitoring the changes in the apparent DNA contour length (ΔL_C^{ap}) (Fig. 1, D and E). A representative packaging trace (Fig. 1E) shows a net decrease in L_C^{ap} equivalent to approximately 1000 bases over roughly 3 min, indicating the active packaging of ssDNA by Cp dimers. The net number of packed nucleotides measured over time was found to be heterogeneous, as expected, due to the stochastic nature of the self-assembly process. In particular, Fig. 1E shows that the process may combine long periods of inactivity (from 20 to 80 s) with others of high cooperativity (from 88 to 96 s). Despite the heterogeneity of the packaging traces, most of them showed that packaging occurs through sequential assembly and disassembly steps, highlighting the reversibility of the process. As a control, when the experiments were performed with an assembly-incompetent version of Cp, the mutant Cp-Y132A (*28*), no packaging events were found.

The step sizes were extracted from the packaging traces by a custom-written step-finding algorithm (see Fig. 1E and Materials and Methods). Figure 1F shows the step size distribution of all measured assembly/disassembly steps ($N = 287$ steps, from 24 traces). Two populations emerged at approximately −70 bases (assembly) and +64 bases (disassembly) (Fig. 1F). This indicates that Cp presents a most preferred assembly configuration, the assembly footprint, which results in the packaging of ~70 nucleotides (nt) (~37 nm, using a 0.53-nm nucleotide spacing as described in Materials and Methods). These experiments reveal that Cp is able to condense nucleic acids, even when the genome is under tension. The work per condensation step can be estimated (*29*) at $F\delta = 100 \pm 4\ k_BT$, with the external conservative force F (11 pN), step size δ, Boltzmann constant k_B, and temperature T. We note that this condensation step corresponds to ~1.4 k_BT/nt, which is substantially smaller than the value reported for the adenosine triphosphate–dependent packaging motor of phage φ29 [4.8 k_BT per base pair (bp) (*30*)]. Nevertheless, a condensation step of ~100 k_BT suggests that the participation of multiple Cp dimers is required to achieve such free energy gains.

Cp exhibits nucleic acid folding capabilities

Cp is known to bind to dsDNA, although the formation of closed capsids is inhibited due to stiffness of the dsDNA (*26*). To test whether Cp has the ability to generate and stabilize dsDNA regions during packaging, the DNA intercalating dye SYTOX Orange was added to the self-assembly experiments. SYTOX intercalates between dsDNA base pairs, giving a substantial increase in fluorescence signal (fig. S1) and does not bind to ssDNA. Here, we used SYTOX to detect and localize dsDNA regions on ssDNA molecules that were previously incubated with Cp dimers (Fig. 2A). All ssDNA molecules

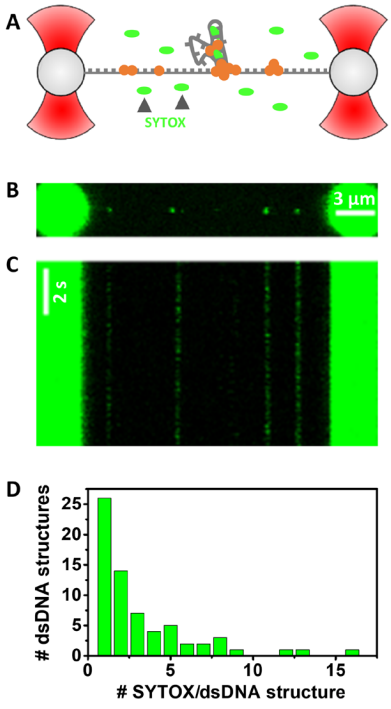

Fig. 2. DNA compaction facilitated by Cp nucleic acid folding function. (**A**) Illustrative cartoon in which a complex Cp-ssDNA is held in the presence of SYTOX. Cp dimers (orange) bound to ssDNA facilitate DNA folding, allowing SYTOX (green ellipsoids) to bind. (**B**) Fluorescence microscopy image of a Cp-ssDNA complex in the presence of 250 nM SYTOX. Molecules of ssDNA (λ DNA) were first incubated with Cp from 30 s to 2 min and then moved to a SYTOX solution to image the complex by CFM. The image shows SYTOX signal localized on the DNA molecule, indicative of the presence of DNA base pairing. (**C**) Kymograph of the same complex shown in (B) stretched at 50 pN. The fluorescence signal remains static, showing the dynamic binding-unbinding of SYTOX to the stabilized dsDNA regions. (**D**) Histogram of the mean number of SYTOX molecules bound per dsDNA structure ($N = 67$) over time, obtained from kymographs in which DNA molecules were stretched at >30 pN and measured for 30 s (see fig. S1).

were stretched with a force of at least 30 pN to confirm the formation of stable dsDNA regions. In a typical CFM image of a Cp-ssDNA complex stained with SYTOX, several fluorescent spots can be seen between the beads (Fig. 2B), revealing the presence of DNA base stacking and indicating the formation of dsDNA structures stabilized by Cp on the ssDNA molecule. Fluorescence spots were observed at different locations in different DNA molecules, suggesting that the Cp stabilizes hairpins and/or other dsDNA structures randomly, i.e., non–sequence specific. No signal was detected when ssDNA molecules were not previously incubated with Cp.

The kymograph presented in Fig. 2C shows that the positions of the dsDNA structures are stable over time. By calibrating the brightness of individual SYTOX molecules (fig. S1), the mean number of SYTOX molecules bound to each dsDNA structure over time can be estimated (Fig. 2D and fig. S1). A correlated force-fluorescence experiment (fig. S2) shows DNA packaging and simultaneous formation of a dsDNA structure in real time. However, these correlated events were rare, as most packaging events did not exhibit fluorescence signal, indicating that DNA packaging does not necessarily involve base pairing. It has been shown by electron microscopy reconstruction of pregenome-filled HBV capsids that single-stranded and double-stranded nucleic acid regions are present inside complete capsids (*31*). Our results show that Cp has the capability to chaperone the formation of dsDNA.

Free energy balance between Cp-Cp and Cp-DNA interactions

To determine the minimal oligomeric arrangement of Cp dimers that is able to condense and fold nucleic acids, Cp-ssDNA interactions were further characterized by applying a single-molecule force spectroscopy approach using OT (fig. S3). Briefly, a single molecule of ssDNA tethered between two beads is sequentially stretched and relaxed in a Cp dimer solution (Fig. 3A). Figure 3B shows three representative force-extension curves (FECs) of ssDNA molecules in the absence of protein (bare ssDNA) and at two different Cp concentrations. A change in protein concentration clearly changes the shape of the curves (Fig. 3B). This behavior indicates that changes in the mechanical properties of DNA are taking place upon Cp binding. Quantitative information on DNA mechanics was extracted from FECs by applying the extensible freely jointed chain (eFJC) model (Fig. 3B and fig. S3). The eFJC analysis revealed that both Cp-ssDNA and Cp-Y132A-ssDNA complexes are stiffer than bare ssDNA; specifically, Fig. 3C shows that persistence length (L_P) increases in a saturable concentration-dependent manner (tables S1 and S2).

The ssDNA saturation curves for both Cp and Cp-Y132A (Fig. 3C) were analyzed using the McGhee–von Hippel (MGvH) model (*32*). Figure 3C shows the best global fit to both Cp and Cp-Y132A curves, and Table 1 lists the fitting parameters and the derived thermodynamic parameters obtained from the analysis. Note that the fitting was performed assuming that Cp and Cp-Y132A share the same intrinsic dissociation constant for the binding of one dimer for DNA ($K_{D,int}$), which was kept as shared parameter during the global fit. This seems justified because the residue Y132 is not located in the CTD (Fig. 1, A and B), and the mutation Y132A should not affect the intrinsic affinity of the protein for DNA. We fixed the MGvH parameter for the binding footprint (n) to 30 nt, as previously estimated (*26*). We note that using $20 \leq n \leq 40$ does not change the quality of the fit and barely changes the extracted thermodynamic parameters (table S3).

The binding parameters reveal that both Cp and Cp-Y132A bind to ssDNA with very similar affinity, $K_D \approx 100$ nM and $\Delta G^{\circ}_{bind} \approx -16\ k_BT$ (Table 1). These results primarily show that single-molecule force spectroscopy experiments in combination with the MGvH model can be used to decouple Cp binding to nucleic acids from the global assembly process, as Cp-Y132A is not able to assemble into capsids (*28*) or perform DNA packaging under our experimental conditions. A closer look at Table 1 shows that both Cp and Cp-Y132A bind to ssDNA with high levels of cooperativity ω ($\Delta G^{\circ}_{prot-prot} \approx -7\ k_BT$). This free energy is in agreement with the ΔG° values reported for the formation of empty capsids using a truncated version of Cp (Cp149) that lacks the CTD (*33*, *34*). To gain more insight into Cp-ssDNA interactions, force spectroscopy experiments were also performed for only the CTD of Cp (fig. S4). However, we found that micromolar concentrations led to CTD aggregation under assembly conditions, preventing quantitative studies. Despite this, the ssDNA saturation curve obtained with the CTD qualitatively matches the values of $K_{D,int}$ found for Cp and Cp-Y132A (fig. S4). The reported values of $\Delta G^{\circ}_{prot-DNA}$ can be seen as a lower bound for the affinity of Cp, as Cp-DNA interactions are electrostatically driven and our

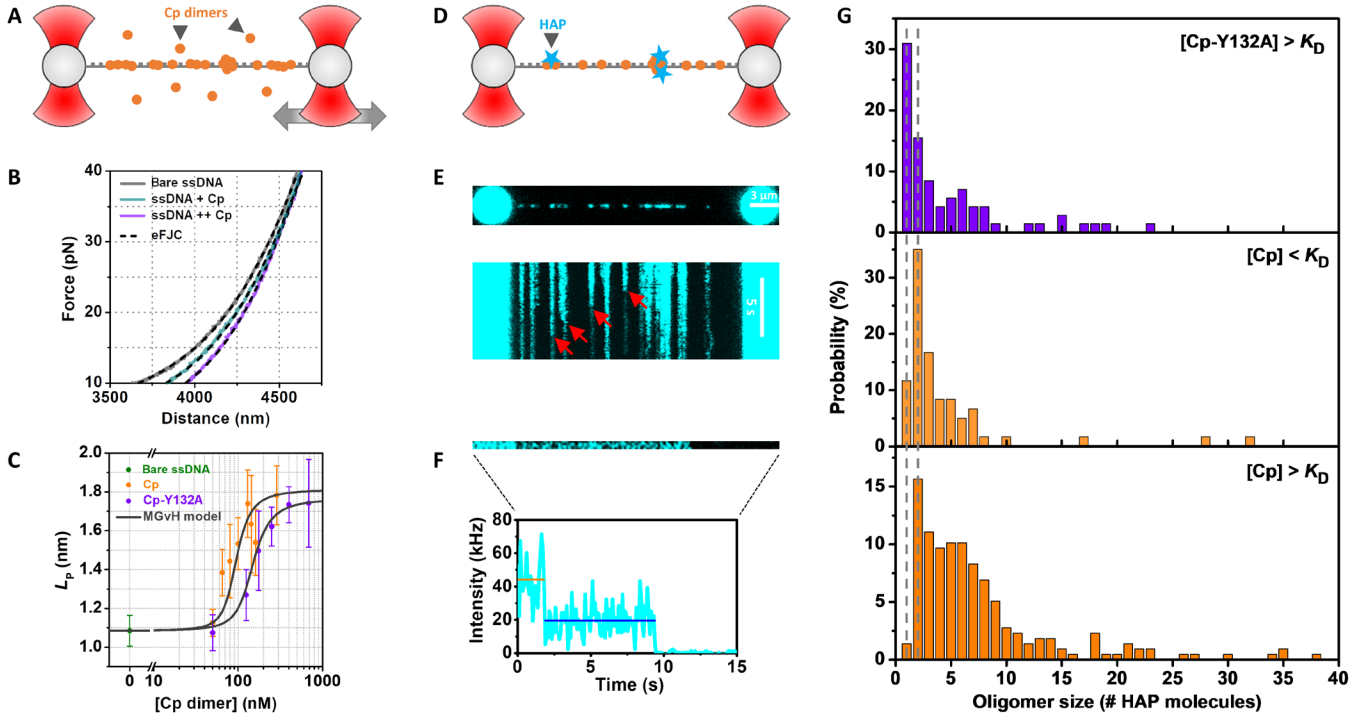

Fig. 3. Characterization of Cp-ssDNA interaction. (**A**) Schematic cartoon of a force spectroscopy experiment by OT of an ssDNA molecule held in a solution of Cp dimers (orange). (**B**) Representative FECs of bare ssDNA (gray; pKYB1) and ssDNA under increasing concentration of Cp (light green and purple). The individual fits to the eFJC model are represented in black dashed lines. (**C**) DNA saturation curves obtained for Cp (orange) and Cp-Y132A (purple) represented as persistence length (L_P) versus Cp dimer concentration. The best global fit to the MGvH model is displayed as gray lines. Error bars represent SD. (**D**) Illustrative cartoon of a Cp-ssDNA complex after its incubation against a HAP-Alexa488 (cyan stars) solution to specifically stain Cp oligomers (orange). (**E**) Top: CFM image of a Cp-ssDNA-HAP-Alexa488 complex (λ DNA). Bottom: Kymograph of a photobleaching experiment of the same complex. Traces were typically static, showing a decrease in intensity over time in a stepwise manner (red arrows), and only few traces showed diffusing behavior (<5%). (**F**) Kymograph of a single Cp oligomer during a photobleaching experiment (top) and the corresponding intensity profile of the trace. Orange and blue lines show the data used to calculate the number of bound HAP molecules (G) and the brightness of HAP-Alexa488, respectively (see fig. S6). (**G**) Histograms of the number of HAP molecules bound per Cp cluster. Cluster size was only measured for diffraction-limited spots. [Cp-Y132A] = 625 nM (top), [Cp-WT] = 50 nM (middle), and [Cp-WT] = 200 to 600 nM (bottom). Total number of oligomers, N = 71, 60, and 217, respectively.

Table 1. MGvH binding parameters (n = 30).

Cp dimer	*$K_{D,int}$ (µM)	ω	†K_D (µM)	‡$\Delta G^0_{prot-DNA}$ (k_BT)	‡$\Delta G^0_{prot-prot}$ (k_BT)	‡ΔG^0_{bind} (k_BT)
Cp	70 ± 30	800 ± 200	0.09 ± 0.04	−10 ± 2	−7 ± 1	−16 ± 2
Cp-Y132A		500 ± 100	0.14 ± 0.06		−6.2 ± 0.9	−16 ± 1

*$K_{D,int}$, the dissociation constant of Cp and ssDNA, was kept as a shared parameter. This ionic strength–sensitive parameter was determined for a buffer of 25 mM Hepes (pH 7.4), 375 mM guanidine HCl, 125 mM LiCl, and 250 µM TCEP. †$K_D = K_{D,int}\,\omega^{-1}$, where ω is the unitless association constant for two nucleic acid–bound ligands. The ω value may be considered a measure of cooperativity. ‡$\Delta G°$ values obtained from $K_{D,int}$, ω, and K_D using the expression $\Delta G° = -k_BT\ln(K)$, where k_BT is the product of the Boltzmann constant and the absolute temperature.

conditions are set to ~500 mM ionic strength, including 375 mM guanidine HCl. These conditions will induce a higher degree of charge screening than under typical intracellular conditions.

The energetics of protein-DNA interaction ($\Delta G^0_{prot-DNA}$) are the same order of magnitude as $\Delta G^0_{prot-prot}$ values in high ionic strength, and they only differ ~3 k_BT (Table 1). Thus, the MGvH model highlights the contribution of cooperativity derived from Cp-Cp interaction to the global ΔG^0_{bind}. Notably, the MGvH analysis allows us to calculate the free energy of recurring arrangements of Cp dimers that could package ssDNA with the disclosed assembly footprint of ~70 nt (~100 k_BT). We estimate that five to six dimers might be sufficient for packaging (fig. S5).

Cp oligomerization behavior around ssDNA

To investigate Cp-Cp interactions for Cp bound to ssDNA, Cp oligomers formed on ssDNA molecules were specifically stained with heteroaryldihydropyrimidine (HAP)–Alexa488 (Fig. 3, D and E). HAP provides specific noncovalent labeling of HBV capsids, as it does not bind to free Cp dimers (*35*, *36*). T = 4 HBV capsids bind up to 120 HAP molecules at dimer-dimer interfaces, with a 1:1 stoichiometry with

respect to Cp dimers under saturating conditions (*37*). Here, we use HAP-Alexa488 to label capsid-like interfaces of higher-order structures formed by Cp dimers under nonpackaging conditions (load, >20 pN). Figure 3E shows the formation of Cp oligomers arranged as stable clusters on an ssDNA molecule. The sizes of such oligomers were estimated by photobleaching experiments (Fig. 3, E and F, and fig. S6).

The distributions obtained for both Cp and Cp-Y132A (Fig. 3G) reveal clear differences between the two versions of the protein that could not be anticipated from the binding parameters discussed above. Cp-Y132A experiments show that the major peak in the histogram corresponds to one bound HAP, indicating that the most probable oligomers to be formed on ssDNA are dimers of dimers, that is, oligomers with only one dimer-dimer interface (Fig. 3G, top). However, under similar conditions ([Cp] > K_D), Cp shows very low probability of small clusters of HAP binding dimers of dimers (below 2%); a distribution of higher-order oligomers is found (Fig. 3G, bottom). Even at a much lower concentration of protein ([Cp] < K_D), Cp shows preference to form oligomers that bind to two or more HAP molecules (Fig. 3G, middle). On the other hand, the Cp oligomer size distribution obtained for [Cp] > K_D (Fig. 3G) points toward stable oligomeric structures binding less than 10 HAP molecules, in agreement with the structures proposed in fig. S5.

Identification of early assembly intermediates

The oligomerization behavior of Cp was studied by HS-AFM, providing structural information on the formation of small Cp oligomers with high temporal resolution. For these experiments, a negatively charged mica surface was used as a proxy for nucleic acid (*25*). Immediately before AFM imaging, Cp dimers were brought to assembly conditions by dilution, at maximum dimer concentrations of 300 nM. Movie S1 shows how Cp dimers arrive to the surface to form higher-order structures until surface saturation. A closer look at the oligomers reveals two major arrangements: a diamond pentamer of dimers and a dodecamer of dimers (Fig. 4A). The latter exhibits quasi-sixfold symmetry suggestive of $T = 4$ twofold symmetric site. The Cp dimers appear to transiently interact with the negatively charged surface, leaving the spike of the dimers facing upward, which facilitates identifying individual Cp dimers (Fig. 4A). When assembly reactions are performed overnight in a test tube and above 1 μM concentration of Cp dimers, we observe fully formed empty HBV capsids (fig. S7), in agreement with the previously reported assembly conditions (*13*). Experiments performed with Cp-Y132A did not show the formation of small oligomers able to diffuse on the surface, but the fast growth of a flat two-dimensional (2D) lattice that is reminiscent of the hexameric planes in Cp-Y132A crystals (fig. S8) (*38*).

Figure 4 (B and C) shows the molecular surface representation of the diamond pentamer and the dodecamer in the context of a full capsid (*39*). The diamond pentamer exhibits five spikes where the middle spike is slightly higher than the surrounding ones (Fig. 4, A and B, and fig. S7). On the other hand, the dodecamer is composed of six high spikes forming the central hexagon and up to six lower spikes surrounding the central hexagon, indicating the initiation of curvature formation (Fig. 4, A and C, and fig. S7). Figure 4 (D and E) shows examples of the formation of both oligomeric structures in real time (see also fig. S9 and movies S2 to S4). Both structures fit within the oligomer size distribution found by HAP staining experiments under nonpackaging conditions (Fig. 3G). We have also observed that most of the dodecamers of Cp dimers are formed starting from at least one diamond pentamer; in some cases, two pentamers directly combine to form an oligomer of 10 Cp dimers with the dodecamer arrangement (see Fig. 4E, fig. S9, and movie S3). The work per condensation step produced by Cp dimers ($F\delta = 100 \pm 4\ k_BT$) correlates with our estimations of free energy provided by the assembly of a diamond pentamer on ssDNA ($\Delta G° = -90 \pm 10\ k_BT$) (fig. S5). It suggests that the formation of a single diamond pentamer might be sufficient to facilitate DNA condensation, supporting the likeliness of this pentamer as a key assembly intermediate. Figure S10 shows that the diamond pentamer can be found either bound to three or two double-stranded regions in the context of a capsid. Moreover, these results show that the HBV assembly pathway proceeds by the formation of diamond pentamers and dodecamers, instead of through the fivefold symmetric pentamer, as also supported by recent electron microscopy experiments (*40*).

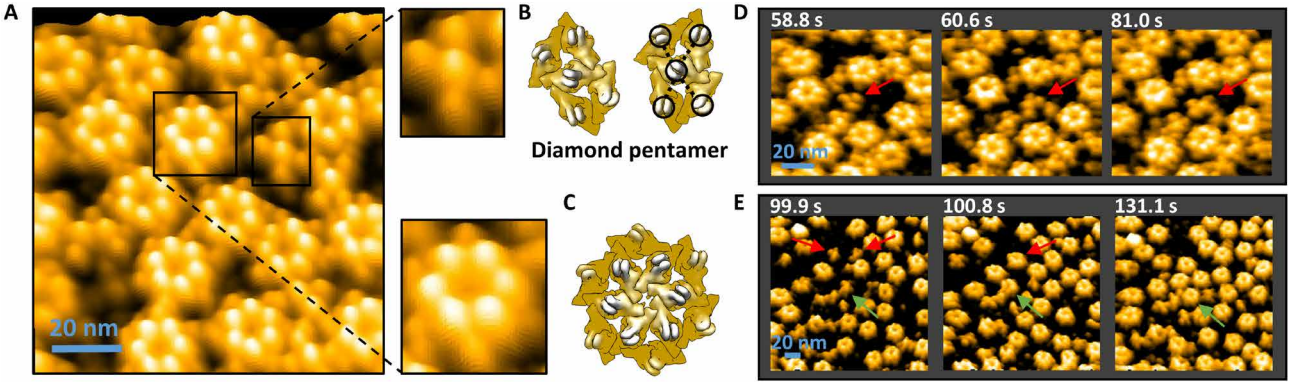

Fig. 4. Cp oligomerization behavior followed by HS-AFM. (**A**) Left: Averaged image of self-assembled Cp oligomers from a solution of Cp dimers at 300 nM. Right: Zoom-in of the most abundant oligomers found: the diamond pentamer, formed by five Cp dimers (top), and the dodecamer, found around the twofold symmetry axes of HBV capsids (bottom). (**B** and **C**) Cartoons of the diamond pentamer (B) and the dodecamer (C), both in the context of the $T = 4$ HBV capsid. The diamond pentamer appears under the AFM with its five dimer spikes arranged as an "X" (B) (right). Low-resolution surfaces are colored by height. (**D**) HS-AFM images following the formation of a diamond pentamer starting from a trimer of dimers (red arrows) (see fig. S9 and movie S2). (**E**) Selected snapshots capturing the formation of two hexagon-mers by the direct combination of two diamond pentamers (red arrows) and through the sequential addition of Cp dimers (green arrows) (see fig. S9 and movies S1, S3, and S4). Images in both (D) and (E) were acquired at 3.33 frames/s. Molecular models presented in (B) and (C) were generated from the x-ray crystal structure [PDB 2G33 (*39*)].

DISCUSSION

We have demonstrated that HBV Cp has the ability to actively package nucleic acid by the mere self-assembly of Cp oligomers. Executing these packaging events requires ~1.4 k_BT/base of work and might involve the base pairing of nucleic acids to form double-stranded regions (Figs. 1 and 2). Although, pregenomic ssRNA is assumed to present secondary structure in vivo, displaying a globular arrangement, we have shown that full-length HBV Cp has the ability to promote nucleic acid folding, consistent with earlier observation of RNA chaperone activity for peptides based on the CTD (*41*). This function could be of particular relevance in the crowded environment of the cytoplasm, where the HBV pregenome could be bound to other cellular factors, or in a capsid where it is likely that the RNA must be properly organized for reverse transcription. In this context, Cp packaging capabilities could be very important to achieve a condensed ternary structure during capsid assembly.

Our results indicate that the diamond-shape pentameric arrangement of Cps is a critical participant in HBV assembly, as a minimal stable structure of Cps able to condense single-stranded nucleic acids. Previous studies have reported the accumulation of small HBV assembly intermediates for truncated versions of Cp (without the CTD) and in the absence of nucleic acids (*34*, *40*, *42*–*44*). The assembly model proposed here (Fig. 5) takes into account the stabilizing contributions of nucleic acids in the formation of small nucleation structures. A key point of our assembly study is that it is performed below pseudocritical concentration of Cp dimers, conditions that do not lead to the assembly of empty HBV capsids (*45*). The addition of a negatively charged substrate for assembly, e.g., nucleic acid or a surface, introduces an additional force driving the reaction toward assembly (Fig. 5). This, in combination with our single-molecule approach, allows us to capture in real time these critical structures on the path to capsid assembly.

The pentameric arrangement of Cp dimers identified here (Fig. 4A) shows important characteristics that feature its potential as a key building block of assembly. When looking at the angles formed between dimer spikes, we notice that the larger angle peaks at ~115° (fig. S7), just in between the angles of a regular pentagon (108°) and a regular hexagon (120°). It highlights its versatility to be part of either the fivefold or quasi-sixfold symmetry axes of complete capsids. In addition, the diamond pentamer represents a local minimum of energy in the oligomerization landscape of Cp, as it is the smallest configuration of dimers that has more interfacial contacts than subunits. Last, we provide experimental evidence that assembly proceeds by reversible reactions, where subunits associated to a growing complex by a single bond have a high probability of dissociating before getting locked in place by an incoming subunit, whereas subunits forming part of contact-rich structures are likely to persist (fig. S11) (*34*, *46*).

Our results complement the recent study by Asor *et al.* (*34*, *44*), which examined the assembly of empty HBV capsids for assembly domain dimers using time-resolved small-angle x-ray scattering (SAXS). The SAXS experiments used much higher protein concentrations than HS-AFM to obtain an interpretable signal at a resolution that allowed distinguishing intermediates. In the SAXS experiments, the protein concentrations were 20 and 30 μM, whereas the HS-AFM experiments in this study were performed at concentrations of 0.1 to 0.3 μM, two orders of magnitude less. The HS-AFM concentrations are below the pseudocritical concentration of assembly (for the assembly domain) and likely depend on the presence of a substrate, be it nucleic acids or mica. Note that the concentrations used in HS-AFM studies are similar to those in a cell (*47*). The high concentrations used in SAXS experiments can support rapid assembly with modest accumulation of small labile species, such as dimers of dimers and trimers of dimers, which are relatively rare at low

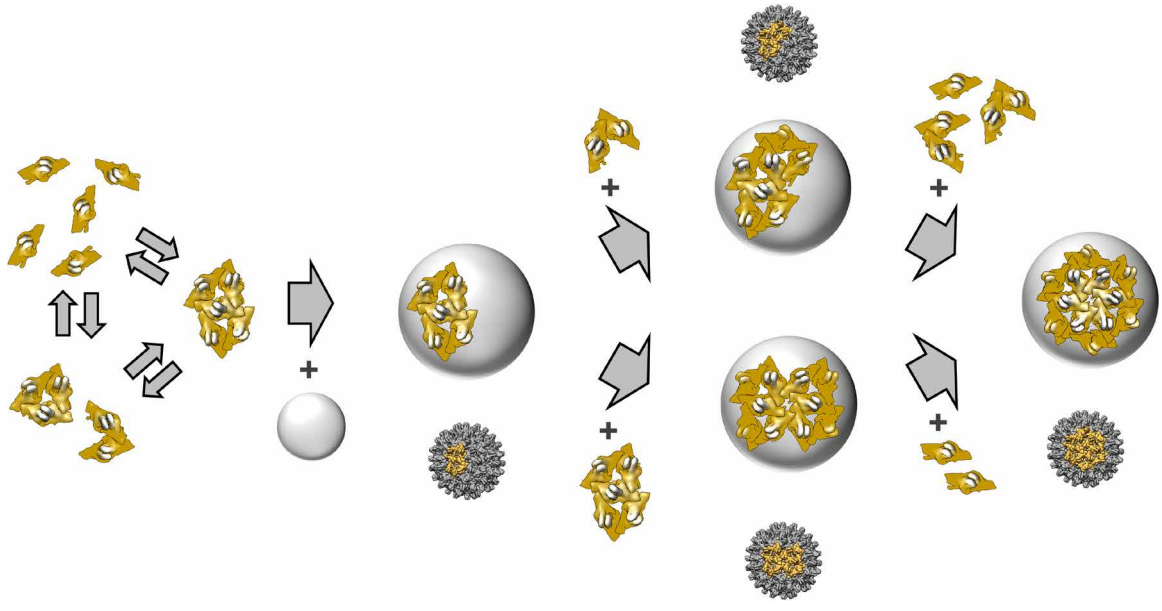

Fig. 5. Model of substrate-assisted Cp assembly. Below pseudocritical concentration of Cp dimers, small oligomers are in equilibrium with free dimers, where interactions are driven by weak dimer-dimer interactions ($\Delta G^0_{prot-prot}$). In the presence of a negatively charged substrate (gray sphere), diamond pentamers are stabilized by the additional contribution of the positively charged CTD ($\Delta G^0_{prot-DNA}$). The substrate, e.g., nucleic acid or surface, introduces movement constrains to the Cp dimers that are forming oligomers, making disassembly unfavorable and, hence, giving the reaction a substantial driving force toward assembly. Once a diamond pentamer is formed on the substrate, the assembly can proceed by its combination with other diamond pentamers and/or the sequential addition of dimers and other Cp oligomers.

protein concentrations. In more aggressive assembly conditions, high concentrations of dimer lead to kinetic bottlenecks and traps dominated by intermediates that are difficult to add to. Conversely, intermediates that are rapidly consumed in the assembly of larger complexes at high concentrations (and thus invisible to SAXS) can occupy a local energy minimum at low concentrations (HS-AFM conditions) that allows them to transiently accumulate. SAXS experiments allow us to examine conditions where we can identify kinetic bottlenecks, while HS-AFM has provided insight into reactions where there are no kinetic traps. Both are important for constraining models of assembly.

To conclude, we disclose previously unknown features of the HBV assembly mechanism, contributing to the more general understanding of the self-assembly of viruses. The binding study presented here provides a full characterization of the energetics involved in the first steps of HBV capsid assembly (Fig. 3), thereby setting a new quantitative reference to scrutinize the generic nucleation-and-growth pathway shown by other viruses (Table 1) (*19*, *48*–*50*). In follow-up studies, the subsequent steps of assembly until capsid closure, including the involved time scales, will be scrutinized to compare with complementary kinetic approaches (*21*, *51*, *52*). Last, the methodology presented here, combining fluorescence OT and HS-AFM, opens new possibilities to study other self-assembly processes of (bio) supramolecular complexes.

MATERIALS AND METHODS

Protein samples

Both Cp and the mutant Cp-Y132A were expressed and purified as previously described (*13*, *28*), based on the protocol reported by Wingfield *et al.* (*12*). Purified Cp capsids were stored in 50 mM tris-HCl, 5% sucrose, 5 mM EDTA, and 2 mM dithiothreitol (pH 7.5) at −80°C. Capsid disassembly and dimer purification were performed on the basis of the protocol described by Porterfield *et al.* (*13*). Briefly, Cp capsids typically between 2 and 4 mg/ml were disassembled by dilution into disassembly buffer [100 mM Hepes, 1.5 M guanidine HCl, 0.5 M LiCl, and 1 mM tris(2-carboxyethyl)phosphine (TCEP) (pH 7.5)], incubating for 16 hours at 4°C. *Escherichia coli* RNA encapsidated during protein expression was precipitated by the LiCl after centrifugation, 16,000*g* for 30 min at 4°C. Cp-Y132A is purified from *E. coli* as dimers and stored as ammonium acetate pellets at −20°C. Dimer solutions of Cp and Cp-Y132A were further purified by size exclusion chromatography using a Superose 6 Increase column (GE Life Sciences) equilibrated in disassembly buffer. Protein concentration was estimated by ultraviolet absorption using a Jasco V-750 spectrophotometer. Purified dimers were kept at 4°C for a few days in disassembly buffer to prevent assembly/aggregation.

Synthesis of HAP-Alexa488

HAP 13 (heteroaryldihydropyrimidine 13) was modified using Alexa Fluor 488 succinimidyl ester (Thermo Fisher Scientific) as previously described (*35*, *36*). HAP-Alexa488 stocks were dissolved in dimethyl sulfoxide and diluted in aqueous buffer immediately before use.

Triggering assembly conditions

Before OT-CFM measurements, Cp dimer solutions in disassembly buffer [100 mM Hepes, 1.5 M guanidine HCl, 0.5 M LiCl, and 1 mM TCEP (pH 7.5)] were diluted four times, using T-connectors, immediately before the sample was introduced into the microfluidic cell to ensure the dimeric state of Cp. A freshly diluted sample was always flushed into the flow cell before every measurement. Therefore, all experiments were performed in 25 mM Hepes, 375 mM guanidine HCl, 125 mM LiCl, and 250 μM TCEP (pH 7.5) (assembly buffer). In the case of HS-AFM experiments, dilutions were performed manually, and assembly buffer without LiCl was used as imaging buffer.

High-speed and conventional AFM

HS-AFM imaging was performed in amplitude modulation mode in liquid using a sample-scanning HS-AFM [Research Institute of Biomolecule Metrology Co., Ltd. (RIBM)] (*53*–*55*). Cantilevers USC-F1.2-k0.15 (NanoWorld) were used for imaging with a nominal spring constant of 0.15 N/m, a resonance frequency ≈ 0.6 MHz, and a quality factor ≈ 2. The cantilever-free amplitude is 3 nm, and the set-point amplitude for the cantilever oscillation was set at ~2.7 nm. Images were recorded from 0.2 to 3.5 frame/s. Measurements were performed in assembly buffer on freshly cleaved mica. Cp dimers were freshly diluted into assembly conditions at final concentrations between 100 and 300 nM before AFM imaging. All AFM experiments were performed by HS-AFM in liquid except for the image presented in fig. S4B, for which a NanoWizard AFM (JPK) was used. For the latter, imaging was performed in air in amplitude modulation mode using ScanAsyst Fluid+ cantilevers (Bruker) with a nominal spring constant of 0.7 N/m and resonance frequency of 150 kHz.

OT with CFM

A commercial setup (LUMICKS) combining dual-trap OT and three-color CFM was used (*22*, *56*). In addition, a five-channel microfluidic cell (LUMICKS) mounted on an automated *XY* stage was used to allow efficient movement of the traps through different solutions. Two DNA sequences were used in this study: bacteriophage λ DNA (Roche) and linearized pKYB1 vector (New England Biolabs), with lengths of 48,502 and 8370 bp, respectively. Both DNA sequences were biotinylated and used in combination with streptavidin-coated polystyrene beads (Spherotech) of 3.11 and 1.76 μm in diameter. Every pair of beads was calibrated before measurements to obtain the stiffness of the trap, typically between 0.2 and 0.5 pN/nm. dsDNA molecules were end labeled with biotin in the same strand, making them suitable to generate ssDNA through mechanical melting of dsDNA (*57*). 2D images and kymographs were constructed from CFM experiments by scanning the confocal volume along the area of interest and collecting the emission by single-photon avalanche photodiodes. Fluorescence intensities are expressed as count•10^3/s or photon•10^3/s (kHz).

Force spectroscopy for ssDNA-protein interaction

FECs were generated in assembly buffer, with or without protein present in the solution, for different short ssDNA molecules. Traces were obtained by moving one of the traps at constant speed (200 nm/s) between 4 and 50 pN. The traces that showed any kind of sudden change in contour length (L_C), due to assembly/disassembly events, or that did not reach the expected bead-to-bead distance for bare ssDNA at 50 pN were discarded for analysis (fig. S3). FECs were individually fitted with the eFJC model (*58*, *59*)

$$x = L_C\left[\coth\left(\frac{2FL_P}{k_BT}\right) - \frac{k_BT}{2FL_P}\right]\left(1 + \frac{F}{S}\right) \quad (1)$$

where x and F are the measured extension and force, respectively; L_C is the contour length; L_P is the persistence length; and S is the stretching modulus of the DNA molecule. To estimate eFJC parameters, FECs generated at 100 Hz were fitted to the eFJC model from 10 to 40 pN, as previously reported (*60*). Fitting was performed by a custom-written MATLAB script adapted from (*61*), keeping all three parameters free to float. Stretching and relaxing curves were analyzed separately for every ssDNA molecule, proving to be indistinguishable (fig. S3); hence, parameters extracted from the fits of forward and backward curves are presented without distinctions (tables S1 and S2).

Analysis of the protein-ssDNA saturation curves

The values of persistence length (L_P) obtained from single-molecule force spectroscopy experiments at different protein concentrations can be related to the degree of binding saturation (Θ) (*62*–*67*). Following the method developed by Farge *et al.* (*62*), FECs of protein-DNA complexes are assumed to be a linear combination of a bare DNA curve and a curve of DNA fully saturated with proteins. This allows Θ to be calculated from the experimental FECs using the expression

$$x_{\text{measured}} = \Theta x_{\text{coated}} + (1 - \Theta)\, x_{\text{naked}} \tag{2}$$

where x_{naked} represents the extension measured for bare DNA curves, x_{coated} is the extension of fully saturated DNA, and x_{measured} is the extension measured at any intermediate protein concentration (*62*). Inserting eFJC model (Eq. 1) in Eq. 2 allows for the direct calculation of Θ as function of the measured L_P values at different protein concentrations

$$\Theta = ({L_{pn}}^{-1} - {L_{pm}}^{-1})/({L_{pn}}^{-1} - {L_{pc}}^{-1}) \tag{3}$$

where L_{pn}, L_{pc}, and L_{pm} represent the persistence length of naked DNA, protein-coated DNA, and DNA at intermediate degrees of saturation, respectively. We notice that the expression presented here (Eq. 3), derived from the eFJC model, differs from the one obtained by Farge *et al.* (*62*) for the extensible worm-like chain model.

To access thermodynamic information from the experimental values of Θ, the isotherm described by the MGvH model (*32*) was used

$$\Theta = \frac{Cn}{K_{D,\text{int}}}(1 - \Theta) \left[\frac{(2\omega - 1)(1 - \Theta) + \Theta/n - R}{2(\omega - 1)(1 - \Theta)}\right]^{n-1} \left[\frac{1 - (n + 1)\Theta/n + R}{2(1 - \Theta)}\right]^{2} \tag{4}$$

where R is defined as

$$R = \sqrt{(1 - (n + 1)\Theta/n)^2 + \frac{4\omega\Theta}{n}(1 - \Theta)} \tag{5}$$

$K_{D,\text{int}}$ is the intrinsic dissociation constant for protein-DNA interaction, n is the binding footprint of the protein (expressed in number of nucleotides), ω is the cooperativity parameter and also the (unitless) equilibrium constant for protein-protein interaction, and C is the protein concentration. Last, combining Eqs. 3 and 4, the experimental curves L_P versus C can be analyzed to extract the binding parameters. Cp-ssDNA and Cp-Y132A-ssDNA saturation curves were fitted globally (Fig. 2). To increase fitting stability, n = 30 nt was kept fixed, as it has been already estimated (*26*). We notice that using slightly different values of n did not involve drastic changes in the reported binding parameters (table S3).

OT isotensional experiments

Real-time assembly experiments were performed by setting an active feedback to one of the optical traps to keep the load applied to the tether constant (force clamp). The force was set to 11 pN with a feedback frequency of 500 Hz. The change in bead-to-bead distance was monitored by bright-field imaging and tracking of the bead positions at 100 Hz. The conversion from distance to apparent DNA contour length (L_C^{ap}) was done by applying the eFJC model assuming that the persistence length (L_P = 1.085 nm), stretching modulus (S = 431 pN), and nucleotide-to-nucleotide distance of 0.53 nm were constant and equal to the ones measured by force spectroscopy experiments for bare DNA (see the "Force spectroscopy for ssDNA-protein interaction" section and tables S1 and S2). Assembly experiments were performed with concentrations of Cp dimers between 40 and 250 nM. The step size analysis was performed by fitting the experimental traces to a custom-written step-finding algorithm.

Assembly footprint and step size analysis

Assembly/disassembly steps were detected by applying a custom-written algorithm (MATLAB) to the real-time assembly traces. The algorithm is based on a moving window algorithm, where the mean of N data points within a window is compared to the average of N data points in an adjacent window. The difference in the found mean is compared with a threshold. This threshold is equal to $k{\cdot}\Delta\sigma$, where Δσ is the difference in SD for both windows, and k is a user-defined multiplication factor. This procedure is then repeated each time the two windows move one data point ahead until the windows reach the end of the experimental trace. The mean calculated from the data points in window 1 is compared to the mean of window 2, and when this difference is larger than the given threshold, it is considered a step. Experimental traces at 10 Hz were analyzed by applying the step-finding algorithm defining $k > 3$. The set of traces was also analyzed with the algorithm developed by Kerssemakers *et al.* (*68*), giving very similar results.

Size estimation of Cp oligomers by HAP-Alexa488 labeling

An ssDNA molecule was first incubated against different concentrations of Cp dimers (~1 min) in assembly buffer. Then, the formed ssDNA-Cp complex was incubated with a 100 nM HAP-Alexa488 solution containing no free Cp dimers for ~30 s. Last, CFM measurements were taken in a HAP-free region of the microfluidic cell to decrease the fluorescence background. The tension applied to the tether was always ≥20 pN, a load at which the Cp is not able to condense ssDNA. The brightness of HAP-Alexa488 was determined by analyzing only the last step of 115 photobleaching traces (fig. S6). In addition, fluorescence correlation spectroscopy with freely diffusing HAP-Alexa488 molecules was used to confirm brightness values under the same conditions as applied for the binding experiments (fig. S6). Oligomer size was estimated from the fluorescence signal of the beginning of the traces that showed the width of diffraction-limited spots (fig. S6).

HAP staining experiments show that Cp-Y132A fails to form many oligomers beyond dimers of dimers. This is consistent with the observation that Cp-Y132A lacks the ability to form stable precursors that lead to the assembly of full capsids (*28*, *69*).

Identification of dsDNA formation by SYTOX labeling

First, an ssDNA molecule was incubated against a solution of Cp dimers (1 to 2 min) in assembly buffer. Then, the formed ssDNA-Cp complex was brought to a solution of 250 nM SYTOX Orange (Thermo Fisher Scientific), where CFM measurements were performed (fig. S1). All ssDNA molecules were stretched to at least 30 pN to confirm the formation of stable dsDNA. The number of SYTOX molecules per dsDNA structure was determined by recording kymographs at 30 pN tension for 30 s and quantifying the mean fluorescence intensity (fig. S1). The brightness of SYTOX was directly obtained from kymographs of dsDNA at 25 pM SYTOX concentration (fig. S1).

Quantitative analysis of SYTOX experiments (Fig. 2D) showed that these dsDNA structures formed by Cp accommodate only a few SYTOX molecules at a time. We anticipate that Cp might be preventing SYTOX binding by steric hindrance; exclusion of SYTOX from potential sites has been shown for other DNA binding proteins (*70*). Moreover, it should be noted that two SYTOX molecules cannot bind to adjacent base pairs (*71*, *72*). Double-stranded regions are found to be in close contact with A-B dimers in an icosahedrally averaged structure (see fig. S10), which make up for half of the dimers of a full capsid. We anticipate that each capsid will contain many SYTOX-binding dsDNA regions. These data suggest that only a small fraction of the condensed DNA is available to SYTOX and/or that DNA folding continues long after the early assembly events occurred.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abg0811

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. C. V. Robinson, A. Sali, W. Baumeister, The molecular sociology of the cell. *Nature* **450**, 973–982 (2007).
2. P. Buzon, S. Maity, W. H. Roos, Physical virology: From virus self-assembly to particle mechanics. *WIREs Nanomedicine & Nanobiotechnology* **12**, e1613 (2020).
3. R. Zandi, B. Dragnea, A. Travesset, R. Podgornik, On virus growth and form. *Phys. Rep.* **847**, 1–102 (2020).
4. J. D. Perlmutter, M. F. Hagan, Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66**, 217–239 (2015).
5. W. H. Roos, R. Bruinsma, G. J. L. Wuite, Physical virology. *Nat. Phys.* **6**, 733–743 (2010).
6. P. Buzon, J. Ruiz-Sanz, J. C. Martinez, I. Luque, Stability, conformational plasticity, oligomerization behaviour and equilibrium unfolding intermediates of the Ebola virus matrix protein VP40. *J. Biomol. Struct. Dyn.* **38**, 4289–4303 (2020).
7. E. D. Holmstrom, D. Nettels, B. Schuler, Conformational plasticity of hepatitis C virus core protein enables RNA-induced formation of nucleocapsid-like particles. *J. Mol. Biol.* **430**, 2453–2467 (2018).
8. P. Lam, N. F. Steinmetz, Plant viral and bacteriophage delivery of nucleic acid therapeutics. *WIREs Nanomedicine & Nanobiotechnology* **10**, e1487 (2018).
9. G. L. Butterfield, M. J. Lajoie, H. H. Gustafson, D. L. Sellers, U. Nattermann, D. Ellis, J. B. Bale, S. Ke, G. H. Lenz, A. Yehdego, R. Ravichandran, S. H. Pun, N. P. King, D. Baker, Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).
10. L. M. Stannard, M. Hodgkiss, Morphological irregularities in Dane particle cores. *J. Gen. Virol.* **45**, 509–514 (1979).
11. R. A. Crowther, N. A. Kiselev, B. Bottcher, J. A. Berriman, G. P. Borisova, V. Ose, P. Pumpens, Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**, 943–950 (1994).
12. P. T. Wingfield, S. J. Stahl, R. W. Williams, A. C. Steven, Hepatitis core antigen produced in Escherichia coli: Subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry* **34**, 4919–4932 (1995).
13. J. Z. Porterfield, M. S. Dhason, D. D. Loeb, M. Nassal, S. J. Stray, A. Zlotnick, Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J. Virol.* **84**, 7174–7184 (2010).
14. J. M. Kenney, C.-H. von Bonsdorff, M. Nassal, S. D. Fuller, Evolutionary conservation in the hepatitis B virus core structure: Comparison of human and duck cores. *Structure* **3**, 1009–1019 (1995).
15. M. Nassal, The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J. Virol.* **66**, 4107–4116 (1992).
16. M. Nassal, Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* **134**, 235–249 (2008).
17. J. C. Wang, D. G. Nickens, T. B. Lentz, D. D. Loeb, A. Zlotnick, Encapsidated hepatitis B virus reverse transcriptase is poised on an ordered RNA lattice. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11329–11334 (2014).
18. A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, D. Endres, A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**, 14644–14652 (1999).
19. D. Endres, A. Zlotnick, Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**, 1217–1230 (2002).
20. M. F. Hagan, O. M. Elrad, Understanding the concentration dependence of viral capsid assembly kinetics—The origin of the lag time and identifying the critical nucleus size. *Biophys. J.* **98**, 1065–1074 (2010).
21. M. G. M. van Rosmalen, D. Kamsma, A. S. Biebricher, C. Li, A. Zlotnick, W. H. Roos, G. J. L. Wuite, Revealing in real-time a multistep assembly mechanism for SV40 virus-like particles. *Sci. Adv.* **6**, eaaz1639 (2020).
22. M. Marchetti, D. Kamsma, E. Cazares Vargas, A. Hernandez Garcia, P. van der Schoot, R. de Vries, G. J. L. Wuite, W. H. Roos, Real-time assembly of viruslike nucleocapsids elucidated at the single-particle level. *Nano Lett.* **19**, 5746–5753 (2019).
23. R. F. Garmann, A. M. Goldfain, V. N. Manoharan, Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22485–22490 (2019).
24. P. Kondylis, C. J. Schlicksup, N. E. Brunk, J. Zhou, A. Zlotnick, S. C. Jacobson, Competition between normative and drug-induced virus self-assembly observed with single-particle methods. *J. Am. Chem. Soc.* **141**, 1251–1260 (2019).
25. A. Valbuena, S. Maity, M. G. Mateu, W. H. Roos, Visualization of single molecules building a viral capsid protein lattice through stochastic pathways. *ACS Nano* **14**, 8724–8734 (2020).
26. M. S. Dhason, J. C. Wang, M. F. Hagan, A. Zlotnick, Differential assembly of hepatitis B virus core protein on single- and double-stranded nucleic acid suggest the dsDNA-filled core is spring-loaded. *Virology* **430**, 20–29 (2012).
27. A. Ben-Shaul, W. M. Gelbart, Viral ssRNAs are indeed compact. *Biophys. J.* **108**, 14–16 (2015).
28. C. R. Bourne, S. P. Katen, M. R. Fulz, C. Packianathan, A. Zlotnick, A mutant hepatitis B virus core protein mimics inhibitors of icosahedral capsid self-assembly. *Biochemistry* **48**, 1736–1742 (2009).
29. C. Bustamante, Y. R. Chemla, N. R. Forde, D. Izhaky, Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **73**, 705–748 (2004).
30. J. R. Moffitt, Y. R. Chemla, K. Aathavan, S. Grimes, P. J. Jardine, D. L. Anderson, C. Bustamante, Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446–450 (2009).
31. J. C. Wang, M. S. Dhason, A. Zlotnick, Structural organization of pregenomic RNA and the carboxy-terminal domain of the capsid protein of hepatitis B virus. *PLOS Pathog.* **8**, e1002919 (2012).
32. J. D. McGhee, P. H. von Hippel, Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* **86**, 469–489 (1974).
33. P. Ceres, A. Zlotnick, Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525–11531 (2002).
34. R. Asor, L. Selzer, C. J. Schlicksup, Z. Zhao, A. Zlotnick, U. Raviv, Assembly reactions of hepatitis b capsid protein into capsid nanoparticles follow a narrow path through a complex reaction landscape. *ACS Nano* **13**, 7610–7626 (2019).
35. S. Nair, L. Li, S. Francis, W. W. Turner, M. VanNieuwenhze, A. Zlotnick, Use of a fluorescent analogue of a HBV core protein-directed drug to interrogate an antiviral mechanism. *J. Am. Chem. Soc.* **140**, 15261–15269 (2018).
36. A. Zlotnick, L. Li, M. S. van Nieuwenzhe, W. W. Turner, Fluorescent HAP: A diagnostic stain for HBV cores in cells. US Patent No. 10006913. (2018).
37. C. J. Schlicksup, J. C.-Y. Wang, S. Francis, B. Venkatakrishnan, W. W. Turner, M. VanNieuwenhze, A. Zlotnick, Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. *eLife* **7**, e31473 (2018).
38. Z. Zhou, T. Hu, X. Zhou, S. Wildum, F. Garcia-Alcalde, Z. Xu, D. Wu, Y. Mao, X. Tian, Y. Zhou, F. Shen, Z. Zhang, G. Tang, I. Najera, G. Yang, H. C. Shen, J. A. Young, N. Qin, Heteroaryldihydropyrimidine (HAP) and sulfamoylbenzamide (SBA) inhibit hepatitis B virus replication by different molecular mechanisms. *Sci. Rep.* **7**, 42374 (2017).
39. C. R. Bourne, M. G. Finn, A. Zlotnick, Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. *J. Virol.* **80**, 11055–11061 (2006).

40. W. Wu, N. R. Watts, N. Cheng, R. Huang, A. C. Steven, P. T. Wingfield, Expression of quasi-equivalence and capsid dimorphism in the Hepadnaviridae. *PLoS Comput. Biol.* **16**, e1007782 (2020).
41. T. H. Chu, A. T. Liou, P. Y. Su, H. N. Wu, C. Shih, Nucleic acid chaperone activity associated with the arginine-rich domain of human hepatitis B virus core protein. *J. Virol.* **88**, 2530–2543 (2014).
42. C. A. Lutomski, N. A. Lyktey, E. E. Pierson, Z. Zhao, A. Zlotnick, M. F. Jarrold, Multiple pathways in capsid assembly. *J. Am. Chem. Soc.* **140**, 5784–5790 (2018).
43. C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. van Duijn, A. J. Heck, Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nat. Chem.* **3**, 126–132 (2011).
44. R. Asor, C. J. Schlicksup, Z. Zhao, A. Zlotnick, U. Raviv, Rapidly forming early intermediate structures dictate the pathway of capsid assembly. *J. Am. Chem. Soc.* **142**, 7868–7882 (2020).
45. A. Zlotnick, To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59–67 (1994).
46. P. Moisant, H. Neeman, A. Zlotnick, Exploring the paths of (virus) assembly. *Biophys. J.* **99**, 1350–1357 (2010).
47. L. Ludgate, K. Liu, L. Luckenbaugh, N. Streck, S. Eng, C. Voitenleitner, W. E. Delaney IV, J. Hu, Cell-free hepatitis B virus capsid assembly dependent on the core protein C-terminal domain and regulated by phosphorylation. *J. Virol.* **90**, 5830–5844 (2016).
48. H. D. Nguyen, V. S. Reddy, C. L. Brooks, Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett.* **7**, 338–344 (2007).
49. J. D. Perlmutter, M. R. Perkett, M. F. Hagan, Pathways for virus assembly around nucleic acids. *J. Mol. Biol.* **426**, 3148–3165 (2014).
50. R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, H. Reiss, Classical nucleation theory of virus capsids. *Biophys. J.* **90**, 1939–1948 (2006).
51. M. Chevreuil, D. Law-Hine, J. Chen, S. Bressanelli, S. Combet, D. Constantin, J. Degrouard, J. Moller, M. Zeghal, G. Tresset, Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nat. Commun.* **9**, 3071 (2018).
52. R. F. Bruinsma, G. J. L. Wuite, W. H. Roos, Physics of viral dynamics. *Nat. Rev. Phys.* **3**, 76–91 (2021).
53. S. Maity, C. Caillat, N. Miguet, G. Sulbaran, G. Effantin, G. Schoehn, W. H. Roos, W. Weissenhorn, VPS4 triggers constriction and cleavage of ESCRT-III helical filaments. *Sci. Adv.* **5**, eaau7198 (2019).
54. T. Uchihashi, N. Kodera, T. Ando, Guide to video recording of structure dynamics and dynamic processes of proteins by high-speed atomic force microscopy. *Nat. Protoc.* **7**, 1193–1206 (2012).
55. S. Maity, J. Ottelé, G. M. Santiago, P. W. J. M. Frederix, P. Kroon, O. Markovitch, M. C. A. Stuart, S. J. Marrink, S. Otto, W. H. Roos, Caught in the act: Mechanistic insight into supramolecular polymerization-driven self-replication from real-time visualization. *J. Am. Chem. Soc.* **142**, 13709–13717 (2020).
56. M. Hashemi Shabestari, A. E. C. Meijering, W. H. Roos, G. J. L. Wuite, E. J. G. Peterman, Recent advances in biological single-molecule applications of optical tweezers and fluorescence microscopy. *Methods Enzymol.* **582**, 85–119 (2017).
57. A. Candelli, T. P. Hoekstra, G. Farge, P. Gross, E. J. Peterman, G. J. Wuite, A toolbox for generating single-stranded DNA in optical tweezers experiments. *Biopolymers* **99**, 611–620 (2013).
58. C. Bustamante, S. B. Smith, J. Liphardt, D. Smith, Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* **10**, 279–285 (2000).
59. S. B. Smith, Y. Cui, C. Bustamante, Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
60. A. Bosco, J. Camunas-Soler, F. Ritort, Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.* **42**, 2064–2074 (2014).
61. O. D. Broekmans, G. A. King, G. J. Stephens, G. J. Wuite, DNA twist stability changes with magnesium(2+) concentration. *Phys. Rev. Lett.* **116**, 258102 (2016).
62. G. Farge, N. Laurens, O. D. Broekmans, S. M. van den Wildenberg, L. C. Dekker, M. Gaspari, C. M. Gustafsson, E. J. Peterman, M. Falkenberg, G. J. Wuite, Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nat. Commun.* **3**, 1013 (2012).
63. G. A. King, M. Hashemi Shabestari, K. H. Taris, A. K. Pandey, S. Venkatesh, J. Thilagavathi, K. Singh, R. Krishna Koppisetti, D. Temiakov, W. H. Roos, C. K. Suzuki, G. J. L. Wuite, Acetylation and phosphorylation of human TFAM regulate TFAM-DNA interactions via contrasting mechanisms. *Nucleic Acids Res.* **46**, 3633–3642 (2018).
64. M. J. McCauley, E. M. Rueter, I. Rouzina, L. J. Maher III, M. C. Williams, Single-molecule kinetics reveal microscopic mechanism by which high-mobility group B proteins alter DNA flexibility. *Nucleic Acids Res.* **41**, 167–181 (2013).
65. D. Murugesapillai, M. J. McCauley, L. J. Maher III, M. C. Williams, Single-molecule studies of high-mobility group B architectural DNA bending proteins. *Biophys. Rev.* **9**, 17–40 (2017).
66. M. S. Rocha, Revisiting the neighbor exclusion model and its applications. *Biopolymers* **93**, 1–7 (2010).
67. M. S. Rocha, Extracting physical chemistry from mechanics: A new approach to investigate DNA interactions with drugs and proteins in single molecule experiments. *Integr. Biol.* **7**, 967–986 (2015).
68. J. W. Kerssemakers, E. L. Munteanu, L. Laan, T. L. Noetzel, M. E. Janson, M. Dogterom, Assembly dynamics of microtubules at molecular resolution. *Nature* **442**, 709–712 (2006).
69. C. Packianathan, S. P. Katen, C. E. Dann III, A. Zlotnick, Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J. Virol.* **84**, 1607–1615 (2010).
70. A. E. C. Meijering, A. S. Biebricher, G. Sitters, I. Brouwer, E. J. G. Peterman, G. J. L. Wuite, I. Heller, Imaging unlabeled proteins on DNA with super-resolution. *Nucleic Acids Res.* **48**, e34 (2020).
71. A. S. Biebricher, I. Heller, R. F. Roijmans, T. P. Hoekstra, E. J. Peterman, G. J. Wuite, The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics. *Nat. Commun.* **6**, 7304 (2015).
72. H. Ihmels, D. Otto, Intercalation of organic dye molecules into double-stranded DNA—Generalprinciples and recent developments, in *Supermolecular Dye Chemistry*, F. Würthner, Ed. (Springer Berlin Heidelberg, 2005), pp. 161–204.
73. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
74. L. He, Z. Porterfield, P. van der Schoot, A. Zlotnick, B. Dragnea, Hepatitis virus capsid polymorph stability depends on encapsulated cargo size. *ACS Nano* **7**, 8447–8454 (2013).

Acknowledgments: We thank G. King and M. van Rosmalen (Vrije Universiteit) for input on experimental and analysis procedures. **Funding:** This work was supported by a VIDI grant and an STW Perspectief grant "Cancer-ID" (project 14192) to W.H.R. from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) and an MSCA Individual fellowship (INTERACT 751404) to S.M., and the work from the Zlotnick laboratory is funded by NIH R01-AI144022. **Author contributions:** P.B. designed and performed the OT-CFM and AFM experiments, analyzed and interpreted the data, and wrote the initial draft of the paper. S.M. performed and analyzed the HS-AFM experiments. P.C. and M.J.W. helped produce and analyze the OT-CFM data. S.D., C.K., and A.Z. provided protein and HAP samples. P.B., G.J.L.W., A.Z., and W.H.R. wrote the manuscript. G.J.L.W., A.Z., and W.H.R. conceived the study. W.H.R. supervised the study. **Competing interests:** G.J.L.W. declares financial interest in the fluorescence OT setup used in this work, patented and licensed to LUMICKS B.V. A.Z. is a founder of Assembly Biosciences and a founder and officer of Door Pharmaceuticals. A.Z. has equity in both companies. These are biotech companies that are focused on developing antivirals. All the other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The HBV proteins can be provided by the Zlotnick laboratory pending scientific review. Requests for the material should be sent to A.Z., Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN 47405-7003.

Submitted 9 December 2020
Accepted 14 September 2021
Published 3 November 2021
10.1126/sciadv.abg0811

Virus self-assembly proceeds through contact-rich energy minima

Pedro BuzónSourav MaityPanagiotis ChristodoulisMonique J. WiertsemaSteven DunkelbargerChristine KimGijs J.L. WuiteAdam ZlotnickWouter H. Roos

Sci. Adv., 7 (45), eabg0811. • DOI: 10.1126/sciadv.abg0811