ACS NANO

www.acsnano.org

ARTICLE

Multiscale Modeling of Hepatitis B Virus Capsid Assembly and Its Dimorphism

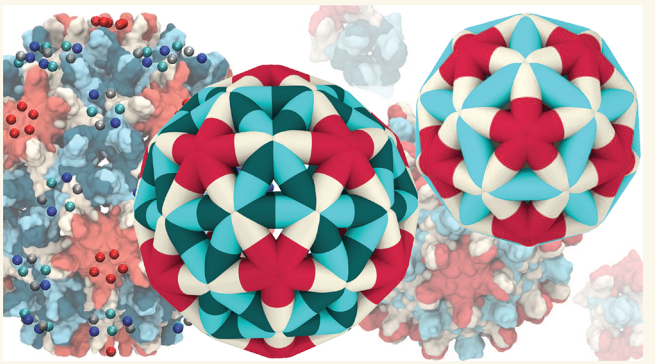
Farzaneh Mohajerani, Botond Tyukodi, Christopher J. Schlicksup, Jodi A. Hadden-Perilla, Adam Zlotnick, and Michael F. Hagan*

Cite This: *ACS Nano* 2022, 16, 13845−13859

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: **Hepatitis B virus (HBV) is an endemic, chronic virus that leads to 800000 deaths per year. Central to the HBV lifecycle, the viral core has a protein capsid assembled from many copies of a single protein. The capsid protein adopts different (quasi-equivalent) conformations to form icosahedral capsids containing 180 or 240 proteins: $T = 3$ or $T = 4$, respectively, in Caspar−Klug nomenclature. HBV capsid assembly has become an important target for recently developed antivirals; nonetheless, the assembly pathways and mechanisms that control HBV dimorphism remain unclear. We describe computer simulations of the HBV assembly, using a coarse-grained model that has parameters learned from all-atom molecular dynamics simulations of a complete HBV capsid and yet is computationally tractable. Dynamical simulations with the resulting model reproduce experimental observations of HBV assembly pathways and products. By constructing Markov state models and employing transition path theory, we identify pathways leading to $T = 3$, $T = 4$, and other experimentally observed capsid morphologies. The analysis shows that capsid polymorphism is promoted by the low HBV capsid bending modulus, where the key factors controlling polymorphism are the conformational energy landscape and protein−protein binding affinities.**



KEYWORDS: *Hepatitis B virus, viral capsid, self-assembly, antivirals, computational modeling, multiscale simulations*

During the lifecycles of many viruses, hundreds of protein subunits self-assemble into a protein capsid that packages the viral nucleic acid and delivers it to a new host cell. In many cases a particular capsid structure is required for infectivity, and capsid proteins form this structure with high fidelity. Yet, capsid proteins can also exhibit striking polymorphism and adaptability, assembling into structures with different sizes and morphologies in response to changes in solution conditions, the presence of antiviral drugs, or encapsulation of nucleic acids or nanoparticles of varying sizes. Elucidating the factors that control assembly pathways and products could identify important targets for antiviral drugs and would advance our fundamental understanding of the viral lifecycle.

Approximately half of the characterized virus families have icosahedral symmetry. That is, their capsids are comprised of 60 identical asymmetric units. Caspar and Klug (CK) showed how multiples of 60 proteins can form icosahedral capsids, by subtriangulating each asymmetric unit so that individual proteins are forced to adopt slightly different (quasi-equivalent) conformations.[1−3] An icosahedral capsid has $60T$ subunits, where the "triangulation number" T specifies the number of different protein conformations and is restricted to certain integer values ($T = h^2 + hk + k^2$, where h and k are non-negative integers).[2]

Hepatitis B virus (HBV) is an important example of self-assembly as a model system, and since chronic HBV infection is a serious public health issue, assembly of its capsid is a target for a recently developed generation of antivirals that may contribute to a cure.[6] About 300 million people have chronic HBV, which contributes, by cirrhosis, liver failure, and liver cancer, to about 800000 deaths each year.[7] Within an infected host cell, the HBV capsid protein assembles around the HBV pregenomic RNA and viral polymerase to form a core particle. While capsid structures with $T = 4$ symmetry (comprising 120 dimer protein subunits) are appropriate to accommodate the

Received: March 1, 2022
Accepted: August 8, 2022
Published: September 2, 2022

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

genome, many secreted particles are empty (containing no RNA)[8] and a small fraction (~5%) have $T = 3$ capsids (with 90 dimers).[9,10] This behavior can be recapitulated *in vitro* using the assembly domain (residues 1−149 of the capsid protein, Cp149) in the absence of nucleic acids, which leads to a similar mix of $T = 4$ and $T = 3$ particles (all of which are empty). Moreover, several classes of small-molecule antiviral agents referred to as "core protein allosteric modulators" (CpAMs) have been identified that can bind to HBV capsid proteins and drive inappropriate assembly, leading to failure to package RNA or redirection of the assembly to alternative non-infectious products.[11−18]

Recently, Cp149 assembly size distributions have been measured near to or at single-subunit precision using resistive pulse sensing,[19,20] mass spectrometry,[21] and charge detection mass spectrometry,[22−25] and small intermediates were identified by high-speed AFM.[26] These experiments identified a number of long-lived intermediate structures with sizes between 85 and 140 dimers. Notably, some pathways also include "overgrown" intermediates with more than 120 dimers, which eventually dissociate some dimers and rearrange into icosahedral capsids.[27] Similarly, small-angle X-ray scattering (SAXS) experiments identified long-lived complexes at sizes between 90 and 120 dimers,[28] as well as a three-phase assembly kinetics, with nucleation and elongation phases followed by a slow rearrangement of capsid subunits into icosahedral structures.[29]

Despite these experimental advances, the mechanisms underlying HBV dimorphism and the factors that separate assembly pathways leading to $T = 4$, $T = 3$, or asymmetric products remain unclear.

Computational models can reveal details of the assembly process that are not accessible to experiments. However, while atomistic and near-atomistic simulations have revealed the dynamics of complete viruses,[4,30−33] the long time scales (milliseconds to hours) prohibit simulating capsid assembly with atomic resolution, except for specific steps.[34,35] Therefore, coarse-grained models have been used to study the assembly dynamics of icosahedral shells.[36−48] Of particular relevance to our work, elastic interactions between 5-fold defects can funnel the elastic energy landscape of an assembling crystalline shell toward icosahedral structures.[49−52] However, previous works on empty capsids with $T \geq 4$ have focused on material properties (parametrized by the dimensionless ratio of stretching to bending moduli, Föppl von Kármán number) that are orders of magnitude below those of virus capsids. Thus, different mechanisms may be important for HBV assembly.

A significant challenge of coarse-grained (CG) models is linking their predictions to specific experimental systems. It is therefore desirable to inform CG models with atomic-resolution information. For example, multiscale modeling approaches have elucidated the stability and dynamics of viral capsids[30,53−55] and other protein structures (e.g., refs 56−60), and CG models based on protein structures have led to insights about specific virus assembly processes (e.g., refs 61−63).

Here, we employ data from atomic-resolution simulations and solution experiments to constrain and build a multiscale model for HBV capsid assembly. We start from a minimal CG model that captures the geometric features of HBV capsid proteins and their quasi-equivalent conformations. We parametrize this model using data from all-atom molecular (AA) dynamics simulations of a complete HBV capsid, along with atomic-resolution structures of the proteins in different assembly morphologies. We focus on a small number of parameters: binding affinity, subunit length, subunit dihedral angle, binding angles, and associated stretching and bending moduli. These data provide estimates for the relative strengths of the different types of protein−protein interactions that drive HBV assembly, the fluctuations of proteins around their mean interaction geometries, and the corresponding larger-scale elastic properties of the assembled structures.

We then perform dynamic Monte Carlo (MC) simulations to simulate dynamical assembly trajectories and compare these to experimentally observed trajectories. Analyzing these computed trajectories using Markov state model and transition path theory analysis explains the origins of trapped or metastable species observed in experiments.[22−25,28,29,64] These results provide the confidence to examine the early steps of the reaction, where the diversity of intermediates shows how nucleation can be described as a much more complex series of intermediates than could possibly be isolated in experiments. We identify pathway "hubs", or intermediates from which pathways diverge toward $T = 4$, $T = 3$, or asymmetric assembly products. Pathway analysis isolates factors, such as the local symmetry and relative stability of associating subunits, that select which pathway emerges from a hub state. The analysis also shows that the key factors controlling polymorphism are the elastic moduli and the extent of conformation dependence of protein−protein binding affinities. For bending modulus values relevant to virus capsids ($\sim(40-400)k_BT$), the elastic strain alone cannot guide assembly toward large symmetric structures, and the robust assembly of icosahedral capsids with sizes $T \geq 4$ strongly depends on the presence of multiple protein conformations and their conformation-dependent binding affinities. We find that such specificity of interactions plays a key role in the assembly of natural homopolymeric icosahedral capsids by reducing the number of accessible assembly pathways and, thus, guiding assembly toward the target geometry. Finally, the results in the paper are qualitatively applicable to other viruses, and the approach is generalizable.

RESULTS AND DISCUSSION

Overview of Model and Parameters. Capsid protein dimers are the basic assembly subunits for HBV.[65,66] Based on the quasi-equivalent conformations of the monomers (denoted as A−D), there are two dimer conformations in the $T = 4$ capsid (AB and CD; see Figure 1A), which each make four interdimer interfaces. Dimer−dimer binding is primarily driven by hydrophobic stereospecific contacts at these interfaces. In each such interaction, a monomer of one dimer is capped by a monomer of a second dimer, with the second fitting against a hydrophobic patch on the first. For example, in the D−B contact shown in Figure 1A, the D monomer in the CD dimer is capped by the B monomer in the AB dimer. These interactions are asymmetric—the inverse, a B−D contact with B of an AB dimer capped by D of a CD dimer, is structurally different from a D−B contact and, thus, has a different binding affinity.

We construct a CG model that enables simulating experimentally relevant time scales as follows. Since HBV dimers assemble with locally triangular lattices,[65,66] the model represents an assembling capsid by an elastic triangular sheet, with edges corresponding to dimers. Dimer diffusion from the

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

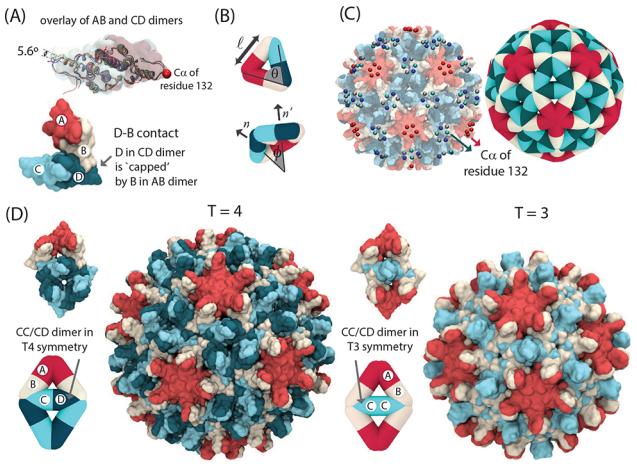


Figure 1. Description of the multiscale model. (A) Overlay of AB and CD conformations of the capsid protein from the crystal structure of the $T = 4$ HBV capsid (top). Example of a dimer−dimer contact in the $T = 4$ HBV capsid, where the D monomer in a CD dimer is "capped" by a B monomer in an AB dimer (bottom). Colors correspond to different monomers in each dimer. (B) Each edge in the elastic sheet coarse-grained model represents an HBV dimer. Model parameters include the equilibrium edge length l_c^0 and associated force constant κ_l, which controls the shell stretching modulus, the equilibrium binding angle between two edges $\theta_{c,c'}^0$ and associated force constant κ_θ, and the equilibrium dihedral angle between the normals of the two adjacent triangles ϕ_c^0 and force constant κ_ϕ, which sets the capsid bending modulus. Here c and c' are the conformations of a given edge and the dimer that it interacts with, respectively. (C) Model parameters are optimized by minimizing the difference between the probability distributions of fluctuations of edge lengths, edge−edge angles, and dihedral angles computed from the coarse-grained (CG) model and an all-atom (AA) molecular dynamics simulation of a complete $T = 4$ HBV capsid in explicit water.[4] We coarse-grain the data from the all-atom simulation by selecting the C-α atom of residue 132 of each monomer and clustering these 240 points to 5-fold and 6-fold vertices based on proximity using the scikit DBSCAN clustering algorithm.[5] The center of mass of each cluster is then assigned to a vertex of the CG model capsid (see Figure S3 and Movie S1 in the Supporting Information). (D) The configurations of dimers in a $T = 4$ capsid (left) and in a $T = 3$ capsid (right). A CD/CC dimer in $T = 4$ has two interactions with CD dimers and two interactions with AB dimers and is referred to a CD dimer. A CD/CC dimer in a $T = 3$ capsid, which has 4 interactions with AB dimers, is referred to as a CC dimer (see Model section and the Supporting Information).

bulk and association (dissociation) to (from) capsid structures are represented by the reversible addition/deletion of edges. Thus, the model and dynamics coarse-grain over subunit diffusion and high-frequency fluctuations of interacting dimers. The approach is similar to existing models for a capsid and tubule assembly based on triangular subunits of a single type,[52,67−69] but in our model growth occurs via addition/deletion of edges (dimers), and the model accounts for dimer asymmetry, as well as how the dimer structure and interactions depend on their conformational state (see the Model section, Figure 1, and Figure S6 in the Supporting Information). We see below that these physical features are essential to achieve robust assembly of $T = 4$ shells when accounting for the material properties of HBV proteins.

We consider two edge types with different equilibrium lengths and dihedral angles (Figure 1B and Figure S6 in the Supporting Information). The first type represents AB dimers present in $T = 4$ and $T = 3$ capsids,[70] while the second, CD/CC, represents both CD and CC dimers in $T = 4$ and $T = 3$ capsids, respectively. This choice is based on the similarity of the structure and local symmetry environment (i.e., set of conformations of neighboring subunits) of CD and CC dimers in $T = 4$ and $T = 3$ capsids. For identifying assembly pathways, we will denote such a dimer as CD (CC) if its neighboring subunits are consistent with the $T = 4$ ($T = 3$) local symmetry (see Figure 1D and Figure S6 in the Supporting Information).

To model dilute, noninteracting capsid structures typical of productive capsid assembly reactions,[71,72] each simulation considers a single assembling capsid. The dynamics, including dimer association, obeys microscopic reversibility. The addition/deletion events occur with probability according to the grand canonical ensemble with a fixed chemical potential, which sets the bulk dimer concentration C_{tot}.[67,69] Technical details are provided in the Model section and Model Details in the Supporting Information.

Optimizing CG Model Parameters against AA Simulations Estimates Protein−Protein Interaction Geometries and Elastic Moduli. The model potential energy function includes terms that represent subunit−subunit binding interactions and the elastic energy associated with deviations from the ground-state sheet geometry, corresponding to harmonic potentials for stretching of edges, deviations of dimer−dimer binding angles, and intradimer strains associated with deviation of dihedral angles between pairs of triangular faces (see Methods and eq 2). Denoting a dimer conformation for edge i as $c(i)$, the key model parameters are the set of equilibrium edge lengths $\{l_c^0\}$, dimer−dimer binding angles, and dihedral angles $\{\theta_{c,c'}^0\}$ and $\{\phi_c^0\}$ (Figure 1B). The associated moduli κ_l, κ_θ, and κ_ϕ are independent of edge conformation, with κ_l and κ_ϕ setting the continuum-limit stretching and bending moduli of the shell (see Model Details in the Supporting Information).

To estimate the model parameter values that correspond to the HBV dimers, we mapped our CG dimers to an AA structure of an HBV capsid (see Methods, Figure 1C and Movie S1). We then used AA simulations of a complete HBV capsid[4] to optimize the model parameters. We estimated the equilibrium edge lengths, dihedral and binding angles, and associated stretching and bending moduli of the CG model by minimizing the difference between the distributions of edge lengths, dihedral angles, and binding angles obtained from the AA MD and CG MC simulations (see Methods and Figure S8). Optimization against the AA data resulted in high values for the stretching modulus $\kappa_l \approx 4200 k_B T/\sigma^2$ (with the simulation unit length scale $\sigma \approx 8$ nm) and binding angle modulus $\kappa_\theta \approx 800 k_B T$ and a relatively low value for the bending modulus $\kappa_\phi \approx 40 k_B T$. The 2D Young's modulus $\varepsilon \approx 6000$ can be estimated from $\varepsilon = \kappa_l + \kappa_\theta/l_0^2$, where $l_0 \approx \sigma$ is the average dimer length.

Mapping the three bending and elastic moduli to the standard Helfrich elastic energy of an elastic sheet[73,74] estimates the Föpple-von Kármán number, FvK $= \varepsilon R_{cap}^2/\kappa_\phi \approx 500$ with $R_{cap} \approx 2.1\sigma$, the radius of the HBV capsid, which is consistent with results of nanoindentation measurements.[75] As shown below, using these optimized elastic moduli in our model results in dynamical assembly behaviors that are consistent with experiments on HBV protein assembly, including the distribution of $T = 4$ and $T = 3$ icosahedral capsids at optimal assembly conditions, as well as large

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

aberrant structures with reduced curvature observed under nonoptimal conditions (including in the presence of CpAMs that strengthen dimer−dimer interactions[76]).

Binding Affinities and Conformational Free Energies. Experiments identify mean dimer−dimer binding affinity values for $T = 4$ capsids in the range $-7.0 \leq g^{\text{bind}}_{T=4} \leq -4.2k_BT$ (with $k_BT \approx 0.6$ kcal/mol at $T = 300$ K) for different solution conditions and temperatures.[65] To estimate the binding energies at each quasi-equivalent site, we set the CD binding affinity $g^{\text{bind}}_{\text{C-D}}$ as a reference value g^{bind}_0 and set the relative values of binding affinities based on their relative buried surface areas (using PDBePISA, Table 1). We define

Table 1. Relative Dimer−Dimer Binding Affinities[a]

	$\bar{g}^{\text{bind}}_{\text{A−A}}$	$\bar{g}^{\text{bind}}_{\text{B−C}}$	$\bar{g}^{\text{bind}}_{\text{C−D}}$	$\bar{g}^{\text{bind}}_{\text{D−B}}$	$\bar{g}^{\text{bind}}_{\text{D−C}}$
2G33 $T = 4$ Capsid[16]	1.434	1.247	1	0.854	0.337
6UI7 $T = 4$ Capsid[70]	1.282	1.127	1	0.936	NA
CG model	1.3	1.1	1	0.9	0.2

[a]Relative binding affinities between different dimers in our CG model are set based on the interdimer interface analysis of HBV $T = 4$ capsid (in PDBePISA[101,102]).

$\bar{g}^{\text{bind}}_{m-m'} \equiv g^{\text{bind}}_{m-m'}/g^{\text{bind}}_0$ with m, m' as the conformations of the two interacting monomers. Table 1 shows these estimates for $\bar{g}^{\text{bind}}_{\text{A−A}}$, $\bar{g}^{\text{bind}}_{\text{B−C}}$, $\bar{g}^{\text{bind}}_{\text{D−B}}$ and $\bar{g}^{\text{bind}}_{\text{C−D}}$ (the dimer−dimer contacts that are present in the HBV $T = 4$ capsid), as well as $\bar{g}^{\text{bind}}_{\text{D−C}}$ (only observed in nonicosahedral structures). We find that simulated assembly behaviors match those observed in experiments with these estimated values, and deviating from the PDBePISA estimates leads to poor agreement. Specifically, the assembly morphology is very sensitive to the value of $\bar{g}^{\text{bind}}_{\text{D−C}}$. The D−C contacts do not occur within icosahedral structures but are a prominent feature of flat hexagonal sheets that occur for nonassembly-competent conditions[77,78] (Figure 4C,D). Our simulations predict the optimal value for this parameter $\bar{g}^{\text{bind}}_{\text{D−C}} = 0.2$ (close to the PDBePISA estimate of ≈0.3) to be much less than the affinities of all the contacts in a $T = 4$ capsid(Figure 4(C-D)).

The model also specifies the equilibrium population distribution of the two dimer conformations AB and CD/CC according to $K_{\text{AB-CD}} \equiv \frac{[\text{AB}]}{[\text{CD}]} = e^{-\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}/k_BT}$, where $K_{\text{AB-CD}}$ is the equilibrium constant for interconversion between the AB and CD conformations and $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$ is the corresponding free energy difference between the two conformations. As noted above, we consider the CD and CC conformations to be equivalent, i.e. $K_{\text{CC-CD}} \equiv \frac{[\text{CC}]}{[\text{CD}]} = 1$, based on the high degree of structural similarity between CD and CC conformations and reduction in the number of model parameters. The parameter $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$ has not been directly estimated, but our results described below suggest that it depends on ionic strength, as previously suggested for the interconversion of HBV capsid protein between conformations that are active or inactive for assembly.[65,79] Experiments suggest that $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$ includes additional entropic contributions, such as different degrees of disorder in the C-terminal region of the assembly domain (Cp149) for the different conformations.[80]

To summarize, we consider two control parameters: g^{bind}_0 (which controls the mean interdimer binding affinity) and the intradimer conformational free energy $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$. To enable direct comparison against experiments, we present our results in terms of the mean value of the dimer−dimer binding affinity (per dimer−dimer contact) in a $T = 4$ capsid, which depends on the binding affinities and conformational free energy provided in eq 8 of the Model section. With the binding energy choices from Table 1, the mean dimer−dimer binding affinity in eq 8 can be written as

$$g^{\text{bind}}_{T=4} = (4.3g^{\text{bind}}_0 + \Delta g^{\text{conf}}_{\text{CD}\to\text{AB}})/4 \quad (1)$$

Assembly Products. *The Model Reproduces HBV Polymorphism.* Figure 2A−C shows how the distribution of

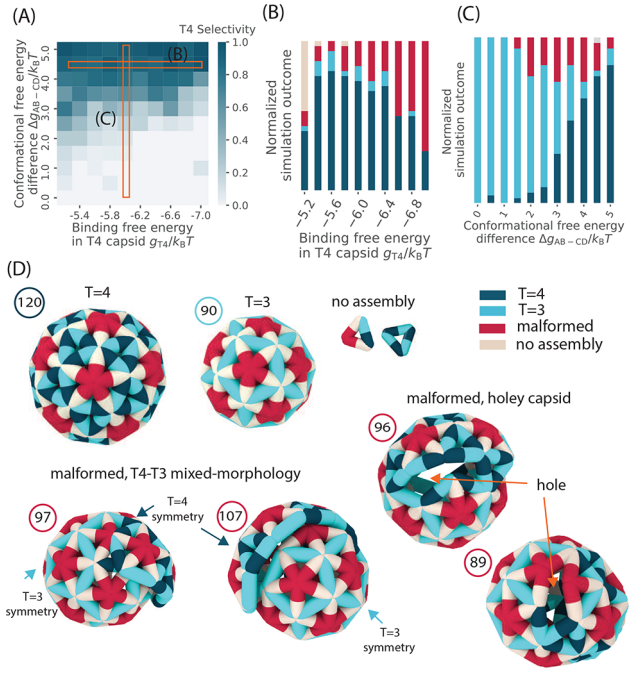


Figure 2. Dependence of assembly product morphologies on binding affinities and the intradimer conformational free energy landscape. (A) Dependence of selectivity for $T = 4$ capsids on the mean dimer−dimer binding affinity in a $T = 4$ capsid ($g^{\text{bind}}_{T=4}$, eq 1) and the conformational equilibrium between AB and CD dimers, parametrized by $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}} = -k_BT \log K_{\text{AB-CD}}$. The selectivity is defined as the ratio of the number of complete $T = 4$ capsids to the number of all closed shells. (B, C) The fraction of assembly product morphologies as a function of (B) the dimer−dimer binding affinity for fixed $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}} = 3.5k_BT$ and (C) the AB/CD conformational equilibrium for fixed $g^{\text{bind}}_{T=4} = -5.8k_BT$. The four categories of product morphologies are shown in (D). Capsids with $T = 4$ and $T = 3$ icosahedral symmetry have 120 and 90 dimers, respectively, while most shells with mixed morphology have sizes between 85 and 140 dimers. (D) Snapshots of the three categories of assembly products are shown: $T = 4$, $T = 3$, and long-lived mixed-morphology capsids that have two distinct parts in $T = 4$ and $T = 3$ symmetries. Long-lived mixed morphologies (left) fail to close due to incompatible curvatures of the two morphologies, resulting in open boundaries that are sterically blocked from addition of new dimers (see Figure S2 in the Supporting Information). Holey capsids (right) frequently form under aggressive assembly conditions, when the two capsid regions with different symmetries bind imperfectly.

assembly products depends on the mean dimer−dimer binding energy $g^{\text{bind}}_{T=4}$ and the AB/CD conformational free energy difference $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$. We classify the assembly products into three categories (with representative snapshots shown in Figure 2D): $T = 4$ capsids, $T = 3$ capsids, and *malformed* structures, which do not have icosahedral symmetry but are

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

(meta)stable on simulation time scales. Figure 2A shows the selectivity for $T = 4$ capsids (defined as ratio of $T = 4$ capsids to all closed shells) as a function of $g^{bind}_{T=4}$ and $\Delta g^{conf}_{CD\rightarrow AB}$, while Figure 2B,C shows histograms of assembly product distributions for varying $g^{bind}_{T=4}$ and $\Delta g^{conf}_{CD\rightarrow AB}$, respectively.

The model reproduces the experimental observations that a fraction of $T = 3$ capsids assemble, despite the fact that the model is geometrically optimized for the $T = 4$ icosahedral symmetry. Further, the malformed products have a broad size distribution with most shells being between ~85−140 dimers, which is consistent with size distribution measurements by CDMS.[24] Malformed structures are typically unclosed, with a mixed morphology that comprises two distinct parts with $T = 4$ and $T = 3$ morphologies, although we observe strained holey capsids at high binding energies. Examples of mixed-morphology and holey capsids are shown in Figure 2D.

The Intradimer Conformational Equilibrium Strongly Affects the Ratio of T = 4/T = 3 Capsids. As shown in Figure 2A, the model robustly assembles icosahedral capsids over a broad range of g^{bind}_0 and $\Delta g^{conf}_{CD\rightarrow AB}$, but the selectivity for $T = 4$ capsids depends strongly on the conformational free energy. For a relatively narrow range $\Delta g^{conf}_{CD\rightarrow AB}/k_BT \in [3.5,5.5]$, we observe selectivity values that are consistent with experimental observations (on the order of 5−30%). This suggests bounds for the parameter $\Delta g^{conf}_{CD\rightarrow AB}$.

Overly Strong Binding Affinities Increase the Fraction of Malformed Structures. Increasing the base binding affinity (g^{bind}_0) decreases the proportion of assembled shells that have $T = 4$ symmetry (x axis of Figure 2A). Interestingly, the yield of $T = 3$ capsids is relatively constant over this range, as observed in ref 81; the reduction in $T = 4$ morphologies occurs due to an increase in malformed structures, especially the mixed-morphology class. These structures typically have open boundaries, and under moderate assembly conditions (moderate binding affinities or subunit concentrations) they are able to "edit" or reconfigure their assembly geometry to result in icosahedral $T = 3$ or $T = 4$ capsids (e.g., see Figure 5B).

Simulated Assembly Product Size Distributions Are Consistent with Experiments if the Intradimer Conformational Equilibrium Depends on Ionic Strength. Figure 3A shows the assembly product size distribution measured in CDMS experiments at four different parameter sets, corresponding to two subunit concentrations and two ionic strengths, from ref 24. The fraction of $T = 3$ capsids dramatically increases with ionic strength, from ~5% at $I = 210$ mM (top) to ~30% at $I = 510$ mM (bottom). In contrast, increasing the subunit concentration from $C_{tot} = 10$ μM (left) to 20 μM (right) does not significantly change the $T = 4/T = 3$ ratio but does increase the prevalence of nonicosahedral structures with sizes of between 85 and 140 dimers.

We find that the simulations qualitatively reproduce both of these behaviors (Figure 3A) if the conformational free energy $\Delta g^{conf}_{CD\rightarrow AB}$ decreases ($K_{AB\text{-}CD}$ increases) with increasing ionic strength. In particular, decreasing the free energy difference from $\Delta g^{conf}_{CD\rightarrow AB} = 4.5$ (top) to $\Delta g^{conf}_{CD\rightarrow AB} = 3.5$ (bottom) changes the $T = 4/T = 3$ ratio, similar to the effect of increasing the salt concentration in experiments. Here we have also slightly increased the binding affinity, $g^{bind}_{T=4} = -5.8$ (top) and $g^{bind}_{T=4} = -6.2$ (bottom), to match the experimental observation that $g^{bind}_{T=4}$ increases with increasing salt concentration. Increasing the protein concentration (from left to right) results in more mixed-morphology products in simulations and CDMS experimental observations. These results are consistent with

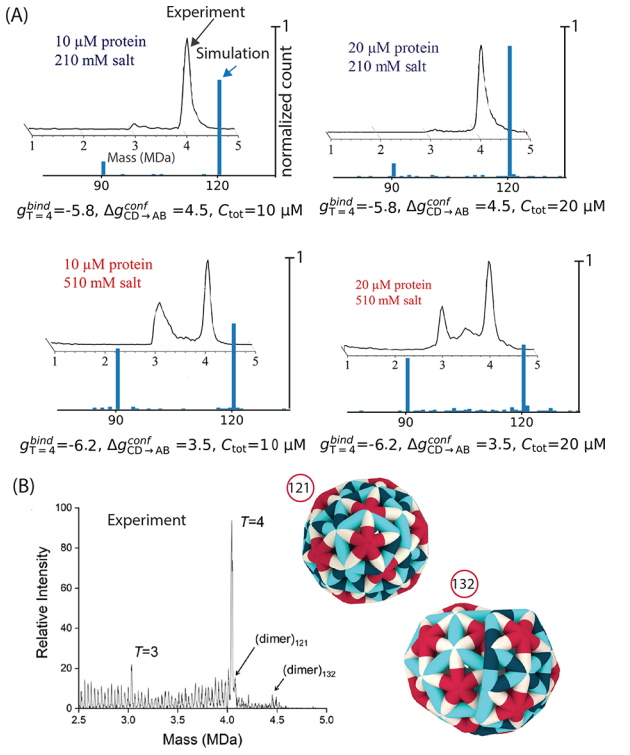


Figure 3. (A) Comparison of the simulation assembly product size distribution with CDMS experiments (after 10 min). Experimental results are shown for ionic strengths of $I = 210$ mM (top) and $I = 510$ mM (bottom) and dimer concentrations $C_{tot} = 10$ μM (left) and $C_{tot} = 20$ μM (right). The peaks near 3 and 4 MDa correspond to $T = 3$ and $T = 4$ capsids, respectively. Simulation assembly size distributions are shown for $\Delta g^{conf}_{CD\rightarrow AB} = 4.5k_BT$, $g^{bind}_{T=4} = -5.8k_BT$ (top) and $\Delta g^{conf}_{CD\rightarrow AB} = 3.5k_BT$, $g^{bind}_{T=4} = 6.2k_BT$ (bottom) and $C_{tot} = 10$ μM (left) and $C_{tot} = 20$ μM (right). Insets of CDMS spectra in the four panels are respectively from Figure 1A−D of Lutomski et al.[24] (B) High-resolution size distribution of assembly products in CDMS experiments[25] for $I = 200$ mM and $C_{tot} = 10$ μM. Snapshots show the mixed-morphology capsids with 121 and 132 dimers, which are among the most abundant nonicosahedral structures observed in both simulations and experiments. While the 121-dimer shell has 111 of its dimers consistent with $T = 4$ symmetry and a small region of 10 dimers with local $T = 3$ symmetry, the 132-dimer shell contains an approximately equal mix (~60 dimers in $T = 3$ and ~62 dimers in $T = 4$ environments).

previous experimental evidence for a relationship between dimer conformation and ionic strength.[65,79]

Figure 3B shows a high-resolution CDMS measurement of the assembly product size distribution for $I = 200$ mM and $C_{tot} = 10$ μM. The assembly product size distributions in our higher concentration simulations are qualitatively consistent with the experimental observations, and we observe prevalent nonicosahedral assembly products at consistent sizes: e.g., the 121-dimer and 132-dimer structures shown in the figure.

The Conformational Dependence and Asymmetry of HBV Interdimer Interactions Are Highly Optimized. The results discussed up to this point have focused on the relative values of dimer−dimer binding affinities estimated from buried surface area (PDBePISA) for each pair of dimer conformations (Table 1). Since we expect these to roughly correspond to the wild-type HBV capsid protein, changes in these affinities would correspond to amino acid substitutions at the dimer−dimer assembly interface or the presence of antiviral agents that bind

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

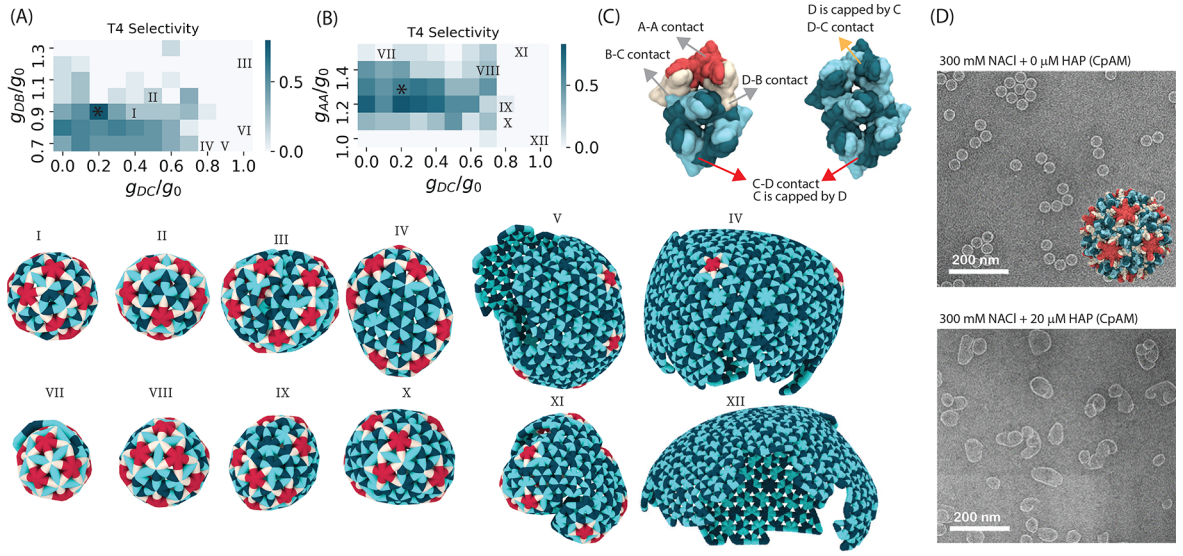


Figure 4. Conformational specificity of interdimer binding affinities is important for assembling $T = 4$ capsid morphologies. (A, B) Dependence of selectivity for $T = 4$ capsids on binding affinities between dimers with different conformations: (A) g^{bind}_{D-B} and g^{bind}_{D-C}; (B) g^{bind}_{A-A} and g^{bind}_{D-C}. Selectivity is defined as the ratio of $T = 4$ capsids to all closed shells. The asterisk symbols in (A) and (B) indicate the relative binding affinities estimated from PDBePISA, showing that these values result in $T = 4$ selectivities close to those observed in experiments. Snapshots of typical morphologies at parameter sets indicated on the plots (I–XII) are shown at the bottom. (C) A diamond from a $T = 4$ capsid (left) and a small piece of the hexameric lattice (right) that forms in flat sheets, showing different contacts between quasi-equivalent conformations. The D–C contact is required to form the hexameric sheets and aberrant structures but does not occur in $T = 3$ or $T = 4$ morphologies. (D) Transmission electron micrographs of the *in vitro* assembly of HBV dimers in the absence (top) and presence (bottom) of CpAM antiviral molecules. Images adapted from Kondylis et al.[76] with permission, copyright 2018 ACS Publications.

to the assembly interface.[11−18,82,83] Note that the dimer−dimer interactions are asymmetric—within a capsid, each dimer is "capped" from one side by another dimer, and thus, for example, the C-D contact (with binding energy g^{bind}_{C-D}) where C is capped by D is different from the D−C contact (with binding energy g^{bind}_{D-C}) where D is capped by C.

Figure 4A,B shows the assembly products that result when some of the relative binding affinities ($\overline{g}^{bind}_{A-A}$, $\overline{g}^{bind}_{D-B}$, and $\overline{g}^{bind}_{D-C}$) deviate from the PDBePISA estimate. Moreover, these results suggest that the binding affinities are highly optimized for the $T = 4$ capsid assembly. Selectivities for $T = 4$ on the order of those observed experimentally occur only for relative binding affinities that are close to those estimated from buried surface area (denoted by the asterisk symbols in Figure 4A,B).

Deviations in $\overline{g}^{bind}_{A-A}$ and $\overline{g}^{bind}_{D-B}$ have qualitatively different effects on assembly morphology than $\overline{g}^{bind}_{D-C}$. First, optimal values of $\overline{g}^{bind}_{A-A}$ and $\overline{g}^{bind}_{D-B}$ are ∼1, whereas the optimal value of $\overline{g}^{bind}_{D-C} \approx 0.2-0.3$ (i.e., the D−C contact is much weaker than the mean binding affinity). Second, increasing $\overline{g}^{bind}_{A-A}$ and $\overline{g}^{bind}_{D-B}$ above their optimal values typically results in the mixed-morphology malformed structures described above for overly strong mean binding affinities. In contrast, increasing $\overline{g}^{bind}_{D-C}$ above its optimal value, $0.3 < \overline{g}^{bind}_{D-C} \lesssim 0.7$, leads to assemblies that are closed but lack even partial icosahedral symmetry. In most cases these are asymmetric, but at some parameter sets we observe capsids with D_{5h} symmetry (e.g., image I at $\overline{g}^{bind}_{D-C} = 0.4$ in Figure 4). Further increasing $\overline{g}^{bind}_{D-C} \gtrsim 0.7$ results in large aberrant structures that contain extended hexagonal lattices (V, VI, XI and XII in Figure 4). Interestingly, these structures bear a strong resemblance to those observed in experiments on assembly of HBV dimers in the presence of core protein allosteric modulator (CpAM) molecules (Figure 4E).[76] CpAMs in general are assembly agonists, increasing dimer−dimer binding affinities; HAPs in particular favor a flattening of quasi-six-folds by binding preferentially to B−C and C−D interdimer interfaces.[14] Increasing the D−C affinity qualitatively mimics this effect.

The strong dependence of assembly morphologies on conformational specificity can be attributed to the relatively low bending modulus estimated from the atomistic simulations (and consistent with experiments), $\kappa_\phi \approx 40 k_B T$. The difference in elastic energy per subunit between $T = 3$ and $T = 4$ morphologies is $\Delta\mathcal{H}_{elastic}/n_{dimer} \approx 0.005 k_B T$, and even the large aberrant structures (V, VI, XI, and XII in Figure 4) incur only modest elastic energy costs relative to a $T = 4$ structure ($\Delta\mathcal{H}_{elastic}/n_{dimer} \approx 0.4 k_B T$). Thus, gradients in elastic energy alone are not strong enough to guide assembly trajectories toward the $T = 4$ morphology, and conformational dependence is required as well.

For the remainder of this article, all results use the default values for $\overline{g}^{bind}_{A-A}$, $\overline{g}^{bind}_{D-B}$, and $\overline{g}^{bind}_{D-C}$ given in Table 1.

Assembly Pathways. *Error Correction during Assembly and Recovery from Overgrown Structures.* Figure 5A,B and Figure S5A,B show example assembly trajectories that respectively result in $T = 3$ or $T = 4$ capsids or asymmetric assembly products. Snapshots are shown along each trajectory, along with labels indicating the numbers of dimers in long-lived intermediates. Here we have defined a long-lived intermediate as a state with a population fraction (of the ensemble of all states along dynamical trajectories) that exceeds a threshold value of 1% (Figure S5A). Qualitatively, this corresponds to the fraction of protein in detectable intermediates in SAXS.[28,84]

While the assembly products can be classified into different categories as described above, the underlying assembly pathways have a key feature in common. At multiple points along assembly trajectories, subunits can bind with a local geometry that is incompatible with the larger-scale geometry of

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

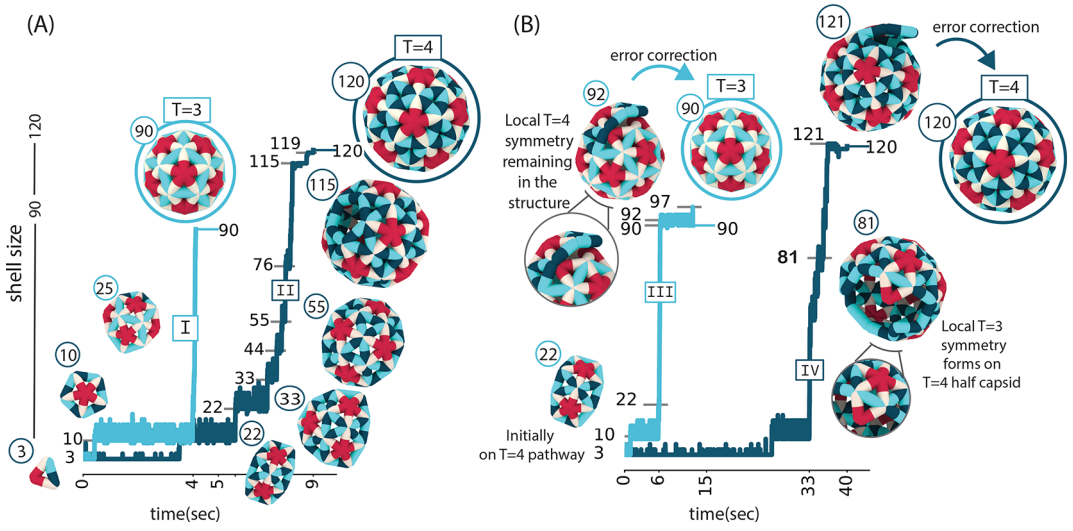


Figure 5. Assembly pathways. (A) Examples of assembly trajectories of $T = 4$ and $T = 3$ capsids, with snapshots of assembly intermediates at $g^{bind}_{T=4} = -5.8k_BT$, $\Delta g^{conf}_{CD\rightarrow AB} = 4.5k_BT$, and $C_{tot} = 20\ \mu M$. (B) Examples of assembly trajectories in which overgrown (unclosed) capsids form, followed by "error correction", or shedding of excess subunits and reconfiguration into complete $T = 4$ (blue) and $T = 3$ (cyan) capsids. Parameters for the $T = 4$ and $T = 3$ trajectories are respectively $g^{bind}_{T=4} = -5.8k_BT$, $\Delta g^{conf}_{CD\rightarrow AB} = 4.5k_BT$, $C_{tot} = 15\ \mu M$ and $g^{bind}_{T=4} = -6.0k_BT$, $\Delta g^{conf}_{CD\rightarrow AB}=3.5k_BT$, and $C_{tot} = 10\ \mu M$. Enlarged views of assembly intermediates of trajectories show the formation of locally incompatible morphologies, which are later corrected. Animations corresponding to these trajectories are provided in Movies S2−S5 in the Supporting Information.

the existing shell, resulting in regions corresponding to different morphologies (e.g., a combination of $T = 4$ and $T = 3$ symmetries). This morphology mismatch results in different behaviors depending on the extent of the mismatch and the number of subunits required to dissociate or rearrange to achieve a uniform morphology.

$T = 4$ assembly trajectories (Figure 5-II,IV) typically involve several long-lived intermediates, which have geometries such that the next dimer to associate can only form a single bond and is thus relatively unstable. In addition, many of these states have several CD/CC dimers on the boundary, which can act as a "seed" for mixed-morphology excursions in which dimers bind with local $T = 3$ symmetry. In this case, successful assembly of a $T = 4$ structure requires error correction: i.e., unbinding or reconfiguration of the dimers in local $T = 3$ arrangements to recover global $T = 4$ symmetry. Thus, trajectories may surmount multiple free energy barriers before assembly proceeds rapidly.

While both $T = 3$ and $T = 4$ assembly exhibit a long-lived 10-dimer intermediate, we do not observe larger long-lived intermediates in $T = 3$ pathways. This trend reflects the fact that the on-pathway $T = 3$ intermediates with sizes $n > 10$ have either a convex boundary or multiple AB dimers and thus avoid high barriers to additional subunit association or seeds with local $T = 4$ symmetry. This suggests that the relatively high value of g^{bind}_{A-A} plays a key role in the facile completion of $T = 3$ capsids.

Overgrown Capsids. In some cases, regions with mismatched local symmetry remain in the partial capsid structure until the final stages of the assembly, resulting in the formation of "overgrown" structures with more than 120 or 90 dimers for $T = 4$ or $T = 3$ capsids, respectively. Figure 5B shows an example of a trajectory (blue line, IV) for which a region of local $T = 3$ symmetry causes a mismatch between the orientations of boundary dimers as the capsid nears completion, resulting in overgrowth to an unclosed 121-dimer structure. However, eventually the dimers in the region with $T = 3$ symmetry disassemble due to their less favorable elastic energy and binding affinities, and then the capsid rapidly forms a complete $T = 4$ structure. We observe similar pathways involving correction of overgrown structures that lead to $T = 3$ capsids. For example, Figure 5B (cyan line, III) shows a trajectory in which an overgrown mixed-morphology structure with about 97 dimers eventually closes with $T = 3$ symmetry. This pathway starts with formation of a 22-dimer structure in $T = 4$ symmetry and continues with addition of dimers in $T = 3$ symmetry, until reaching a metastable intermediate with 90−92 dimers that retains the initial $T = 4$ region. This long-lived intermediate has ∼9 dimers on the boundary, but further growth is prevented by sterics and elastic strain due to the unfavorable curvature (see Figure S4A). Finally, the structure breaks interactions and undergoes conformational changes in the $T = 4$ region, leading to the $T = 3$ morphology. Importantly, such error correction only occurs under moderate assembly conditions: i.e., when the net driving force for assembly (determined by binding affinities and dimer concentrations) is such that assembly is nearly reversible (e.g., Figure 2B).

The trajectories and error correction described above are qualitatively consistent with the recent CDMS observations of overgrowth followed by shedding of excess subunits.[27] Moreover, slow capsid closure during the final stages of HBV assembly was observed in SAXS experiments by Chevreuil et al.[29] Figure S7B shows that $T = 4$ trajectories on average spend up to ∼20% of the assembly time in the preclosure states.

Long-Lived Off-Pathway Intermediates in Simulations and Experiments. At higher concentrations or higher binding affinities, we observe other assembly products (meaning they are (meta)stable on all simulation time scales) with sizes between 85 and 140 dimers. Despite an open boundary that allows further subunit association and dissociation, the sizes of these structures remain relatively constant, due to the mismatch in the curvature of two parts. Simulation trajectories

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

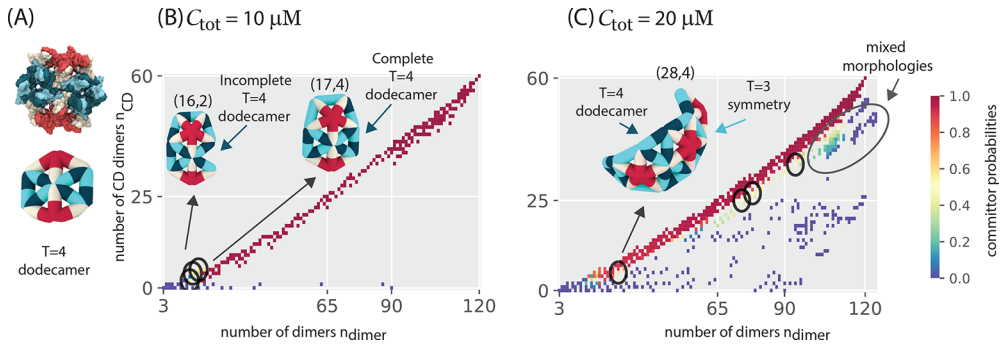


Figure 6. T = 4 pathway "hubs". (A) The structure of the "T = 4 dodecamer". (B) Committor probabilities for T = 4 capsids $q_i^{T=4}$ computed from a Markov state model (MSM) analysis, shown as a function of the number of CD dimers n_{CD} and the total number of dimers n_{dimer} for C_{tot} = 10 μM at $g_{T=4}^{bind}$ = −5.8 k_BT and $\Delta g_{CD\rightarrow AB}^{conf}$ = 4.5k_BT. The committor probability $q_i^{T=4}$ corresponds to the probability that a trajectory initiated at a given structure will form a complete T = 4 shell before disassembling. Circles indicate the T = 4 pathway "hubs" ($q_i^{T=4} \approx 0.5$), from which a trajectory is equally likely to form a T = 4 capsid or other product morphology. Other snapshots are labeled as (n_{dimer}, n_{CD}). The (17, 4) intermediate in (B) is the smallest structure that has a complete T = 4 dodecamer. (C) T = 4 committor probabilities for C_{tot} = 20 μM at parameters similar to those in (B). At C_{tot} = 10 μM (B), the smallest hub states lack a complete T = 4 dodecamer, while at the higher concentration (C) the smallest hub state has a mixed morphology that includes a T = 4 dodecamer. The region indicated by an oval at large sizes in (C) shows that at high concentrations there is a high propensity to form long-lived mixed-morphology structures with sizes between 85 and 140 dimers.

(Figure S3A,B) show that these structures form because of unsuccessful correction when a local mismatch in the geometry forms, resulting in a metastable structure that has two distinct parts with respectively T = 4 and T = 3 morphologies. Further details of this mechanism are provided in the Long-lived T = 3-T = 4 mixed morphologies section in the Supporting Information and Figures S3 and S4.

Inferring Mechanisms of Dimorphism and Path Selection. Following the method introduced in ref 85 we built Markov state models (MSMs) from the simulated assembly trajectories and then used transition path theory[86] to enable further insight into factors that control morphology selection during assembly and when pathways resulting in T = 4 or T = 3 morphologies diverge from each other. To this end, we decomposed the set of simulated capsid structures into a state space defined by two order parameters: (1) the number of dimers n_{dimer} in a partial capsid structure and (2) an order parameter that distinguishes the assembly morphology. We chose different quantities for the second order parameter depending on the morphology of interest: namely, the number of CD dimers n_{CD} or the number of CC dimers n_{CC} for pathways leading to T = 4 or T = 3. From the set of simulation trajectories at a given parameter set, we estimated the transition probability matrix $\mathbf{T}(\tau)$, which has elements T_{ij} that give the probability of a transition between a pair of states i,j in a lag time τ (see Methods).

To evaluate the accuracy of the constructed MSMs, we compare the MSM prediction for the fraction of assembled structures with sizes corresponding to T = 4 and T = 3 capsids as a function of time to the unbiased MC dynamics (Figure S1A,B). This comparison provides a stringent test of the MSM, as it requires an accurate estimate of all statistically relevant elements of the transition matrix. The results closely agree and show that T = 3 capsids form much earlier in the reaction than T = 4 capsids, in agreement with experimental observations.[81] The constructed MSM also predicts that, on longer time scales, T = 3 sized structures convert to T = 4 capsids, as seen in experiments.[24] We also validated the state decomposition by the Chapman−Kolmogorov test (see Markov State Models (MSMs) in the Supporting Information and Figure S2).

With the constructed MSMs, we use transition path theory to identify dominant pathways, key structural intermediates, and points at which pathways are committed to certain morphology products as follows.

Committor Probabilities Identify Pathway Selection Hub States. The committor probability, q_i^B, is the conditional probability that a trajectory that is in state i will visit a set of outcome states (denoted as B) before returning to the initial unassembled state[85,86] (see Markov State Models (MSMs) in the Supporting Information). Figure 6 shows the committor probabilities for the complete T = 4 capsid $q_i^{T=4}$ at two different dimer concentrations, C_{tot} = 10 μM (left) and C_{tot} = 20 μM (right). The most notable difference between the two concentrations involves the committor probabilities of the states close to T = 4 capsid (n_{dimer} > 100, n_{CD} > 25). While the majority of states end up in T = 4 capsids ($q_i^{T=4} \approx 1$) at the lower concentration, a fraction of these states have $q_i^{T=4} \approx 0$ at higher dimer concentration: i.e., they remain trapped in other morphologies for the finite simulation time.

We identify the intermediates at which assembly pathways are most likely to diverge from formation of a T = 4 capsid, resulting in an alternative assembly product, by identifying *hub states* as those for which the T = 4 committor probability $q_i^{T=4} \approx 1/2$. The five of these states with the highest forward flux (discussed next), as well as the smallest hub state, are shown with circles on the plots in Figure 6B,C. At the lower concentration, pathways diverging at the hub states typically result in T = 3 capsids, whereas at the higher concentration assembly pathways are significantly more likely to diverge from T = 4 to mixed-morphology states.

A comparison of the committor probabilities and assembly products at the different concentrations reveals that formation of a "T = 4 dodecamer" (snapshot in Figure 6A) is the key event that determines whether pathways are more likely to proceed to T = 4 or T = 3 products. The T = 4 dodecamer is the smallest relatively stable partial capsid intermediate that occurs in T = 4 capsids but not T = 3 capsids. The hub states at C_{tot} = 10 μM (Figure 6B) that lead to T = 3 capsids occur before formation of a structure with a complete T = 4 dodecamer, whereas hub states at C_{tot} = 20 μM (e.g., the 28-

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

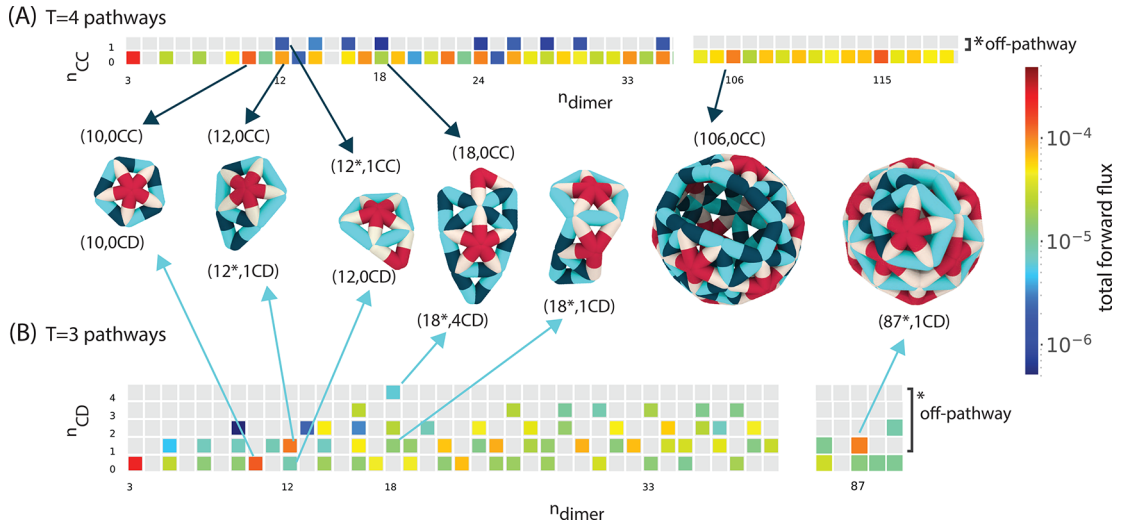


Figure 7. MSM analysis of intermediate stabilities and assembly pathway probabilities. Total forward flux to (A) $T = 4$ and (B) $T = 3$ capsids as the function of the number of dimers n_{dimer} and the number of CC (n_{CC}) or CD (n_{CD}) dimers for $T = 4$ or $T = 3$ pathways, respectively. States with CC dimers ($n_{CC} > 0$) are off-pathway for $T = 4$ assembly, and states with CD dimers ($n_{CD} > 0$) are off-pathway for $T = 3$ assembly. For the $T = 4$ pathway all the states with high forward flux are on-pathway states, while on $T = 3$ pathways many of the states with high forward flux are off-pathway, and there are off-pathway states with up to 5 CD dimers in structures with high forward flux to a $T = 3$ capsid. The 18-dimer (18*,4CD) state on the $T = 3$ pathway is actually an on-pathway $T = 4$ intermediate, but it can also generate pathways that form mixed-morphologies that undergo annealing to form $T = 3$ capsids. Simulation parameters are $g^{bind}_{T=4} = -5.8\ k_BT$, $\Delta g^{conf}_{CD\rightarrow AB} = 4.5 k_BT$, and $C = 20\ \mu M$. Examples of assembly intermediates in the initial and final stages of the assembly are shown. Middle states are not shown for clarity.

dimer structure shown in Figure 6C) occur after formation of a complete $T = 4$ dodecamer and consequently diverge to mixed-morphology malformed states. We find that, with increasing concentration, pathways typically do not commit to the $T = 4$ morphology until larger sizes and the size of the smallest hub state increases. Thus, increasing concentration results in a higher probability of forming mixed-morphology structures across a range of large sizes (85−140 dimers). The increasing proportion of long-lived intermediates in this size range with increasing concentration is consistent with CDMS measurements.[24]

The predicted significance of the $T = 4$ dodecamer to the formation of $T = 4$ capsids qualitatively agrees with AFM experiments of of HBV dimer assembly on mica plates, which frequently observed this intermediate.[26]

Forward Flux Reveals Most Probable Assembly Pathways. While the committor probabilities in Figure 6B,C identify the "hub" states for $T = 4$ assembly, the *forward flux* f^+_{ij} shows the relative probability of different reaction channels and hence the relative importance of different classes of assembly pathways. Figure 7A,B shows the forward flux toward $T = 4$ and $T = 3$ morphologies, respectively. The first order parameter is the number of capsid dimers. The second order parameter in each plot is chosen to show deviations from the target stucture pathway: i.e., the y axis is the number of CC dimers in a given state for $T = 4$ pathways (Figure 7A) and number of CD dimers for $T = 3$ pathways (Figure 7B).

States with the largest flux between the initial and final configurations identify the most probable assembly pathways for each morphology. Interestingly, these states coincide with the long-lived intermediates discussed above (Figure 5). The states with high forward flux toward $T = 4$ capsids (Figure 7A) can be understood by visualizing individual trajectories (Movie S2). Typically, growth from one of these states begins with multiple off-pathway excursions, in which dimers bind with configurations incompatible with the final morphology, followed by stalling of growth, dissociation, and a return to the long-lived intermediate state, before eventual on-pathway growth.

While the states with highest forward flux on $T = 4$ pathways are all on-pathway states (with no CC dimers), the $T = 3$ pathway exhibits several off-pathway states with high forward flux, and we do not observe any trajectories that transition from a $T = 3$ pathway to a complete $T = 4$ capsid.

CONCLUSIONS

We have investigated the assembly pathways of HBV capsid proteins using a coarse-grained model, informed by atomic-resolution data from molecular dynamics simulations and structure-based estimates of binding affinities for different protein conformations. The simulations reproduce key assembly products observed in experiments, including polymorphic assembly into $T = 3$ and $T = 4$ icosahedral structures and nonicosahedral complexes. Notably, our simulations predict the structural characteristics of the off-pathway intermediates, which could not be inferred from the CDMS experiments.[22−25] At the optimized values of the elastic moduli and binding energies, the off-pathway products have a mixed morphology comprising a combination of local $T = 4$ and $T = 3$ symmetry environments. The curvature of such structures is not geometrically compatible with self-closure, leading to shells with defects or holes, as well as overgrown structures in which the assembling structure spirals around itself until assembly is stalled due to excluded volume and elastic strain. These observations elucidate the experimental finding that metastable asymmetric intermediates can convert into icosahedral capsids over several days,[24] as well as slow capsid closure during the final stages of HBV assembly observed using SAXS experiments.[29] The simulations also predict key factors in the assembly of large aberrant structures

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

with lower curvature observed in HBV experiments with CpAM modulators and demonstrate that the combination of specific molecular-scale interactions and flexibility, in particular the relatively small bending modulus in comparison to other virus capsids, enable HBV to undergo high-fidelity assembly into the infectious capsid structure while retaining the capability to form other polymorphs.

The simulations suggest that, due to this flexibility, a coupling between the protein conformational state and its protein−protein interactions is critical for determining assembly pathways and products. These results suggest additional interpretations of existing experimental data. For example, existing models have not been able to explain the experimental observations that higher proportions of $T = 3$ capsids assemble at higher ionic strengths and in ammonium acetate buffer compared to NaCl.[29,84]

The observation from our simulations that $\Delta g^{conf}_{CD\rightarrow AB}$ has a strong effect on morphology selectivity suggests that the ionic strength shifts the conformational equilibrium toward CD dimer conformations. That is, increasing the salt concentration decreases $\Delta g^{conf}_{CD\rightarrow AB}$ and thereby increases the driving force for 5-fold dimer coordination during the assembly process, thus favoring pathways that lead to the $T = 3$ morphology. These results are also consistent with recent suggestions that coupling between conformational interconversion and interaction strengths provides an important means of regulatory control over the timing and robustness of assembly.[80,87−89] Moreover, Biela et al.[90] recently demonstrated that assembly of larger capsids from MS2 capsids ($T = 4$ and D_5) could be achieved by engineering insertions into the capsid protein that likely shift the capsid protein conformational equilibrium (analogous to changing our parameter $\Delta g^{conf}_{CD\rightarrow AB}$).

The need for conformational specificity when accounting for HBV material properties sheds light on previous results for a model of icosahedral capsid assembly with only one subunit species, which found that much higher values of the bending modulus, corresponding to much smaller values of the Föppl von Kármán number number (FvK < 0.25, compared to FvK ≈ 500 for HBV), were required to observe assembly into $T = 4$ structures.[51,91] Here we find that $T = 4$ capsids assemble with high yields at FvK ≈ 500, provided that the dimer subunits can adopt quasi-equivalent conformations with different binding affinities.

Trajectory analysis shows that, while the products of HBV assembly can be qualitatively classified into $T = 4$ capsids, $T = 3$ capsids, and long-lived malformed structures, the assembly pathways all have key features in common. We identified hub states or intermediates at which assembly pathways frequently diverge from $T = 4$ pathways, as well as structural features of such hub states that determine the classes of products they will form. A key feature is whether a hub state intermediate contains a dodecamer of dimers in the quasi-6-fold arrangement found in $T = 4$ capsids (the $T = 4$ dodecamer in Figure 6A).

Trajectories that form such an intermediate will go on to form either $T = 4$ capsids or malformed structures, whereas trajectories that diverge from the $T = 4$ pathway before forming a complete $T = 4$ dodecamer form either $T = 3$ capsids or malformed structures. This prediction is consistent with recent AFM observations of HBV capsid protein assembly.[26] Knowledge of hub states could suggest previously unexplored antiviral drug targets.

Outlook. The simulation results and buried surface area estimates suggest that, although the different quasi-equivalent conformations of HBV capsid proteins have high structural similarity, their differences play a key role in guiding assembly pathways. This conclusion is further supported by the fact that the antiviral assembly effectors (CpAMs) bind preferentially to interfaces between certain conformational pairs (e.g., within the quasi-6-fold interfaces). It would thus be of great interest to extend recent atomic-resolution simulations of CpAM-HBV capsid protein interactions[92,93] to investigate the molecular-scale factors that couple the protein conformational state to its interactions. This information could facilitate designing more effective assembly effectors.

The importance of conformational heterogeneity in the HBV assembly highlights the inherent tradeoff between minimizing the complexity of a self-assembly reaction and maximizing the selectivity for target structures. Recent experiments and simulations of synthetic subunits designed to assemble into capsids and tubules demonstrated that multiple species with species-specific subunit−subunit interactions can significantly increase specificity for a target geometry, by blocking assembly pathways that would lead to other geometries with similar thermodynamic stabilities.[94−96] However, the additional information content associated with encoding for multiple species or conformations incurs extra costs, such as material and design costs in the synthetic realm, or additional selective pressure on protein sequences in natural systems. Understanding how this tradeoff has shaped evolution in other natural systems would provide important information for developing treatments for pathogenic diseases that work by redirecting assembly pathways and could guide a more efficient design of synthetic self-assembly systems.

METHODS

Coarse-Grained (CG) Model. In our CG model, an assemblage is represented as a triangulated elastic sheet. Each edge in the elastic network corresponds to a dimer, the basic assembly unit for HBV.[65,66] It is straightforward to generalize the model to other common basic assembly subunits, such as monomers, trimers, and pentamers/hexamers, by changing the graph structure of the elastic sheet (e.g., refs 52, 67, and 69). Each HBV dimer is modeled as an edge with two asymmetric half-edges. Half-edges have orientations (e.g., CD conformation edges comprise CD and DC half-edges in the model), and the edge−edge bindings are type- and order-dependent (e.g., $g^{bind}_{C-D} \neq g^{bind}_{D-C}$). This is an important feature for modeling viruses that assemble from protein dimers (e.g., HBV, MS2, BMV) because the dimer−dimer interfaces in these viral capsids are asymmetric—one dimer is "capped" at one end by the other dimer. The model HBV $T = 4$ capsid has 120 edges corresponding to the 120 dimers in a complete $T = 4$ HBV capsid.

The energy for a model capsid configuration is given by

$$\mathcal{H}_{capsid} = \mathcal{H}_{elastic} + \mathcal{H}_{steric} + \mathcal{H}_{bind} + \mathcal{H}_{conf} \quad (2)$$

The elastic energy accounts for the harmonic potentials for edge length fluctuations l, dihedral angles ϕ between the planes of each pair of adjacent triangles, and binding angles θ between each pair of edges meeting at a vertex, with moduli κ_l, κ_θ, and κ_ϕ respectively:

$$\mathcal{H}_{elastic} = \mathcal{H}_{stretch} + \mathcal{H}_{bend} + \mathcal{H}_{angle} \quad (3)$$

$$\mathcal{H}_{stretch} = \frac{\kappa_l}{2} \sum_{i}^{\text{all edges}} (l_i - l^0_{c(i)})^2 \quad (4)$$

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

$$\mathcal{H}_{\text{bend}} = \kappa_\phi \sum_{i}^{\text{bound edges}} 1 - \cos[\phi_i - \phi^0_{c(i)}] \tag{5}$$

$$\mathcal{H}_{\text{angle}} = \kappa_\theta \sum_{\langle ij \rangle}^{\text{bound edge pairs}} 1 - \cos[\theta_{ij} - \theta^0_{c(i),c(j)}] \tag{6}$$

with $c(i)$ being the dimer conformation for edge index i and $c(i)$,$c'(j)$ the conformations of two bound edges, with i,j the edge indices.

The term $\mathcal{H}_{\text{steric}}$ represents the excluded volume interaction of dimers and is represented by hard-sphere excluders with positions relative to the edge axis that are based on the dimer structure (see Figure S4). Excluders move as a rigid body with the edge position and orientation. Configurations in which excluders overlap have infinite energy and thus are forbidden.

The binding energy $\mathcal{H}_{\text{bind}}$ is the sum of binding energies between pairs of bound edges (i,j), with a binding affinity that depends on the edge conformations, $g_{c(i),c'(j)}$. The half-edge data structure used in our implementation has an orientation and thus efficiently represents the conformation- and order-dependent binding energies in HBV capsids.

The term $\mathcal{H}_{\text{conf}}$ represents the intradimer conformational free energy landscape: i.e., the relative equilibrium populations of different conformational states for the capsid protein dimers (see Overview of Model and Parameters). Since CD and CC monomers are set as the zero free energy reference conformational states, the total conformational energy of an assembly is given by the number of AB dimers n_{AB} in the structure, $\mathcal{H}_{\text{conf}} = n_{\text{AB}}\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$. Although there is evidence that the conformations of HBV dimers in solution differ from those in the capsid,[80,87−89,97] NMR and crystallography experiments show that the free dimer has the same fundamental structure as in the capsid.[80,98] Therefore, while it is straightforward to allow for additional dimer conformations, we have not done so here to minimize the number of parameters.

Monte Carlo (MC) Simulations. The initial configuration for each simulation is three edges (dimers) bound in a triangular geometry, with the initial conformation of each triangle being chosen randomly. The MC algorithm includes 7 moves, which include association or dissociation of either single dimers or pairs of dimers, binding or unbinding of the dimers in the shell, conformational switches of individual dimers, and thermal relaxation of the shell by vertex displacement moves (Figure S7).

Estimating CG Model Parameters from an AA Simulation of a Complete HBV Capsid. We estimate the parameters for $\mathcal{H}_{\text{elastic}}$, κ_l, κ_θ, κ_ϕ, $\{l^0_c\}$, $\{\theta^0_{c,c'}\}$, and $\{\phi^0_c\}$, by comparison against AA molecular dynamics simulations of a complete $T = 4$ HBV capsid.[4] We coarse-grain the data from the AA simulations in two steps. First, we select the C-α atom of residue 132 of each capsid protein monomer. The rationale for using residue 132 as an anchor residue is 2-fold. One, it provides a convenient point of reference due to its position near opposite ends of the dimers principle axis. Two, it is a well-studied residue; amino acid substitutions at this site have shown measurable effects on capsid assembly and capsid stability.[99] These 240 points are clustered based on proximity using the scikit DBSCAN algorithm,[5] which identifies the 5-fold and quasi-6-fold axes. The center of mass of each cluster is then assigned to a vertex in the CG capsid. Calculating the set of CG edge lengths, dihedral angles, and binding angles for each residue based on 50,000 conformers from the 1 μs AA simulation provides the data for estimating the equilibrium distribution of these parameters in the CG model (see Figure 1C and Movie S1).

We perform MC simulations on complete $T = 4$ capsids to estimate the corresponding distribution of edge lengths and angles in the CG model. In particular, we use the complete $T = 4$ capsid mapped from a frame in the AA simulation as the initial condition and only perform vertex displacement moves to relax the structure. In each simulation, we initially perform 50000 MC sweeps to equilibrate the system and then an additional 50000 sweeps to estimate the edge length and angle distributions. We optimize parameter values by minimizing the Kullback−Leibler divergence of the CG and AA distributions for each quantity, using the Tree-structured Parzen Estimator (TPE) algorithm in the Hyperopt package.[100]

As shown in Figure S8,the distribution of edge lengths is bimodal, and thus the edge-length distribution can only be fit by a CG model with at least two edge types (corresponding to two different conformations, AB and CD). Thus, in addition to the elastic moduli, we fit equilibrium edge lengths for both conformations l^0_{AB}, l^0_{CD}, dihedral angles ϕ^0_{AB}, ϕ^0_{CD}, and binding angles $\{\theta^0_{c,c'}\}$ with $(c,c') \in \{(\text{BA,AB}), (\text{AB,CD}), (\text{CD,BA}), (\text{DC,DC})\}$.

In addition to the binding angle conformation pairs listed above for $T = 4$ and $T = 3$ capsids, structures of HBV dimers in drug-mediated assembled structures (and similarly hexagonal sheets of HBV dimers) include a binding angle pair CD−CD, for which we set $\theta^0_{\text{CD,CD}} = \pi/3$ since the sheet is composed of equilateral triangles.

Estimating Binding Energy Parameter Values. Atomic-resolution structures of different HBV capsids and assemblies[16,70] show that there are significant structural differences between the binding interfaces for different dimer conformations. Therefore, in our model the dimer−dimer binding affinities depend on the dimer conformations. Since the dimers are not head−tail symmetric, the binding affinities also depend on the relative orientations of the monomers within each dimer. Thus, there are 16 possible binding arrangements for two dimer types AB and CD. However, only 7 of these binding arrangements can be found in available structures of HBV capsids: (BA−AB, AB−CD, and CD−BA in both $T = 4$ and $T = 3$ capsids, DC−DC in $T = 4$ capsids, AB−CC and CC−BA in $T = 3$ capsids, and CD(CC)−CD(CC) in drug-mediated assembled structures (and similarly hexagonal sheets of HBV dimers).[16,70,77,78]

To estimate the relative difference in binding affinity between different dimer−dimer conformation pairs, we use the buried surface area computed from atomic-resolution structures using PDBePISA, a tool for examining macromolecular interfaces.[101,102]Table 1 shows relative binding energies calculated from the two different $T = 4$ HBV capsid crystal structures 2G33[16] and 6UI7,[70] where we have set the reference binding energy parameter $g^{\text{bind}}_0 = \bar{g}^{\text{bind}}_{\text{C−D}}$. In $T = 3$ capsids, AB−CC and CC−BA have buried surface areas similar to those of AB−CD and CD−BA in $T = 4$ capsids[70] (Table S2), and thus we consider them to be the same in the model. The CD−CD interaction is not present in $T = 4$ or $T = 3$ capsids, but calculation of the CD−CD binding affinity for CD conformations in the $T = 4$ structure[16] gives $\bar{g}^{\text{bind}}_{\text{D−C}} \approx 0.3$. We have optimized this parameter in Figure 2A,B and set $\bar{g}^{\text{bind}}_{\text{D−C}} = 0.2$. Similarly, estimates for all other binding affinities result in small values of $(0.01-0.1)g^{\text{bind}}_0$; thus, to limit the number of model parameters we set all other interactions to $0.05g^{\text{bind}}_0$.

Mapping the Free Model Parameters to Experimental Values. The structure and dynamics-based procedures described thus far to set model parameters leave two free parameters, g^{bind}_0, which controls the mean interdimer binding affinity, and $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$, the intradimer conformational interconversion free energy. These parameters together set the mean free energy of a dimer in a capsid ground-state configuration (meaning subunits are at their equilibrium position so that $\mathcal{H}_{\text{elastic}} = 0$), according to

$$K_{\text{bind}} = \nu_{\text{b}} C_0 e^{-\beta g^{\text{bind}}_{T=4}} \tag{7}$$

with

$$g^{\text{bind}}_{T=4} = (\mathcal{H}^{T=4}_{\text{bind}} + \mathcal{H}^{T=4}_{\text{conf}})/240 \tag{8}$$

where $\mathcal{H}^{T=4}_{\text{bind}} = 60g^{\text{bind}}_{\text{A−A}} + 60g^{\text{bind}}_{\text{B−C}} + 60g^{\text{bind}}_{\text{D−B}} + 60g^{\text{bind}}_{\text{C-D}}$ and $\mathcal{H}^{T=4}_{\text{conf}} = 60\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$, ν_{b} is the binding volume parameter (see the Supporting Information), and $c_0 = 1$ M is the standard state volume. $C_0 = 1$ M is also used for calculating the chemical potential $\mu = \ln(C_{\text{tot}}/C_0)$.

The mean dimer−dimer binding free energy corresponding to K_{bind} has been estimated from experimental measurement of the equilibrium capsid assembly yields as a function of total subunit concentration under different experimental conditions (e.g., ref 65). However, the assembly equilibrium depends on both g^{bind}_0 and $\Delta g^{\text{conf}}_{\text{CD}\to\text{AB}}$, and it is likely that both of these parameters depend on solution conditions.[65,79,80,87,103] Thus, it is not straightforward to

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859

estimate both of these parameters from the experimental data. Instead, we consider g_0^{bind} and $\Delta g_{CD\rightarrow AB}^{conf}$ but note that their ranges and relative values are constrained by experimental data according to eq 7.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsnano.2c02119.

Additional figures and tables as described in the text and additional details of the computation model and Markov State models (PDF)

Animation, optimizing CG model parameters against AA simulations of an HBV capsid (MP4)

Animation, assembly trajectory of a $T = 4$ capsid (MP4)

Animation, assembly trajectory of a $T = 3$ capsid (MP4)

Animation, assembly trajectory of a $T = 4$ capsid from an overgrown structure (MP4)

Animation, assembly trajectory of a $T = 3$ capsid from an overgrown structure (MP4)

Animation, assembly trajectory of a long-lived overgrown structure with 121 dimers (MP4)

AUTHOR INFORMATION

Corresponding Author

Michael F. Hagan − *Martin A. Fisher School of Physics, Brandeis University, Waltham, Massachusetts 02453, United States;* orcid.org/0000-0002-9211-2434; Email: hagan@brandeis.edu

Authors

Farzaneh Mohajerani − *Martin A. Fisher School of Physics, Brandeis University, Waltham, Massachusetts 02453, United States;* orcid.org/0000-0002-9340-3305

Botond Tyukodi − *Martin A. Fisher School of Physics, Brandeis University, Waltham, Massachusetts 02453, United States; Department of Physics, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania;* orcid.org/0000-0001-6868-1234

Christopher J. Schlicksup − *Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, Indiana 47405, United States*

Jodi A. Hadden-Perilla − *Department of Chemistry & Biochemistry, University of Delaware, Newark, Delaware 19716, United States;* orcid.org/0000-0003-4685-8291

Adam Zlotnick − *Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, Indiana 47405, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsnano.2c02119

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Institute Of General Medical Sciences under Award Numbers R01GM108021 (F.M., B.T, M.F.H.) and P20-GM-104316 (J.A.H.-P), the National Institute Of Allergies and Infectious Diseases, R01 AI118933 (C.J.S., A.Z.), the Brandeis Center for Bioinspired Soft Materials, an NSF MRSEC, DMR-2011846 (F.M., B.T., M.F.H.), and the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project no. PN-III-P1-1.1-PD-2019-0236, within PNCDI III (B.T.). Computational resources were provided by NSF XSEDE computing resources (XStream, Bridges, and Comet) and the Brandeis HPCC, which is partially supported by DMR-2011846. Animations of assembly trajectories in the Supporting Information were provided using Ovito[104] and VMD.[105]

REFERENCES

(1) Johnson, J. E.; Speir, J. A. Quasi-equivalent Viruses: A Paradigm For Protein Assemblies. *J. Mol. Biol.* **1997**, *269*, 665−675.

(2) Caspar, D. L. D.; Klug, A. Physical Principles in Construction of Regular Viruses. *Cold Spring Harbor Symp. Quant. Biol.* **1962**, *27*, 1−24.

(3) Natarajan, P.; Lander, G. C.; Shepherd, C. M.; Reddy, V. S.; Brooks, C. L.; Johnson, J. E. Exploring Icosahedral Virus Structures With Viper. *Nat. Rev. Microbiol.* **2005**, *3*, 809−817.

(4) Hadden, J. A.; Perilla, J. R.; Schlicksup, C. J.; Venkatakrishnan, B.; Zlotnick, A.; Schulten, K. All-atom Molecular Dynamics of The HBV Capsid Reveals Insights Into Biological Function and Cryo-em Resolution Limits. *Elife* **2018**, *7*, e32478.

(5) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in python. *Journal of Machine Learning Research* **2011**, *12*, 2825−2830.

(6) Viswanathan, U.; Mani, N.; Hu, Z.; Ban, H.; Du, Y.; Hu, J.; Chang, J.; Guo, J.-T. Targeting the Multifunctional HBV Core Protein As A Potential Cure for Chronic Hepatitis B. *Antiviral Res.* **2020**, *182*, 104917.

(7) Who Global Hepatitis Report. Accessed January 05, 2022 [Online], 2017.

(8) Ning, X.; Nguyen, D.; Mentzer, L.; Adams, C.; Lee, H.; Ashley, R.; Hafenstein, S.; Hu, J. Secretion of Genome-free Hepatitis B Virus−single Strand Blocking Model for Virion Morphogenesis of Para-retrovirus. *PLoS pathogens* **2011**, *7*, e1002255.

(9) Crowther, R.; Kiselev, N.; Böttcher, B.; Berriman, J.; Borisova, G.; Ose, V.; Pumpens, P. Three-dimensional Structure of Hepatitis B Virus Core Particles Determined by Electron Cryomicroscopy. *Cell* **1994**, *77*, 943−950.

(10) Dryden, K. A.; Wieland, S. F.; Whitten-Bauer, C.; Gerin, J. L.; Chisari, F. V.; Yeager, M. Native Hepatitis B Virions and Capsids Visualized by Electron Cryomicroscopy. *Molecular cell* **2006**, *22*, 843−850.

(11) Kim, C.; Schlicksup, C.; Barnes, L.; Jarrold, M.; Patterson, A.; Bothner, B.; Zlotnick, A. HBV Core-directed Antivirals and Importin β Can Synergistically Disrupt Capsids. *Microscopy and Microanalysis* **2021**, *27*, 1130−1131.

(12) Kondylis, P.; Schlicksup, C. J.; Katen, S. P.; Lee, L. S.; Zlotnick, A.; Jacobson, S. C. Evolution of Intermediates During Capsid Assembly of Hepatitis B Virus with Phenylpropenamide-based Antivirals. *ACS Infect. Dis.* **2019**, *5*, 769−777.

(13) Ruan, L.; Zlotnick, A.; Hadden, J. Assembly Properties of Hepatitis B Virus Core Protein Mutants Correlate with Their Resistance to Assembly-directed Antivirals. *Journal of Virology* **2018**, *92*, 1.

(14) Schlicksup, C. J.; Wang, J. C.-Y.; Francis, S.; Venkatakrishnan, B.; Turner, W. W.; VanNieuwenhze, M.; Zlotnick, A. Hepatitis B Virus Core Protein Allosteric Modulators Can Distort and Disrupt Intact Capsids. *Elife* **2018**, *7*, e31473.

(15) Katen, S. P.; Tan, Z.; Chirapu, S. R.; Finn, M. G.; Zlotnick, A. Assembly-directed Antivirals Differentially Bind Quasiequivalent Pockets to Modify Hepatitis B Virus Capsid Tertiary and Quaternary Structure. *Structure* **2013**, *21*, 1406−16.

(16) Bourne, C. R.; Finn, M. G.; Zlotnick, A. Global Structural Changes in Hepatitis b Virus Capsids Induced by the Assembly Effector Hap1. *J. Virol.* **2006**, *80*, 11055−11061.

(17) Stray, S. J.; Johnson, J. M.; Kopek, B. G.; Zlotnick, A. An in Vitro Fluorescence Screen to Identify Antivirals That Disrupt

Hepatitis b Virus Capsid Assembly. *Nat. Biotechnol.* **2006**, *24*, 358−362.
(18) Stray, S. J.; Zlotnick, A. Bay 41−4109 Has Multiple Effects on Hepatitis b Virus Capsid Assembly. *J. Mol. Recognit.* **2006**, *19*, 542−548.
(19) Zhou, K.; Li, L.; Tan, Z.; Zlotnick, A.; Jacobson, S. C. Characterization of Hepatitis b Virus Capsids by Resistive-pulse Sensing. *J. Am. Chem. Soc.* **2011**, *133*, 1618−1621.
(20) Zhou, J.; Kondylis, P.; Haywood, D.; Harms, Z.; Jacobson, S.; Lee, L.; Zlotnick, A. Characterization of Virus Capsids and Their Assembly Intermediates by Multicycle Resistive-pulse Sensing with Four Pores in Series. *Anal. Chem.* **2018**, *90*, 7267.
(21) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; Van Duijn, E.; Heck, A. J. Interrogating Viral Capsid Assembly with Ion Mobility-mass Spectrometry. *Nat. Chem.* **2011**, *3*, 126−132.
(22) Pierson, E. E.; Keifer, D. Z.; Selzer, L.; Lee, L. S.; Contino, N. C.; Wang, J. C. Y.; Zlotnick, A.; Jarrold, M. F. Detection of Late Intermediates in Virus Capsid Assembly by Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136*, 3536−3541.
(23) Pierson, E. E.; Keifer, D. Z.; Kukreja, A. A.; Wang, J. C.-Y.; Zlotnick, A.; Jarrold, M. F. Charge Detection Mass Spectrometry Identifies Preferred Non-icosahedral Polymorphs in the Self-assembly of Woodchuck Hepatitis Virus Capsids. *Journal of molecular biology* **2016**, *428*, 292−300.
(24) Lutomski, C. A.; Lyktey, N. A.; Pierson, E. E.; Zhao, Z.; Zlotnick, A.; Jarrold, M. F. Multiple Pathways in Capsid Assembly. *J. Am. Chem. Soc.* **2018**, *140*, 5784−5790.
(25) Todd, A. R.; Barnes, L. F.; Young, K.; Zlotnick, A.; Jarrold, M. F. Higher Resolution Charge Detection Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 11357−11364.
(26) Buzon, P.; Maity, S.; Christodoulis, P.; Wiertsema, M. J.; Dunkelbarger, S.; Kim, C.; Wuite, G. J.; Zlotnick, A.; Roos, W. H. Virus Self-assembly Proceeds through Contact-rich Energy Minima. *Science advances* **2021**, *7*, eabg0811.
(27) Lutomski, C. A.; Lyktey, N. A.; Zhao, Z.; Pierson, E. E.; Zlotnick, A.; Jarrold, M. F. Hepatitis B Virus Capsid Completion Occurs through Error Correction. *J. Am. Chem. Soc.* **2017**, *139*, 16932−16938.
(28) Asor, R.; Schlicksup, C. J.; Zhao, Z.; Zlotnick, A.; Raviv, U. Rapidly Forming Early Intermediate Structures Dictate the Pathway of Capsid Assembly. *J. Am. Chem. Soc.* **2020**, *142*, 7868−7882.
(29) Chevreuil, M.; Lecoq, L.; Wang, S.; Gargowitsch, L.; Nhiri, N.; Jacquet, E.; Zinn, T.; Fieulaine, S.; Bressanelli, S.; Tresset, G. Nonsymmetrical Dynamics of the HBV Capsid Assembly and Disassembly Evidenced by Their Transient Species. *J. Phys. Chem. B* **2020**, *124*, 9987−9995.
(30) Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Ieong, P. U.; Dommer, A. C.; Amaro, R. E. Mesoscale All-atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent Sci.* **2020**, *6*, 189−196.
(31) Ghaemi, Z.; Gruebele, M.; Tajkhorshid, E. Molecular Mechanism of Capsid Disassembly in Hepatitis B Virus. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2102530118.
(32) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-electron Microscopy and All-atom Molecular Dynamics. *Nature* **2013**, *497*, 643−646.
(33) Jana, A. K.; May, E. R. Structural and Dynamic Asymmetry in Icosahedrally Symmetric Virus Capsids. *Current Opinion in Virology* **2020**, *45*, 8−16.
(34) Salas, G. G. S.; Hernandez, A. E. L.; He, J.; Karki, C.; Xie, Y.; Sun, S.; Xian, Y.; Li, L. Using Computational Approaches to Study Dengue Virus Capsid Assembly. *Computational and Mathematical Biophysics* **2019**, *7*, 64−72.
(35) Pavlova, A.; Bassit, L.; Cox, B. D.; Korablyov, M.; Chipot, C.; Verma, K.; Russell, O. O.; Schinazi, R. F.; Gumbart, J. C.Mechanism of Action of HBV Capsid Assembly Modulators Predicted from Binding to Early Assembly Intermediates*bioRxiv*2020, DOI: 10.1101/2020.03.23.002527, accessed July 30, 2022.
(36) Hagan, M. F. Modeling Viral Capsid Assembly. *Adv. Chem. Phys.* **2014**, *155*, 1−68.
(37) Hagan, M. F.; Zandi, R. Recent Advances in Coarse-grained Modeling of Virus Assembly. *Current Opinion in Virology* **2016**, *18*, 36−43.
(38) Nguyen, H. D.; Reddy, V. S.; Brooks, C. L. Invariant Polymorphism in Virus Capsid Assembly. *J. Am. Chem. Soc.* **2009**, *131*, 2606−14.
(39) Rapaport, D. C. Molecular Dynamics Simulation of Reversibly Self-assembling Shells in Solution Using Trapezoidal Particles. *Phys. Rev. E* **2012**, *86*, 051917.
(40) Mendoza, C. I.; Reguera, D. Shape Selection and Mis-assembly in Viral Capsid Formation by Elastic Frustration. *eLife* **2020**, *9*, e52525.
(41) Wagner, J.; Zandi, R. The Robust Assembly of Small Symmetric Nanoshells. *Biophys. J.* **2015**, *109*, 956.
(42) Elrad, O. M.; Hagan, M. F. Mechanisms of Size Control and Polymorphism in Viral Capsid Assembly. *Nano Lett.* **2008**, *8*, 3850−3857.
(43) Fejer, S. N.; Chakrabarti, D.; Wales, D. Emergent Complexity from Simple Anisotropic Building Blocks: Shells, Tubes, and Spirals. *ACS Nano* **2010**, *4*, 219−228.
(44) Perlmutter, J. D.; Mohajerani, F.; Hagan, M. F. Many-molecule Encapsulation by An Icosahedral Shell. *eLife* **2016**, *5*, e14078.
(45) Mohajerani, F.; Hagan, M. F. The Role of the Encapsulated Cargo in Microcompartment Assembly. *PLoS Computational Biology* **2018**, *14*, e1006351.
(46) Mohajerani, F.; Sayer, E.; Neil, C.; Inlow, K.; Hagan, M. F. Mechanisms of Scaffold-mediated Microcompartment Assembly and Size Control. *ACS Nano* **2021**, *15*, 4197−4212.
(47) Twarock, R.; Bingham, R. J.; Dykeman, E. C.; Stockley, P. G. A Modelling Paradigm for Rna Virus Assembly. *Current opinion in virology* **2018**, *31*, 74−81.
(48) Zandi, R.; Dragnea, B.; Travesset, A.; Podgornik, R. On virus Growth and Form. *Phys. Rep.* **2020**, *847*, 1−102.
(49) Li, S.; Roy, P.; Travesset, A.; Zandi, R. Why Large Icosahedral Viruses Need Scaffolding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10971−10976.
(50) Li, S.; Zandi, R.; Travesset, A.; Grason, G. M. Ground States of Crystalline Caps: Generalized Jellium on Curved Space. *Physical review letters* **2019**, *123*, 145501.
(51) Panahandeh, S.; Li, S.; Zandi, R. The Equilibrium Structure of Self-assembled Protein Nano-cages. *Nanoscale* **2018**, *10*, 22802−22809.
(52) Panahandeh, S.; Li, S.; Marichal, L.; Leite Rubim, R.; Tresset, G.; Zandi, R. How A Virus Circumvents Energy Barriers to Form Symmetric Shells. *ACS Nano* **2020**, *14*, 3170−3180.
(53) Arkhipov, A.; Freddolino, P. L.; Schulten, K. Stability and Dynamics of Virus Capsids Described by Coarse-grained Modeling. *Structure* **2006**, *14*, 1767−1777.
(54) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-grained Model of the Sars-cov-2 Virion. *Biophysical journal* **2021**, *120*, 1097−1104.
(55) Parton, D. L.; Tek, A.; Baaden, M.; Sansom, M. S. P. Formation of Raft-like Assemblies Within Clusters of Influenza Hemagglutinin Observed by Md Simulations. *PLoS Computational Biology* **2013**, *9*, e1003034−e1003034.
(56) Ayton, G. S.; Noid, W. G.; Voth, G. A. S Ystematic Coarse Graining of Biomolecular and Soft-matter Systems. *MRS Bull.* **2007**, *32*, 929−934.
(57) Durumeric, A. E. P.; Voth, G. A. Adversarial-residual-coarse-graining: Applying Machine Learning Theory to Systematic Molecular Coarse-graining. *J. Chem. Phys.* **2019**, *151*, 124110.
(58) Sidky, H.; Chen, W.; Ferguson, A. L. Molecular Latent Space Simulators. *Chem. Sci.* **2020**, *11*, 9459−9467.
(59) Wang, J.; Olsson, S.; Wehmeyer, C.; Perez, A.; Charron, N. E.; de Fabritiis, G.; Noe, F.; Clementi, C. Machine Learning of Coarse-

grained Molecular Dynamics Force Fields. *ACS Cent Sci.* **2019**, *5*, 755−767.
(60) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deepcg: Constructing Coarse-grained Models Via Deep Neural Networks. *J. Chem. Phys.* **2018**, *149*, 034101.
(61) Yu, A.; Skorupka, K. A.; Pak, A. J.; Ganser-Pornillos, B. K.; Pornillos, O.; Voth, G. A. Trim5α Self-assembly and Compartmentalization of the HIV-1 Viral Capsid. *Nat. Commun.* **2020**, *11*, 1307.
(62) Tan, A.; Pak, A. J.; Morado, D. R.; Voth, G. A.; Briggs, J. A. G. Immature HIV-1 Assembles from Gag Dimers Leaving Partial Hexamers at Lattice Edges As Potential Substrates for Proteolytic Maturation. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2020054118.
(63) Zhu, F.; Chen, B. Monte Carlo Simulations of HIV Capsid Protein Homodimer. *J. Chem. Inf. Model.* **2015**, *55*, 1361−1368.
(64) Oliver, R. C.; Potrzebowski, W.; Najibi, S. M.; Pedersen, M. N.; Arleth, L.; Mahmoudi, N.; André, I. Assembly of Capsids from Hepatitis B Virus Core Protein Progresses through Highly Populated Intermediates in the Presence and Absence of Rna. *ACS Nano* **2020**, *14*, 10226−10238.
(65) Ceres, P.; Zlotnick, A. Weak Protein-protein Interactions Are Sufficient to Drive Assembly of Hepatitis B Virus Capsids. *Biochemistry* **2002**, *41*, 11525−11531.
(66) Wingfield, P. T.; Stahl, S. J.; Williams, R. W.; Steven, A. C. Hepatitis Core Antigen Produced in Escherichia Coli-subunit Composition, Conformational Analysis, and In Vitro Capsid Assembly. *Biochemistry* **1995**, *34*, 4919−4932.
(67) Rotskoff, G. M.; Geissler, P. L. Robust Nonequilibrium Pathways to Microcompartment Assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 6341−6346.
(68) Li, S.; Orland, H.; Zandi, R. Self Consistent Field Theory of Virus Assembly. *J. Phys.: Condens. Matter* **2018**, *30*, 144002.
(69) Tyukodi, B.; Mohajerani, F.; Hall, D. M.; Grason, G. M.; Hagan, M. F. Thermodynamic size control in curvature-frustrated tubules: Self-limitation with open boundaries. *ACS Nano* **2022**, *16*, 9077−9085.
(70) Wu, W.; Watts, N. R.; Cheng, N.; Huang, R.; Steven, A. C.; Wingfield, P. T. Expression of Quasi-equivalence and Capsid Dimorphism in the Hepadnaviridae. *PLoS Computational Biology* **2020**, *16*, e1007782.
(71) Perlmutter, J. D.; Hagan, M. F. Mechanisms of Virus Assembly. *Annu. Rev. Phys. Chem.* **2015**, *66*, 217−239.
(72) Zlotnick, A.; Johnson, J. M.; Wingfield, P. W.; Stahl, S. J.; Endres, D. A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly. *Biochemistry* **1999**, *38*, 14644−14652.
(73) Helfrich, W. Elastic Properties of Lipid Bilayers: Theory and Possible Experiments. *Zeitschrift für Naturforschung C* **1973**, *28*, 693−703.
(74) Deserno, M. Fluid Lipid Membranes: From Differential Geometry to Curvature Stresses. *Chem. Phys. Lipids* **2015**, *185*, 11−45. Membrane mechanochemistry: From the molecular to the cellular scale.
(75) Roos, W. H.; Bruinsma, R.; Wuite, G. J. L. Physical Virology. *Nat. Phys.* **2010**, *6*, 733−743.
(76) Kondylis, P.; Schlicksup, C. J.; Brunk, N. E.; Zhou, J.; Zlotnick, A.; Jacobson, S. C. Competition Between Normative and Drug-induced Virus Self-assembly Observed with Single-particle Methods. *J. Am. Chem. Soc.* **2019**, *141*, 1251−1260.
(77) Klumpp, K.; Lam, A. M.; Lukacs, C.; Vogel, R.; Ren, S.; Espiritu, C.; Baydo, R.; Atkins, K.; Abendroth, J.; Liao, G.; et al. High-resolution Crystal Structure of A Hepatitis B Virus Replication Inhibitor Bound to the Viral Core Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15196−15201.
(78) Liu, C.; Fan, G.; Wang, Z.; Chen, H.-S.; Yin, C.-C. Allosteric Conformational Changes of Human HBV Core Protein Transform Its Assembly. *Sci. Rep.* **2017**, *7*, 1−9.
(79) Stray, S. J.; Ceres, P.; Zlotnick, A. Zinc Ions Trigger Conformational Change and Oligomerization of Hepatitis B Virus Capsid Protein. *Biochemistry* **2004**, *43*, 9989−9998.
(80) Zhao, Z.; Wang, J. C.-Y.; Gonzalez-Gutierrez, G.; Venkatakrishnan, B.; Asor, R.; Khaykelson, D.; Raviv, U.; Zlotnick, A. Structural Differences Between the Woodchuck Hepatitis Virus Core Protein in the Dimer and Capsid States Are Consistent with Entropic and Conformational Regulation of Assembly. *Journal of virology* **2019**, *93*, e00141.
(81) Harms, Z. D.; Haywood, D. G.; Kneller, A. R.; Selzer, L.; Zlotnick, A.; Jacobson, S. C. Single-particle Electrophoresis in Nanochannels. *Analytical chemistry* **2015**, *87*, 699−705.
(82) Kondylis, P.; Schlicksup, C. J.; Zlotnick, A.; Jacobson, S. C. Analytical Techniques to Characterize the Structure, Properties, and Assembly of Virus Capsids. *Analytical chemistry* **2019**, *91*, 622−636.
(83) Katen, S. P.; Chirapu, S. R.; Finn, M. G.; Zlotnick, A. Trapping of Hepatitis b Virus Capsid Assembly Intermediates by Phenylpropenamide Assembly Accelerators. *ACS Chem. Biol.* **2010**, *5*, 1125−1136.
(84) Asor, R.; Selzer, L.; Schlicksup, C. J.; Zhao, Z.; Zlotnick, A.; Raviv, U. Assembly Reactions of Hepatitis B Capsid Protein into Capsid Nanoparticles Follow A Narrow Path through A Complex Reaction Landscape. *ACS Nano* **2019**, *13*, 7610−7626.
(85) Perkett, M. R.; Hagan, M. F. Using Markov State Models to Study Self-assembly. *J. Chem. Phys.* **2014**, *140*, 214101.
(86) Prinz, J.-H.; Keller, B.; Noé, F. Probing Molecular Kinetics with Markov Models: Metastable States, Transition Pathways and Spectroscopic Observables. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912−16927.
(87) Packianathan, C.; Katen, S. P.; Dann, C. E.; Zlotnick, A. Conformational Changes in the Hepatitis B Virus Core Protein Are Consistent with A Role for Allostery in Virus Assembly. *J. Virol.* **2010**, *84*, 1607−1615.
(88) Lazaro, G.; Hagan, M. Allosteric Control in Icosahedral Capsid Assembly. *J. Phys. Chem. B* **2016**, *120*, 6306−6318.
(89) Grime, J.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. Coarse-grained Simulation Reveals Key Features of HIV-1 Capsid Self-assembly. *Nat. Commun.* **2016**, *7*, 1−11.
(90) Biela, A. P.; Naskalska, A.; Fatehi, F.; Twarock, R.; Heddle, J. G. Programmable Polymorphism of A Virus-like Particle. *communications materials* **2022**, *3*, 1−9.
(91) Wagner, J.; Erdemci-Tandogan, G.; Zandi, R. Adsorption of Annealed Branched Polymers on Curved Surfaces. *J. Phys.: Condens. Matter* **2015**, *27*, 495101.
(92) Perilla, J. R.; Hadden, J. A.; Goh, B. C.; Mayne, C. G.; Schulten, K. All-atom Molecular Dynamics of Virus Capsids As Drug Targets. *journal of physical chemistry letters* **2016**, *7*, 1836−1844.
(93) Pérez-Segura, C.; Goh, B. C.; Hadden-Perilla, J. A. All-atom Md Simulations of the HBV Capsid Complexed with At130 Reveal Secondary and Tertiary Structural Changes and Mechanisms of Allostery. *Viruses* **2021**, *13*, 564.
(94) Videbæk, T. E.; Fang, H.; Hayakawa, D.; Tyukodi, B.; Hagan, M. F.; Rogers, W. B. Tiling A Tubule: How Increasing Complexity Improves the Yield of Self-limited Assembly. *J. Phys.: Condens. Matter* **2022**, *34*, 134003.
(95) Hayakawa, C.; Videbæk, C.; Fang, C.; Hall, C.; Sigl, C.; Feiglc, C.; Dietz, C.; Fraden, C.; Hagan, C.; Grason, C.; Rogers, C.Geometrically Programmed Self-limited Assembly of Tubules Using DNA-origami Colloids*arXiv* accessed July 30, 2022 2021 DOI: 10.48550/arXiv.2203.01421.
(96) Sigl, C.; Willner, E. M.; Engelen, W.; Kretzmann, J. A.; Sachenbacher, K.; Liedl, A.; Kolbe, F.; Wilsch, F.; Aghvami, S. A.; Protzer, U.; et al. Programmable Icosahedral Shell System for Virus Trapping. *Nature materials* **2021**, *20*, 1281−1289.
(97) Starr, C. A.; Barnes, L. F.; Jarrold, M. F.; Zlotnick, A. Hysteresis in Hepatitis B Virus (HBV) Requires Assembly of Near-perfect Capsids. *Biochemistry* **2022**, *61*, 505−513.
(98) Freund, S. M.; Johnson, C. M.; Jaulent, A. M.; Ferguson, N. Moving Towards High-resolution Descriptions of the Molecular Interactions and Structural Rearrangements of the Human Hepatitis B Core Protein. *Journal of molecular biology* **2008**, *384*, 1301−1313.

(99) Bourne, C. R.; Katen, S. P.; Fulz, M. R.; Packianathan, C.; Zlotnick, A. A Mutant Hepatitis B Virus Core Protein Mimics Inhibitors of Icosahedral Capsid Self-assembly. *Biochemistry* **2009**, *48*, 1736−1742.
(100) Bergstra, J.; Yamins, D.; Cox, D. D.et al.Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. *Proceedings of the 12th Python in Science Conference*. 2013; p 20.
(101) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of molecular biology* **2007**, *372*, 774−797.
(102) Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute; http://www.ebi.ac.uk/pdbe/prot_int/pistart.html Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774−797.
(103) Zlotnick, A.; Mukhopadhyay, S. Virus Assembly, Allostery and Antivirals. *Trends Microbiol.* **2011**, *19*, 14−23.
(104) Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with Ovito−The Open Visualization Tool. *Modell. Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.
(105) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

NOTE ADDED AFTER ASAP PUBLICATION

Originally published ASAP September 2, 2022; Supporting Information links revised September 7, 2022.

https://doi.org/10.1021/acsnano.2c02119
ACS Nano 2022, 16, 13845−13859