Biophysical Journal
Article

Biophysical Society

Should Virus Capsids Assemble Perfectly? Theory and Observation of Defects

Justin Spiriti,[1] James F. Conway,[2,*] and Daniel M. Zuckerman[1,*]
[1]Department of Biomedical Engineering, Oregon Health and Science University, Portland, Oregon and [2]Department of Structural Biology, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT Although published structural models of viral capsids generally exhibit a high degree of regularity or symmetry, structural defects might be expected because of the fluctuating environment in which capsids assemble and the requirement of some capsids for disassembly before genome delivery. Defective structures are observed in computer simulations, and are evident in single-particle cryoelectron microscopy studies. Here, we quantify the conditions under which defects might be expected, using a statistical mechanics model allowing for ideal, defective, and vacant sites. The model displays a threshold in affinity parameters below which there is an appreciable population of defective capsids. Even when defective sites are not allowed, there is generally some population of vacancies. Analysis of single particles in cryoelectron microscopy micrographs yields a confirmatory $\gtrsim$15% of defective particles. Our findings suggest structural heterogeneity in virus capsids may be under-appreciated, and also points to a nontraditional strategy for assembly inhibition.

SIGNIFICANCE Structural biology technologies have enabled the visualization of virus capsid structures at high resolution, showcasing the beauty and symmetry of spherical capsids. However, viral capsids form in fluctuating environments and often must disassemble in the infection process. This report shows that both mathematical modeling and direct electron microscopy evidence indicate that defects in capsid structures may be substantially under-appreciated and could play an important role in basic and translational virology.

INTRODUCTION

Capsid assembly is a critical step in the life cycle of all viruses (1,2). During each infection cycle, capsids must assemble for each new viral replicate before it leaves the cell, and often must go through maturation steps that involve significant changes in conformation as well. One consequence of the dynamic and intrinsically stochastic nature of viral capsid assembly is that the process does not always proceed to a precise end point and that, occasionally, defective capsids form. For example, there are three forms of the herpesvirus capsid that have been isolated in the nucleus, two of which do not proceed to form fully infectious virions (3,4). The human hepatitis B virus (HBV) capsid can assemble in vivo into two structures with different numbers of quasiquivalent subunits (5). The symmetry may be either $T = 3$ (with 90 subunits per capsid) or $T = 4$ (with 120 subunits per capsid) depending on the conformational behavior of seven amino acid residues on the C-terminal region of the HBcAg capsid subunit (6–8). The nature of the structures that form may also depend strongly on the ambient conditions. For example, the cowpea chlorotic mosaic virus can form a variety of structures, including multiwalled shells, tubes, and "rosettes," depending on the pH and ionic strength (9).

The limitations of conventional structural biology techniques make it very difficult to verify the formation of defective capsid structures directly. Although structures of individual viral capsid subunits can be obtained by x-ray crystallography, most structures of complete capsids are obtained using cryoelectron microscopy (cryo-EM) (10,11). Cryo-EM images often include capsids that are distorted or incomplete, or that have additional layers of subunits (Fig. 1; (12)). However, the signal/noise ratio of individual cryo-EM images is low, so obtaining atomistic structures requires collecting, filtering, reorienting, and averaging many such images (13–16). This in turn requires discarding images of capsids that appear to be defective and exploiting

Submitted April 10, 2020, and accepted for publication September 8, 2020.
*Correspondence: james.conway@pitt.edu or zuckermd@ohsu.edu
Justin Spiriti present address is Department of Chemistry, University of Missouri-St. Louis, St. Louis, Missouri 63121.
Editor: Yuji Sugita.
https://doi.org/10.1016/j.bpj.2020.09.021

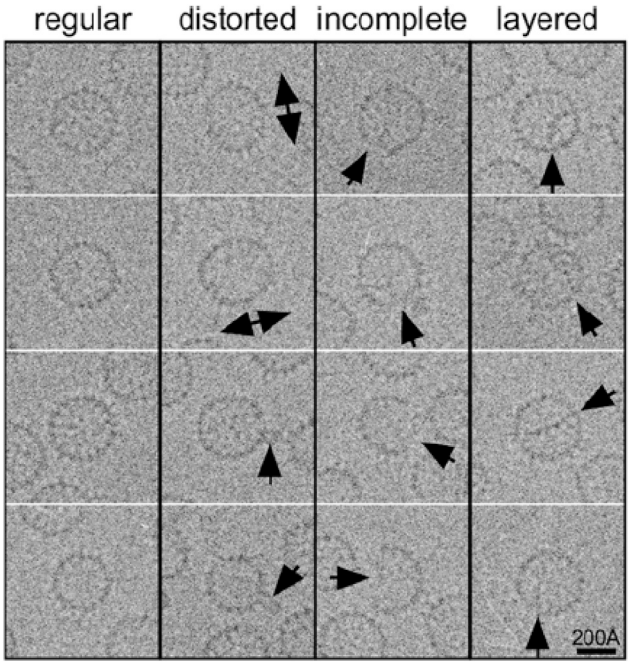


FIGURE 1 Heterogeneity in raw cryo-EM data. Shown are selected cryo-EM images of hepatitis B capsid particles. Arrows indicate structural variation.

symmetry in the averaging process, which artificially imposes a corresponding symmetry on the resulting structure. Because defective capsids are not uniform, this method cannot be applied to calculate high-resolution structures of defective capsids. Similarly, x-ray crystallography exploits the perfect repetition of the crystal lattice, thus excluding observation of irregular or infrequent capsids resulting from defects in assembly.

Computer simulations can enable detailed examinations of the mechanism of virus capsid assembly and the types of structures that can form (17–21). Several simulations produced unusual, malformed structures, such as oblate capsids lacking icosahedral symmetry or aggregates composed of several partially formed capsids joined together in an irregular fashion (22–26). A recent study of the assembly of the $T = 4$ form of the HBV capsid using a semi-atomistic model yielded trajectories with persistent defects, despite the lowest energy conformation corresponding to a defect-free fully symmetrical structure (27).

Defective capsid formation can also be studied from a more theoretical perspective via thermodynamics and statistical mechanics (1,2,28–33). A still-influential early model of equilibrium assembly for icosahedral capsids used 12 pentagonal subunits that unite to form a dodecahedron following a single, minimal-energy pathway for assembly (28). This 12-pentamer model indicated that, for a range of plausible concentrations of free subunits, the population of intermediates was very low compared to that of either free subunits or complete capsids—i.e., two-state behavior: fully formed or monomeric. However, capsid subunits were considered rigid and able to exist in only two states: either fully attached to the capsid or completely detached from it. Consequently, the only possible defects were vacancies. In addition, the assumption of a single minimal-energy pathway breaks down for large viruses because of the increased complexity of the structure and the larger number of sites in which a subunit can be attached to a growing capsid.

To develop an improved understanding of potential defects in capsids, we use the HBV capsid as a demonstration system because cryo-EM images of this capsid available to us show a variety of defects. Some of the defects represented include distorted, nonspherical capsids; incomplete capsids with gaps in their surface; and layered capsids with subunits forming additional surfaces inside an otherwise fully formed capsid. Several instances of these defects are shown in Fig. 1. Importantly, HBV capsids assemble spontaneously from a solution of their subunit proteins, without genetic material, and have been the subject of extensive study (12,34–36).

Extensive experimental studies of the thermodynamics and kinetics of HBV capsid assembly have been performed (28,35,37). At low salt concentration, significant concentrations of species with molecular masses intermediate between individual subunits and complete capsids were not reported, apparently confirming the two-state picture of (28). However, the resolution of size exclusion chromatography and light scattering as mass-measurement techniques is insufficient to distinguish between complete, perfect capsids and those that have small defects or vacancies. Furthermore, at high salt concentrations, charge depletion mass spectrometry (CDMS) experiments suggest the presence of significant numbers of trapped intermediates with ~90 dimers. These experiments also suggest the presence of defective, overgrown capsids whose masses are greater than that of a perfect capsid (38,39). However, these experiments likewise do not have the resolution to identify small vacancies and do not provide detailed structural information. A study of the assembly of woodchuck HBV capsids using both CDMS and cryo-EM identified significant numbers of capsids with more than 120 dimers, where the additional dimers result in elongation along a fivefold axis or spiral-like deformities (40).

Here, we take a statistical-mechanical approach to the problem of determining the equilibrium distribution of capsid states, defining a lattice model (Fig. 2). Our model is similar to the lattice gas model for studying phase transitions in critical fluids (41,42) but defined on a contact graph that reflects the geometry of the HBV capsid. Unlike prior studies, this model incorporates the possibility of capsid subunits being partially attached or being in a conformational state such that their interactions with a partially formed capsid are not as strong as the intersubunit

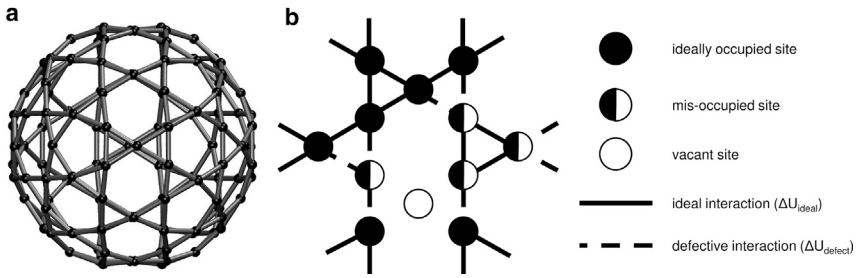


FIGURE 2 Lattice model for Hepatitis B virus capsid. (*a*) The perfect capsid, with all sites ideally occupied. (*b*) Close-up view of a subset of sites showing one possible mixture of ideally occupied sites (*black circles*), misoccupied sites (*black half-circles*), and vacant sites (*white circles*). The interaction energy between ideally occupied sites or between misoccupied sites is ΔG_{ideal} (*solid line*), whereas the interaction energy between an ideally occupied and a misoccupied site is ΔG_{mixed} (*dashed line*).

interactions in a fully formed capsid. The model can be simulated numerically using the Metropolis Monte Carlo method (43) and also solved analytically under the approximation that defects in capsids are isolated. Because of our primary interest in near-perfect capsids, the approximate calculation is highly useful.

MATERIALS AND METHODS

Model for capsid formation

We construct a statistical-mechanical model for capsid assembly and examine its equilibrium behavior, which depends only on the energy function via its parameters. Given a constant chemical potential for free subunits, we can determine the distribution of sizes of partially formed capsids. We emphasize that we are not estimating the free-energetic parameters of the model, but rather exploring the consequences of a physically plausible range of parameter choices.

Physically, each discrete state of our model represents a "set" of continuum configurations in the underlying configuration space. Therefore, all state energies are intrinsically free energies that account for both energetic and entropic characteristics of the underlying configurational subensembles (44). The entropic interactions include both configurational entropy due to minor structural fluctuations as well as solvation entropy changes including hydrophobic effects.

We employ a simple lattice-like graphical model with N sites (Fig. 2), in which each site i can be in one of three states: unoccupied, "ideally" occupied (i.e., in the orientation suitable for the symmetric capsid), or misoccupied (i.e., occupied but misoriented compared to the symmetric configuration). A contact graph $\mathcal{G}$ (Fig. 2) shows which sites interact with each other. Thus, the free energy of a capsid configuration $\mathbf{x}$ is given by

$$G(\mathbf{x}) = \sum_{\text{edges } ij \in \mathcal{G}} \Delta G_{\text{int}}(x_i, x_j), \tag{1}$$

where $\Delta G_{\text{int}}(x_i, x_j)$ is an interaction free energy that depends on the occupation states x_i and x_j of the two sites i and j as given below.

The interaction energy ΔG_{int} can take on one of three possible values: ΔG_{ideal} for interactions between ideally occupied sites, ΔG_{mixed} for interactions between a ideally occupied site and a misoccupied site, or ΔG_{defect} for interactions between misoccupied sites. In general, we expect that interactions between ideally occupied and misoccupied sites will be weaker than those between ideally occupied sites. We defined the difference $\Delta\Delta G = \Delta G_{\text{mixed}} - \Delta G_{\text{ideal}}$ to be the difference between the two interaction energies and is always assumed positive, i.e., mixed interactions are weaker than ideal. We also assumed that the interaction between two adjacent misoccupied sites (ΔG_{defect}), which in principle could include ideal pair geometry, would overall be slightly weaker than that between ideally occupied sites (ΔG_{ideal}), so that $\Delta G_{\text{defect}} = \Delta G_{\text{ideal}} + 1.0 k_B T$. Our results were not sensitive to this choice. Note that all of these interaction energies account for average effects of configurational fluctuations and hence represent effective free energies.

The probability of any given configuration $\mathbf{x}$ can be calculated based on its free energy. To probe the distribution of the number of occupied sites both in terms of energetic parameters and concentration effects, we work in the grand-canonical ensemble, in which the chemical potential of capsid subunits μ is fixed and the number of assembled subunits self-adjusts accordingly. This corresponds to considering any partially formed capsid as being in equilibrium with an ideal solution of free subunits. The probability of a given configuration $\mathbf{x}$ is given by

$$\begin{aligned} p(\mathbf{x}) &\propto \exp[-\beta G(\mathbf{x})] \exp[-\beta \mu N(\mathbf{x})] \\ &= \exp\{-\beta[G(\mathbf{x}) - \mu N(\mathbf{x})]\}, \end{aligned} \tag{2}$$

where $N(\mathbf{x})$ is the number of occupied sites for configuration $\mathbf{x}$. The total probability of having M subunits in the capsid is obtained by summing over all configurations having M occupied sites:

$$p(N) = \frac{\sum_{\mathbf{x} \text{ s.t. } N(\mathbf{x}) = M} \exp\{-\beta[G(\mathbf{x}) - \mu N(\mathbf{x})]\}}{\sum_{\mathbf{x}} \exp\{-\beta[G(\mathbf{x}) - \mu N(\mathbf{x})]\}} \tag{3}$$

The chemical potential μ is usually approximated by an ideal relation to concentration

$$\mu = k_B T \ln \frac{[\text{subunit}]}{[\text{subunit}]_{\text{ref}}}, \tag{4}$$

though the true dependence is generally more complicated (45). The functional form of μ does not affect the results presented below.

Application to the hepatitis B virus capsid

We apply the lattice model described above to describe the thermodynamics of defective configurations of the HBV $T = 4$ capsid to a first approximation. The capsid is composed of 120 dimers, each of which consists of a pair of sequentially identical subunits, each contributing a pair of helices to the 4-helix bundle interface and further linked by a disulfide bridge (46). Experimentally, these dimers have been found to be quite robust (35) and act as building blocks for the capsid. Consequently, we considered each dimer to constitute a site in our model. Furthermore, individual missing dimers seem to be a common feature in both the cryo-EM micrographs presented here and the defective conformations produced in our previous simulations (27). In the simulations, we also observed partially attached dimers which retained a significant degree of orientational freedom. Although no clear examples of partially attached dimers were observed in the cryo-EM images, this motivated our choice to define three possible states for each site: fully occupied, empty, or misoccupied.

Each of the dimers that constitute the HBV capsid is a sequentially identical homodimer, with very slight structural differences between dimers. Furthermore, estimates of the free energy of interactions between individual subunits show that they are approximately similar in strength (35). Consequently, as a first order approximation, we chose to define only one chemical potential μ and set of interaction strengths ΔG_{ideal} and ΔG_{mixed} for our model. We also assumed only local interactions between neighboring dimers, constructing the contact graph $\mathscr{G}$ for the virus from the structure of the capsid given in the Protein Data Bank (PDB: 1QGT) (46) by including an edge for each pair of dimers in which there is at least one pair of α-carbons within 10 Å of each other in the complete structure of the capsid. In this graph, there are four nearest neighbors for each site.

Analysis of defective images in cryo-EM micrographs of HBV capsids

Three historical micrographs of the HBV capsid assembled in vitro from expressed HBcAg truncated at residue 147 (6,12) were analyzed to identify and classify images of defective capsids. Briefly, samples were rapidly vitrified on copper grids with lacey carbon film, transferred on a Gatan 626 cryoholder (Gatan, Pleasanton, CA) into a Phillips CM200 cryoelectron microscope (FEI/Thermo Fisher Scientific, Portland, OR) operating at 120 kV, and imaged by low-dose techniques at 38,000x on film. The data were digitized on an SCAI flatbed scanner (Z/I Imaging, Huntsville, AL) to yield a pixel size at the sample of 1.84 Å. Scans of the three micrographs (which included one used in (12)) were then trimmed and processed using ImageMagick (https://imagemagick.org) to remove the edges and to eliminate the data window imprinted on the micrograph to reduce spurious regions of high contrast that would otherwise interfere with subsequent particle picking and extraction. The modified micrograph scans were then converted to MRC format using em2em (Image Science Software, Berlin, Germany) for further processing.

Images of capsids were then picked and extracted from the micrographs using RELION (14,47,48). An initial set of 89 particles was manually picked and extracted from the micrographs, and subjected to two-dimensional (2D) classification into five classes; one of the class means was then used as a template for template-based autopicking. A threshold value of −0.6 was used to maximize the number of particles selected while excluding unwanted debris as much as possible. This resulted in a first stack of 17,790 particle images, which was subjected to 2D classification into 10 classes to remove partial capsids and other debris. Two of the classes were selected to provide a final filtered stack of 4940 images. A random sample of 100 images was then drawn (without replacement) from this stack and examined manually to identify any visible defects. Statistics on the types of defects visible in the sample of images were then collected. The filtered stack was also subjected to 2D classification with 50 initial clusters and 50 iterations of the expectation-maximization (E-M) algorithm used by RELION.

Approximate analytical approach to the model based on assuming isolated defects

It is possible to obtain an analytical expression for the partition function of the above model if we make the approximation that all defects (either empty sites or misoccupied sites) are isolated from each other. This is a good approximation if there are few defects, which is the regime of primary interest in which capsids are generally stable.

The free energy change for each type of defect, a vacancy, or misoccupied site, can easily be obtained for a model with N sites, each of which has k nearest neighbors. Starting from a perfect capsid of free energy $G_{\text{perf}} = -N\mu + (Nk/2)\Delta G_{\text{ideal}}$, changing one site to empty removes the interactions along k edges, and the free energy change is $\Delta G_{\text{vac}} = \mu - k\Delta G_{\text{ideal}}$. If instead we change one site to a misoccupied site, the free energy difference is $\Delta G_{\text{mis}} = k\Delta G_{\text{ideal}} - k\Delta G_{\text{mixed}} \equiv -k\Delta\Delta G$.

The partition function can now be obtained from straightforward combinatorics. The number of configurations with N_v vacant sites and N_m misoccupied sites is $(N!/N_v!N_m!(N-N_v-N_m)!)$. When all defects are assumed to be isolated (not neighboring another defect), the free energy difference from the perfect capsid is simply additive,

$$N_v\Delta G_{\text{vac}} + N_m\Delta G_{\text{mis}} \ . \tag{5}$$

The grand-canonical partition function is then given by

$$\Xi = e^{-G_{\text{perf}}/k_BT} \sum_{N_v,N_m=0}^{N} \frac{N!}{N_v!N_m!(N-N_v-N_m)!} \times \exp\left[-\left(N_v\,\Delta G_{\text{vac}} - N_m\,\Delta G_{\text{mis}}\right)/k_BT\right] \tag{6}$$

which is a trinomial form that can be evaluated exactly:

$$\Xi = e^{-G_{\text{perf}}/k_BT}\left(1 + e^{-\Delta G_{\text{vac}}/k_BT} + e^{-\Delta G_{\text{mis}}/k_BT}\right)^N \tag{7}$$

With the partition function in hand, the probability of the "single" perfect capsid configuration is given by

$$p_{\text{perf}} = \frac{e^{-G_{\text{perf}}/k_BT}}{\Xi} = \left(\frac{1}{1+\exp[-(\mu - k\Delta G_{\text{ideal}})/k_BT] + \exp[-k\Delta\Delta G/k_BT]}\right)^N, \tag{8}$$

which is used below to study more general spherical capsids, beyond HBV. Importantly, the only structural properties of the capsid on which this expression depends are the number of sites N and number of nearest neighbors k. A plot of Eq. 8 is given in Fig. S5.

The approach outlined here may be extended to conditions where defects are not isolated by using additional virial-like corrections to the isolated defect approximation are described in Appendix 2 in the Supporting Material. At higher defect concentrations, other structural properties of the capsid besides N and k, such as the overall topology of the capsid, may become relevant.

Monte Carlo simulations of the model

In addition to the analytical approximation just given, we study the full model described in Model for capsid formation below using Metropolis Monte Carlo (MC) simulation. In Metropolis MC, trial moves attempt to change the configuration and are accepted with probability $\min(1,\exp[-\beta\Delta H(\mathbf{x})])$, where $H(\mathbf{x})$ is the appropriate grand-canonical potential corresponding to Eq. 2) defined by

$$H(\mathbf{x}) = G(\mathbf{x}) - \mu N(\mathbf{x}) = -\mu N(\mathbf{x}) + \sum_{\text{edges } ij \in \mathscr{G}} \Delta G_{\text{int}}(x_i, x_j), \tag{9}$$

where N is the total number of subunits and ΔH is the change in H in going from an old to a new configuration.

We performed a set of MC simulations for our model as applied to the HBV capsid, using the contact graph $\mathcal{G}$ constructed from its structure as described above. MC simulations of the model were performed with varying values of μ, ΔG_{ideal}, and ΔG_{mixed} to map out a "phase diagram" for the system. The MC moves for these simulations consisted of choosing a site i at random and changing its state. Each simulation lasted for 1 billion trial moves, and the average number and distribution of the number of occupied sites were computed. The simulations were performed on a grid in μ-$\Delta G_{\text{ideal}}-\Delta G_{\text{mixed}}$ space in which μ ranged from $-6k_BT$ to $6k_BT$ in steps of $0.2k_BT$, ΔG_{ideal} ranged from $-5k_BT$ to 0 in steps of $0.25k_BT$, and ΔG_{mixed} ranged from ΔG_{ideal} to 0 in steps of $0.25k_BT$. Only simulations in which $|\Delta G_{\text{mixed}}| \leq |\Delta G_{\text{ideal}}|$ were run.

To be considered a "partially formed capsid," a collection of subunits must be touching each other, so that they would be expected to move together because of the forces holding them together but independently of any other collection of subunits. Consequently, for the purposes of analyzing this simulation, a "partially formed capsid" was defined as a set of occupied sites (either ideally occupied or misoccupied) that span a connected subgraph of the contact graph $\mathcal{G}$. Therefore, after each trial move of the simulation, the subgraph of $\mathcal{G}$ spanned by the occupied sites was determined and divided into connected components. The distribution of the sizes of these subcomponents was then used for further analysis. The probability of perfect capsids (in which every site is ideally occupied) was determined separately.

RESULTS

Defective capsids in cryo-EM images

Analysis of cryo-EM image data by random sampling of single-particle images yields populations of different types of defects and provides strong motivation for the theoretical analysis to follow. We collected images from three cryo-EM micrographs of HBV capsids, and examined samples of images manually to determine whether they contained defects and what kinds of defects they were. We examined both the entire stack of images obtained by template-based autopicking in RELION (14,47,48), as well as a filtered stack (subset) obtained by 2D classification. We note that only 600 of 1618 total T = 4 capsid particles were used in the original structure publication (12) due primarily to technological constraints, but likely resulting in a selection of the best one-third of the particle data set and avoiding defective capsids.

About one-sixth of the images from the filtered stack have visible defects (Table 1). A larger proportion of the images in the original stack also show defects; some of these images were filtered out by the classification steps that were used to create the filtered stack. The samples of images themselves are shown in Figs. S1 and S2 and descriptions of which images exhibit what kinds of defects are shown in Tables S2 and S3. The template-based autopicker selected a large number of image centers that were located midway between several capsid particles, resulting in images that consisted of parts of multiple capsids. Almost all of these images were filtered out by the 2D classification.

The cluster mean images obtained by automatic 2D classification of the filtered stack are shown in Fig. 3 and the distribution is shown in Table S1. During the classification, some of the clusters became empty so that RELION produced a total of 22 clusters at the end of the E-M iterations. Almost all the images (98.5%) were classified into a set of clusters representing "perfect" capsids (clusters 11, 16, 24, 26, 29, 30, 32, 41, and 46) despite visual evidence to the contrary. RELION did identify a few clusters that appear to represent defective capsids. For example, cluster 47 appears to contain HBV capsids that are missing a dimer along the outer edge as seen in the micrograph, whereas cluster 42 appears to contain incomplete capsids that are missing multiple such dimers, similar to those shown in the "incomplete" column of Fig. 1. Clusters 35 and 45 appear to contain elongated capsids similar to those portrayed in the "distorted" column of Fig. 1. However, these clusters together account for only 1.5% of the filtered stack.

TABLE 1 Classification of Defective Capsids Identified from Inspection of Electron Micrographs of the HBV Capsid

Stack	Entire	Filtered
Total number of images in stack	17,790	4940
Sample size	100	100
Images containing parts of multiple capsids	64	0
Images not containing a visible capsid, or partially obscured	3	1
Images showing extra lines, additional dimers, or "buds"	0	8
Images with one missing dimer	8	7
Images with multiple missing dimers	2	0
Images with an elongated or noncircular capsid	2	0
Images with unusually small capsids	4	0
Images with small capsids that also have a missing dimer	5	0
Total fraction of defective capsids	17%	16%

Descriptions of individual images are shown in Fig. S2 for the entire stack and Fig. S3 for the filtered stack. The images themselves are shown in Fig. S2 for the entire stack and Fig. S2 for the filtered stack.

Calculated structural heterogeneity in a hepatitis B viral capsid lattice model

To assess the propensity for defects in a tractable model suited to distinguishing perfect capsids from those with

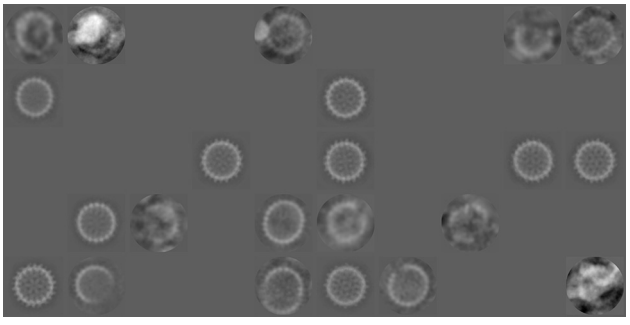


FIGURE 3 Cluster means obtained from automatic 2D classification of the filtered stack. During the classification, some of the clusters became empty so that RELION produced a total of 22 clusters at the end of the E-M iterations.

moderate defects, we analyze the lattice model described above in the Model for capsid formation section (Fig. 2).

In brief, each discrete site of the model can either be empty, occupied ideally as in a perfect capsid, or "misoccupied" in a deviated conformation or orientation. Each site interacts only with its nearest neighbors. Semimicroscopic free energy parameters account for average atomistic interactions: ΔG_{ideal} is the interaction energy between two ideally occupied sites, ΔG_{mixed} for ideal-misoccupied neighbors, and ΔG_{defect} for two misoccupied neighbors. As our analysis shows, the dominating parameter is the difference between ideal and ideal-misoccupied pairs: $\Delta\Delta G = \Delta G_{\text{mixed}} - \Delta G_{\text{ideal}}$, which generally will be positive because of weaker nonideal interactions. The chemical potential μ controls the overall subunit concentration, and we confirmed the expected increased favorability of larger (partial) capsids with higher subunit concentration and stronger ideal interactions (data not shown).

The model makes informative predictions regarding structural heterogeneity. Fig. 4 shows the distribution of the number of occupied sites, and importantly, there is always some population of imperfect capsids, which becomes dominant in some regions of parameter space. For the chosen sets of parameters, the distributions have two peaks, one corresponding to a nearly fully formed capsid, the other corresponding to an isolated subunit. In a model with fewer sites, or in typical experiments, the heterogeneity within each peak might not be resolved, depending on the experimental technique used.

Interestingly, the probability minimum between these peaks increases with $\Delta\Delta G$, which we ascribe to entropic effects, echoing recent work (49,50). Adding a third state increases the number of possible configurations, and in particular for larger capsids, there are more ways to mix fully occupied and misoccupied states. For example, when $\Delta\Delta G$ is very high the model will effectively be a two-state model, with only empty and fully occupied sites. In adding a third state to the model, the number of possible configurations with M sites occupied will be multiplied by 2^M because there are now two possible ways to occupy each site. If $\Delta\Delta G = 0$ each of these 2^M configurations will have the same energy and consequently the overall probability of observing M sites occupied will be multiplied by 2^M and the entropy of this state will be increased by $k_B M \log 2$. If $\Delta\Delta G$ is low but greater than zero, the additional configurations will have higher energy than those in the two-state model, so the entropy will still increase but not by as much. Therefore, the entropy associated with large capsids is greater when there is a misoccupied state whose energy is close to that of the fully occupied state, and hence larger capsids will be favored.

When do structurally "perfect" capsids predominate, and when are defective capsids significant? Fig. 5 shows the proportion of perfect capsids in which all sites are ideally occupied based on both MC simulations and analytical results obtained using the isolated-defect approximation Eq. 8. The proportion of perfect capsids depends predominantly on the difference $\Delta\Delta G = \Delta G_{\text{mixed}} - \Delta G_{\text{ideal}}$, and shows a critical threshold of $\Delta\Delta G \approx 2.5 k_B T$. Below this threshold (stronger nonideal interactions), the fraction of defective capsids can become significant, whereas above it, and for a sufficient interaction strength, structurally perfect capsids predominate, although the fraction is below this threshold. At $\Delta\Delta G \approx 2 k_B T$, the fraction of defective capsids is ~10%, roughly matching what was observed in the cryo-EM images. As $\Delta\Delta G$ decreases, the many imperfect but complete capsid configurations in which some sites are misoccupied, but no sites are vacant, come to have the same energy and hence entropically dominate the system, as described above. The threshold of $2.5 k_B T$ represents the point at which the lower energy of an ideally occupied site outweighs this entropic effect. We have also confirmed that the value of ΔG_{defect} has little impact on the probability of perfect capsids, as demonstrated by plots of the perfect capsid probability for various values of ΔG_{defect} (Fig. S6). We also note that the approximation is akin to a dilute fluid of noninteracting defects, and matches MC results extremely well because it is accurate in the "near-perfect" regime of interest.

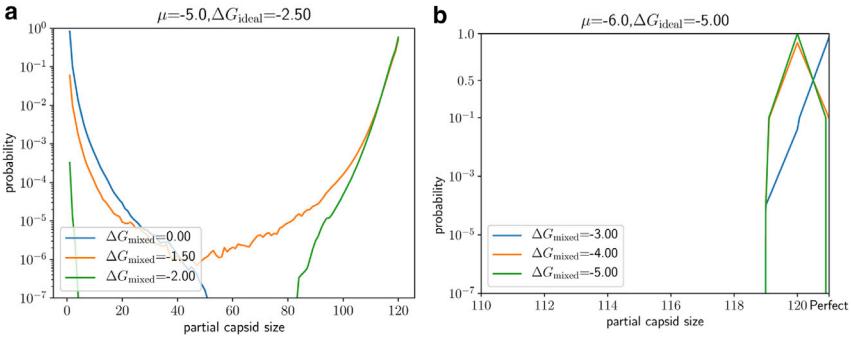


FIGURE 4 The distribution of partial capsid sizes from Monte Carlo simulation. (*a*) When $(\mu, \Delta G_{\text{ideal}}) = (-5.0, -2.5)$, the distribution of capsid sizes shows two peaks corresponding to nearly complete capsids and isolated dimers. Once $\Delta\Delta G = \Delta G_{\text{mixed}} - \Delta G_{\text{ideal}} \lesssim 2 k_B T$, as holds for all three curves, there is an appreciable population of partially formed capsids and even a depletion of full capsids for weaker interaction strengths. (*b*) When $(\mu, \Delta G_{\text{ideal}}) = (-6.0, -5.0)$, which roughly corresponds to experimental value of ΔG_{ideal}, the distribution consists primarily of larger partial and full capsids, shown here in magnified view. Note that ~10% of capsids remain imperfect even for $\Delta G_{\text{mixed}} = -3 k_B T$.

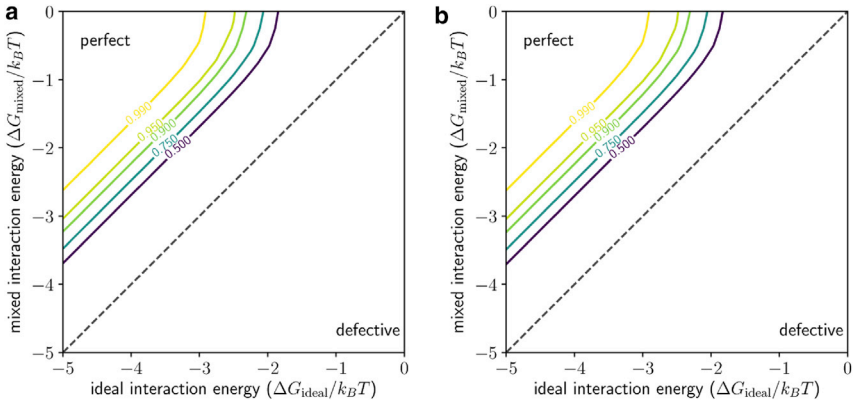


FIGURE 5 When are perfect capsids expected? (*a*) Contour plots are shown of the fraction of perfect capsids p_{perf} as determined from Eq. 8 and (*b*) p_{perf}, as determined from MC simulations, both in the $(\Delta G_{\mathrm{ideal}}, \Delta G_{\mathrm{mixed}})$ plane (chemical potential and ideal interaction energy) for $\mu = -2.0k_BT$. To see this figure in color, go online.

Model prediction for arbitrary spherical capsids

Because the model is specific to HBV only based on the number of sites (subunits) N and the number of nearest neighbors per site k (Materials and Methods), theoretical analysis can readily be extended to arbitrary spherical viral capsids based on those two parameters. Although extending our lattice strategy to nonspherical shapes/defects is not difficult in principle, here, we prioritize tractability in an effort to quantify expectations for defects in the important class of spherical viruses.

In the presumably high-concentration milieu (large chemical potential μ) in which capsids assemble, we can develop an analytical expression Eq. 8, for fractional population of perfect capsids $p_{\mathrm{perf}}(N, k, \Delta\Delta G)$ solely in terms of the two geometric parameters and the difference $\Delta\Delta G$ between mixed and ideal interactions. As before, we have fixed as constant the unimportant parameter $\Delta G_{\mathrm{defect}}$. To graphically interpret the equation, we invert it to obtain the threshold value of $\Delta\Delta G$ below which the fraction of defective capsids p_{perf} exceeds a specified threshold:

$$\Delta\Delta G/k_BT = -\frac{1}{k}\log\left(p_{\mathrm{perf}}^{-1/N} - 1\right) \qquad (10)$$

The analysis allows a theoretical characterization of whether defects should be expected based on geometric considerations. Fig. 6 shows the minimal $\Delta\Delta G$ value needed for an arbitrary spherical virus, defined by N and k, to exhibit an overwhelming population of perfect (defect-free) capsids. Below the $\Delta\Delta G$ threshold, defective interactions are competitive with ideal interactions given the entropic characteristics of the particular lattice. Above the threshold for a particular (N, k) pair, perfect capsids predominate. Hence, if the threshold is smaller, a bigger range of $\Delta\Delta G$ values leads to a near-perfect population and fewer defects should be anticipated. Conversely, a larger $\Delta\Delta G$ threshold requires stricter structural selectivity for ideal versus mixed neighbor interactions. Based on this analysis, perfect capsids are seen to be most favorable at smaller N and larger k, and defects more to be expected at larger N and smaller k.

We note that graph structure enters Eq. 10 only via N and k values for a given $\Delta\Delta G$. This stems from the dilute defect assumption, which is directly pertinent to examining perfect and near-perfect capsids.

DISCUSSION

Our study questions the conventional wisdom that assembled virus capsids are essentially defect-free. A close review of HBV micrographs and analysis of a physiologically motivated statistical-mechanical model suggests significant fractions of defects can indeed be present and are likely under a range of conditions. The defective particles present in cryo-EM micrographs are usually rejected for being quite

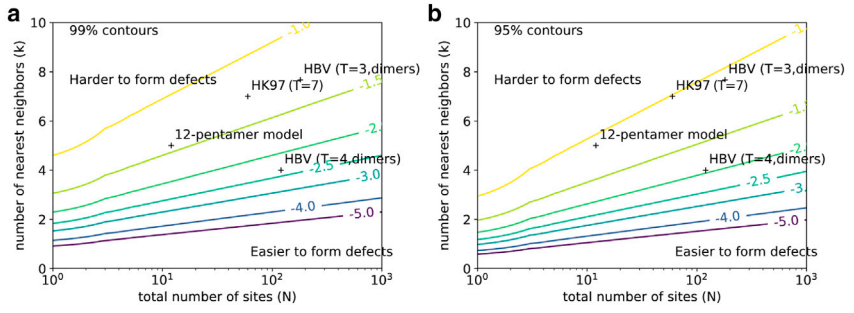


FIGURE 6 Tolerance of defects for arbitrary spherical viruses. Shown are contour plots of the minimal $\Delta\Delta G$-value needed to ensure a high proportion of capsid formation, given the total number of subunits N and the number of nearest neighbors k. If $\Delta\Delta G$ is above the minimal threshold for a given (N, k) pair, perfect capsids predominate; below the threshold, there will be a substantial fraction of defects. Contours are based on 5. (*a*) 99% capsid formation is shown; (*b*) 95% capsid formation is shown. Values of N and k for selected viruses are plotted as well. The "12-pentamer model" refers to (28). To see this figure in color, go online.

variable in structure and presumed to be physiologically off-pathway. Consistent with our findings, one previous study of woodchuck HBV capsid assembly using both CDMS and cryo-EM identified several types of defects similar to those seen here, including elongated particles and broken particles with spiral-like morphologies; ~20% of the images showed elongated particles (40). Likewise, defects are a hallmark of computational models where defects are possible (22–26).

The analysis presented in this work, relying on visual inspection of cryo-EM images, is necessarily somewhat subjective because of the high level of noise in the images, and it is possible that, for example, variations in the noise level across the images could make it appear that a dimer is missing when in fact it is not. Because of the limited number of images and the diversity of possible defects, RELION was only able to classify a small proportion of the defective images together in a class. Even if not every image of a defective capsid is genuine, however, defective capsids may constitute a sizable proportion of the sample under study reflecting the innate assembly fidelity of the complex as well as aspects of sample preparation, and may offer insights into assembly that could be exploited for antiviral or other therapeutic purposes. Such heterogeneity in cryo-EM structures is now taken to be a fundamental part of structural biology (51–54), and we hope that our findings will motivate similar analyses of other systems and other data sets.

Our theoretical model, although still greatly simplified, advances prior theoretical work in several ways. Generally speaking, our model incorporates more realistic and combinatoric possibilities than the seminal early model of Zlotnick (28) and follow-up work (30). We also find two-state behavior, mostly assembled and mostly disassembled, but the greater granularity of our model clearly shows that what appeared to be a uniform assembled state in the prior work consists of an "ensemble" of nearly fully assembled configurations in large parts of parameter space. That is, defects are common, both as vacancies and/or misassembled sites. A model somewhat similar to ours, though without misoccupied sites, also found a substantial imperfect population (32,33) Our theoretical analysis indicates that, in the dilute-defects limit appropriate to analyzing the degree of perfect capsid formation, the number of nearest neighbors (edges per node) and graph size (number of nodes) govern the likelihood of perfect capsid formation for a given set of thermodynamic parameters.

Our model has implications for the design of inhibitors of virus capsid assembly, which have received significant interest as a novel class of therapeutics for HIV (55–57) as well as for HBV (58–60). The model suggests that it is possible to design two distinct types of inhibitors: ones that block or destabilize ideal interactions between capsid subunits, and ones that stabilize nonideal interactions. In the case of HIV, both of these kinds of inhibitors have been developed (55–57). Coarse-grained simulations of the HIV-1 capsid protein suggest that the capsid assembly inhibitor PF74, a tripeptide mimic (61), works by stabilizing trimers of dimers, which then interact with partially assembled capsids in such a way as to encourage them to form incorrect assemblies (57). For HBV, the focus has been on developing modulators of capsid assembly that may either accelerate or inhibit functional capsid assembly, depending on circumstances. For example, the inhibitor BAY-41-4109 exhibits different behaviors depending on the stoichiometric ratio relative to the capsid dimers. At a low ratio, it accelerates capsid growth, whereas at a higher ratio, it causes capsids to misassemble into "voids" or polymers (62). There is a binding site on the capsid for assembly modulators that can either trigger assembly or disassembly, depending on how the modulator is bound and what conformation it adopts (63). Knowledge of the principles behind assembly should be useful for guiding the design of inhibitors that block or hijack the capsid assembly process.

One limitation of the statistical-mechanical model described here is that it considers the formation of only one capsid at a time from an ideal solution of free subunits. The possible defects are limited to vacant sites or sites with partially attached subunits. In other simulations, using structurally more realistic models, more complex structures were formed. For example, the Brooks group obtained a variety of irregular capsids, including twisted, tubular, prolate, and conical capsids, as well as partially formed capsids with missing subunits and open misaggregates consisting of two partial capsids joined together (23,24). Furthermore, the model described here is a thermodynamic model and implicitly assumes that all configurational transitions are reversible. Consequently, it cannot be used to study defective capsids occurring as a result of kinetic trapping or irreversible transitions, as has been observed in other simulations (1,18,25,64–66). This could be remedied by defining kinetic laws for transitions between configurational states.

The parameter $\Delta G_{\mathrm{ideal}}$, the interaction free energy, must include all environmental factors, including temperature and ionic strength. The experimental values for the interaction strength among HBV capsid subunits range from approximately −3.2 to −4.4 kcal/mol (35), which correspond to values of $\Delta G_{\mathrm{ideal}}$ from approximately −5 to −7 k_BT, on the upper end of the values considered here. The lower values are more consistent with physiological conditions; the strength of the intersubunit interaction increases with increasing temperature and ionic strength because of the primarily hydrophobic nature of the forces that hold the capsid together. Experimental studies comparing woodchuck and human HBV also demonstrate that the two capsid proteins, despite having ~65% sequence identity and forming very similar assembled capsid structures, have very different thermodynamics of assembly, with that of woodchuck HBV having a stronger entropic contribution to assembly (36).

There are also many ways in which our model can be modified to accommodate other viruses with protein capsids besides HBV that have different capsid compositions or geometries. For example, the model may be extended to capsid geometries with different T values by specifying different contact graphs $\mathscr{G}$. Also, many capsids have two or more types of subunits with different sequences, although usually, the number of types is limited, and the different types have similar folds (31). To accommodate this, a separate value of the chemical potential μ could be defined for each type of subunit and a separate set of interaction parameters ΔG_{ideal}, ΔG_{mixed}, and ΔG_{defect} could be used for each possible interactions between different types of subunits within the capsid.

CONCLUSIONS

To investigate systematically the potential for defect formation in virus capsids, we developed a statistical-mechanical model that incorporates more realistic capsid geometry than prior work, as well as taking into account the possibility of weaker intersubunit interactions due to partial attachment of subunits. The model, which is applicable to arbitrary spherical capsids, is motivated by the substantial fraction of defective particles ($\sim$1 in 6) occurring in a re-analyzed set of HBV capsid micrographs.

Both analytical approximations and Monte Carlo simulations of the model reveal that significant proportions of defective capsids can form when the difference in interaction strengths between ideal and defective subunits is below a threshold that depends only on geometric properties of the capsid. This implies that capsid assembly can be inhibited equally well by blocking ideal interactions or by stabilizing nonideal ones. Although significant work remains to be done in extending the model to allow a greater variety of defects, and in clarifying the physical meaning of the parameters, these results demonstrate that the concentration of defective capsids can be more significant under certain conditions than previously appreciated. We believe the proposed model provides a paradigm that will be useful in understanding and designing a wide range of experiments.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2020.09.021.

AUTHOR CONTRIBUTIONS

J.S. performed research, analyzed data, wrote computer code, analyzed cryo-EM images, and wrote initial draft of manuscript. J.F.C. designed and oversaw research, analyzed cryo-EM images, and edited manuscript draft. D.M.Z. designed and oversaw research and revised manuscript.

ACKNOWLEDGMENTS

We thank Dr. Alasdair Steven, NIH, for the cryo-EM data used to access to the HBV. We also thank Drs. David Koes, Lillian Chong, Barmak Mostofian, and Jeremy Copperman for helpful discussions and technical assistance. We also thank the Advanced Computing Center at Oregon Health and Science University for computer time.

This work was supported by NIH Grants 1R01-GM115805 and R21-AI130745 and by NSF Grants MCB-1715823 and CNS-1229064.

REFERENCES

1. Perlmutter, J. D., and M. F. Hagan. 2015. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* 66:217–239.
2. Mateu, M. G. 2013. Assembly, stability and dynamics of virus capsids. *Arch. Biochem. Biophys.* 531:65–79.
3. Gibson, W., and B. Roizman. 1972. Proteins specified by herpes simplex virus. 8. Characterization and composition of multiple capsid forms of subtypes 1 and 2. *J. Virol.* 10:1044–1052.
4. Homa, F. L., J. B. Huffman, …, J. F. Conway. 2013. Structure of the pseudorabies virus capsid: comparison with herpes simplex virus type 1 and differential binding of essential minor proteins. *J. Mol. Biol.* 425:3415–3428.
5. Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27:1–24.
6. Zlotnick, A., N. Cheng, …, P. T. Wingfield. 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry.* 35:7412–7421.
7. Watts, N. R., J. F. Conway, …, P. T. Wingfield. 2002. The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface. *EMBO J.* 21:876–884.
8. DiMattia, M. A., N. R. Watts, …, P. T. Wingfield. 2013. Antigenic switching of hepatitis B virus by alternative dimerization of the capsid protein. *Structure.* 21:133–142.
9. Lavelle, L., M. Gingery, …, J. Ruiz-Garcia. 2009. Phase diagram of self-assembled viral capsid protein polymorphs. *J. Phys. Chem. B.* 113:3813–3819.
10. Chang, J., X. A. Liu, …, W. Chiu. 2012. Reconstructing virus structures from nanometer to near-atomic resolutions with cryo-electron microscopy and tomography. *Adv. Exp. Med. Biol.* 726:49–90.
11. Veesler, D., B. M. Kearney, and J. E. Johnson. 2016. Integration of X-ray crystallography and electron cryo-microscopy in the analysis of virus structure and function. *Crystallogr. Rev.* 22:102–127.
12. Conway, J. F., N. Cheng, …, A. C. Steven. 1997. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature.* 386:91–94.
13. Rhodes, G. 2006. Crystallography Made Crystal Clear, Third Edition. Elsevier Academic Press, Amsterdam.
14. Nogales, E., and S. H. W. Scheres. 2015. Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Mol. Cell.* 58:677–689.
15. Cheng, Y., N. Grigorieff, …, T. Walz. 2015. A primer to single-particle cryo-electron microscopy. *Cell.* 161:438–449.
16. Conway, J. F., and A. C. Steven. 1999. Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution. *J. Struct. Biol.* 128:106–118.
17. Schwartz, R., P. W. Shor, …, B. Berger. 1998. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* 75:2626–2636.
18. Baschek, J. E., H. C. R Klein, and U. S. Schwarz. 2012. Stochastic dynamics of virus capsid formation: direct versus hierarchical self-assembly. *BMC Biophys.* 5:22.

19. Xie, L., G. R. Smith, …, R. Schwartz. 2012. Surveying capsid assembly pathways through simulation-based data fitting. *Biophys. J.* 103:1545–1554.
20. Perkett, M. R., and M. F. Hagan. 2014. Using Markov state models to study self-assembly. *J. Chem. Phys.* 140:214101.
21. Dykeman, E. C., P. G. Stockley, and R. Twarock. 2014. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc. Natl. Acad. Sci. USA.* 111:5361–5366.
22. Nguyen, H. D., V. S. Reddy, and C. L. Brooks, III. 2007. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett.* 7:338–344.
23. Nguyen, H. D., and C. L. Brooks, III. 2008. Generalized structural polymorphism in self-assembled viral particles. *Nano Lett.* 8:4574–4581.
24. Nguyen, H. D., V. S. Reddy, and C. L. Brooks, III. 2009. Invariant polymorphism in virus capsid assembly. *J. Am. Chem. Soc.* 131:2606–2614.
25. Hagan, M. F., and D. Chandler. 2006. Dynamic pathways for viral capsid assembly. *Biophys. J.* 91:42–54.
26. Panahandeh, S., S. Li, and R. Zandi. 2018. The equilibrium structure of self-assembled protein nano-cages. *Nanoscale.* 10:22802–22809.
27. Spiriti, J., and D. M. Zuckerman. 2015. Tabulation as a high-resolution alternative to coarse-graining protein interactions: initial application to virus capsid subunits. *J. Chem. Phys.* 143:243159.
28. Zlotnick, A. 1994. To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* 241:59–67.
29. Bruinsma, R. F., W. M. Gelbart, …, R. Zandi. 2003. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* 90:248101.
30. Zandi, R., P. van der Schoot, …, H. Reiss. 2006. Classical nucleation theory of virus capsids. *Biophys. J.* 90:1939–1948.
31. Bruinsma, R. F., and W. S. Klug. 2015. Physics of viral shells. *Annu. Rev. Conden. Ma. P.* 6:245–268.
32. Chen, J., M. Chevreuil, …, G. Tresset. 2017. Investigating the thermal dissociation of viral capsid by lattice model. *J. Phys. Condens. Matter.* 29:474001.
33. Tresset, G., J. Z. Chen, …, Y. Lansac. 2017. Two-dimensional phase transition of viral capsid gives insights into subunit interactions. *Phys. Rev. Appl.* 7:8.
34. Wingfield, P. T., S. J. Stahl, …, A. C. Steven. 1995. Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry.* 34:4919–4932.
35. Ceres, P., and A. Zlotnick. 2002. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry.* 41:11525–11531.
36. Kukreja, A. A., J. C. Y. Wang, …, A. Zlotnick. 2014. Structurally similar woodchuck and human hepadnavirus core proteins have distinctly different temperature dependences of assembly. *J. Virol.* 88:14105–14115.
37. Zlotnick, A., J. M. Johnson, …, D. Endres. 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry.* 38:14644–14652.
38. Lutomski, C. A., N. A. Lyktey, …, M. F. Jarrold. 2017. Hepatitis B virus capsid completion occurs through error correction. *J. Am. Chem. Soc.* 139:16932–16938.
39. Lutomski, C. A., N. A. Lyktey, …, M. F. Jarrold. 2018. Multiple pathways in capsid assembly. *J. Am. Chem. Soc.* 140:5784–5790.
40. Pierson, E. E., D. Z. Keifer, …, M. F. Jarrold. 2016. Charge detection mass spectrometry identifies preferred non-icosahedral polymorphs in the self-assembly of woodchuck hepatitis virus capsids. *J. Mol. Biol.* 428:292–300.
41. Chandler, D. 1987. Introduction to Modern Statistical Mechanics. Oxford University Press, Berkley, CA.
42. Pathria, R. K. 1996. Statistical Mechanics. Elsevier, Amsterdam.
43. Metropolis, N., A. W. Rosenbluth, …, E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.
44. Zuckerman, D. M. 2010. Statistical Physics of Biomolecules: An Introduction. CRC Press, Boca Raton, FL.
45. Hill, T. L. 1986. An Introduction to Statistical Thermodynamics. Dover Publications, Inc., New York.
46. Wynne, S. A., R. A. Crowther, and A. G. W. Leslie. 1999. The crystal structure of the human hepatitis B virus capsid. *Mol. Cell.* 3:771–780.
47. Scheres, S. H. W. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180:519–530.
48. Scheres, S. H. W. 2012. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* 415:406–418.
49. Asor, R., L. Selzer, …, U. Raviv. 2019. Assembly reactions of hepatitis B capsid protein into capsid nanoparticles follow a narrow path through a complex reaction landscape. *ACS Nano.* 13:7610–7626.
50. Asor, R., C. J. Schlicksup, …, U. Raviv. 2020. Rapidly forming early intermediate structures dictate the pathway of capsid assembly. *J. Am. Chem. Soc.* 142:7868–7882.
51. Heymann, J. B., N. Cheng, …, A. C. Steven. 2003. Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. *Nat. Struct. Biol.* 10:334–341.
52. Gao, H., M. Valle, …, J. Frank. 2004. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J. Struct. Biol.* 147:283–290.
53. Leschziner, A. E., and E. Nogales. 2007. Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annu. Rev. Biophys. Biomol. Struct.* 36:43–62.
54. Bilokapic, S., M. Strauss, and M. Halic. 2018. Cryo-EM of nucleosome core particle interactions in trans. *Sci. Rep.* 8:7046.
55. Tang, C., E. Loeliger, …, M. F. Summers. 2003. Antiviral inhibition of the HIV-1 capsid protein. *J. Mol. Biol.* 327:1013–1020.
56. Neira, J. L. 2009. The capsid protein of human immunodeficiency virus: designing inhibitors of capsid assembly. *FEBS J.* 276:6110–6117.
57. Pak, A. J., J. M. A. Grime, …, G. A. Voth. 2019. Off-pathway assembly: a broad-spectrum mechanism of action for drugs that undermine controlled HIV-1 viral Capsid Formation. *J. Am. Chem. Soc.* 141:10214–10224.
58. Deres, K., C. H. Schröder, …, H. Rübsamen-Waigmann. 2003. Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science.* 299:893–896.
59. Stray, S. J., J. M. Johnson, …, A. Zlotnick. 2006. An in vitro fluorescence screen to identify antivirals that disrupt hepatitis B virus capsid assembly. *Nat. Biotechnol.* 24:358–362.
60. Cho, M. H., H. Jeong, …, G. Jung. 2014. 2-amino-N-(2,6-dichloropyridin-3-yl)acetamide derivatives as a novel class of HBV capsid assembly inhibitor. *J. Viral Hepat.* 21:843–852.
61. Shi, J., J. Zhou, …, K. Whitby. 2011. Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. *J. Virol.* 85:542–549.
62. Stray, S. J., and A. Zlotnick. 2006. BAY 41-4109 has multiple effects on Hepatitis B virus capsid assembly. *J. Mol. Recognit.* 19:542–548.
63. Qazi, S., C. J. Schlicksup, …, A. Zlotnick. 2018. An assembly-activating site in the hepatitis B virus capsid protein can also trigger disassembly. *ACS Chem. Biol.* 13:2114–2120.
64. Rapaport, D. C. 2004. Self-assembly of polyhedral shells: a molecular dynamics study. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70:051905.
65. Zhang, T., and R. Schwartz. 2006. Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys. J.* 90:57–64.
66. Rapaport, D. C. 2012. Molecular dynamics simulation of reversibly self-assembling shells in solution using trapezoidal particles. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 86:051917.

Biophysical Journal, Volume *119*

Supplemental Information

Should Virus Capsids Assemble Perfectly? Theory and Observation of Defects

Justin Spiriti, James F. Conway, and Daniel M. Zuckerman

Supporting Material: Should virus capsids assemble perfectly? Theory and observation of defects

J. Spiriti, J.F. Conway and D.M. Zuckerman

October 24, 2020

1 Samples and descriptions of cryo-EM images of HBV capsids

In this section, we provide detailed information on our analysis of cryo-EM images for defective HBV capsids. S1 shows the distribution of images in clusters as determined by RELION, and Fig. S1, S2, Fig. S2, and S3 show samples of images obtained from the micrographs and our descriptions of them based on our manual examination.

Cluster	Number of images
1	1
2	2
5	1
9	1
10	1
11	250
16	358
24	525
26	451
29	569
30	584
32	605
33	1
35	18
36	1
38	1
41	1041
42	34
45	7
46	481
47	7
50	1

Table S1: Distribution of images in clusters obtained from automatic 2D classification of the filtered stack. Clusters are numbered so that clusters 1-10 correspond to the first row of Fig. 3, 11-20 to the second row, and so on.

Description	Images
Partial capsids	1, 3, 5, 6, 8, 10, 11, 13, 15-18, 21-23, 26-36, 39, 40-50, 51, 53, 56, 59, 62, 66, 68, 70, 73, 74, 76, 77, 81-83, 85, 86, 88-91, 95-97, 99, 100
No visible capsid	7, 24, 87
Missing dimer	19, 20, 37, 38, 58, 60, 66, 72
Small	25, 69, 71, 78
Small with missing dimer	9, 65, 67, 84, 93
Elongated	57, 61

Table S2: Description of defects in the sample of 100 images drawn from the entire stack. The images are numbered such that images 1-10 correspond to the top row of Fig. S1, 11-20 to the second row, and so on.

Description	Images
Missing dimer	1, 18, 22, 58, 72, 89, 95
Missing multiple dimers	16, 42
Extra lines across capsid	35
Additionally attached dimer, “bud”, or other extra appendage	7, 13, 17, 27, 53, 69, 86
Partially obscured by gray rectangle	60

Table S3: Description of defects in the sample of 100 images drawn from the filtered stack of 4940 images created by 2D classification. The images are numbered such that images 1-10 correspond to the top row of Fig. S2, 11-20 to the second row, and so on.

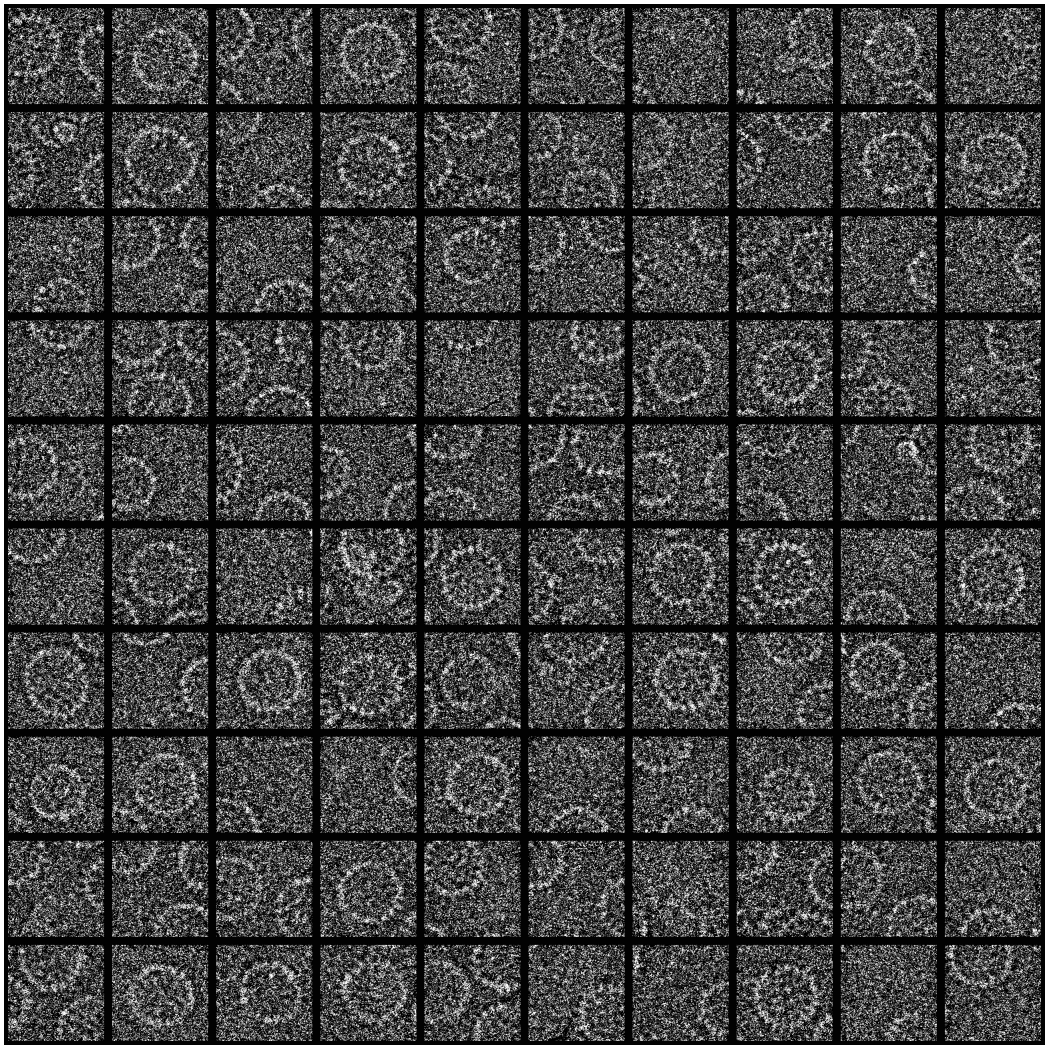


Figure S1: Sample of 100 images of the HBV capsid taken from the entire stack. The larger intact particles are 35 nm in diameter.

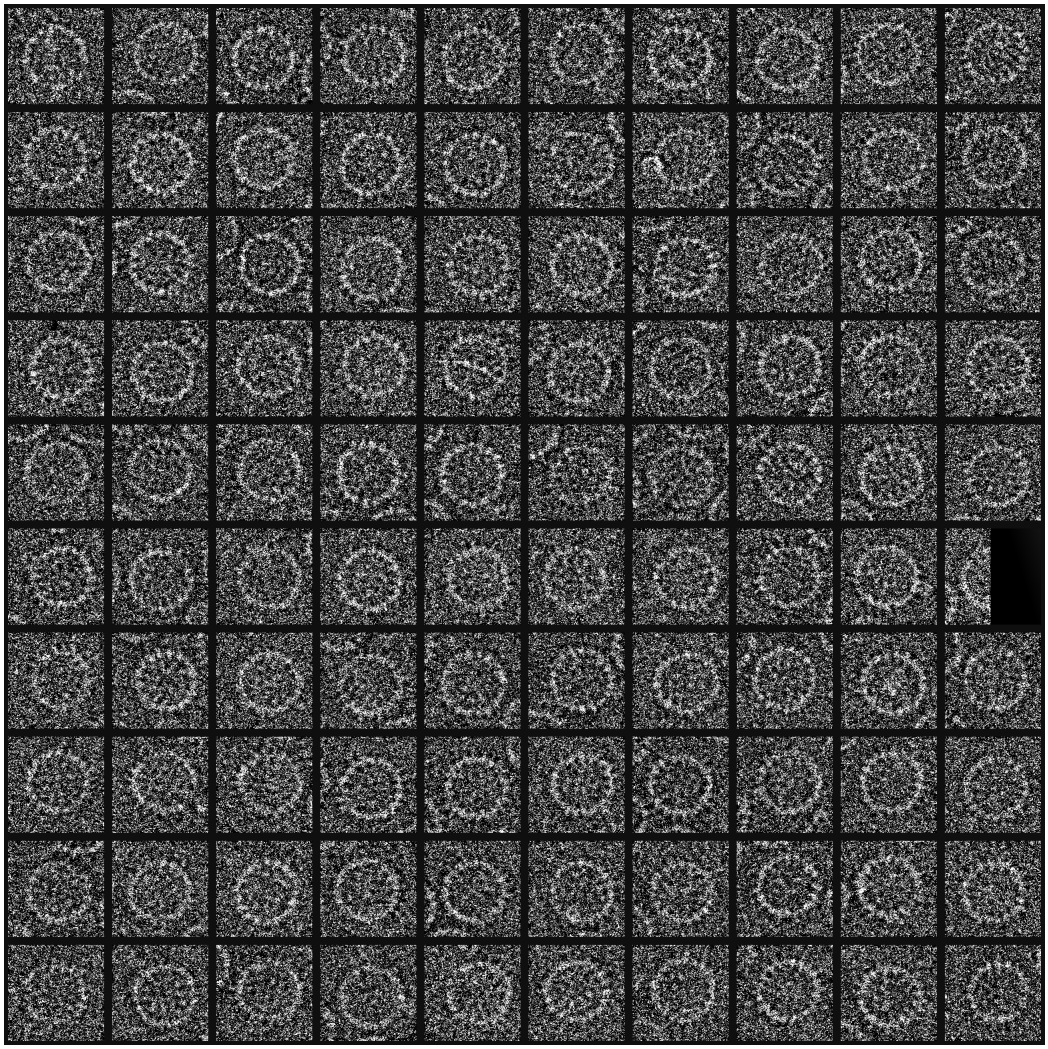


Figure S2: Sample of 100 images of the HBV capsid taken from the filtered stack. The larger intact particles are 35 nm in diameter.

2 Virial-like corrections to the isolated-defect approximation for the lattice model

Given a specific contact graph $\mathcal{G}$, the approximate analytical approach to the lattice model described in the main text of the paper may be extended to incorporate defects that are not isolated. This is done by incorporating additional terms in higher powers of the fugacities z_v and z_m of defects involving vacant or mis-occupied sites, which are defined as $z_v = \exp[-\Delta G_{\rm vac}/k_BT]$ and $z_m = \exp[-\Delta G_{\rm mis}/k_BT]$. For example, in the hepatitis B capsid, of the $120!/(2!\,118!) = 7140$ configurations with exactly two empty sites, 240 have the sites touching each other. The difference in energy between these configurations and the perfect capsid is $2\mu - 7\Delta G_{\rm ideal}$, not $2\mu - 8\Delta G_{\rm ideal}$, which would be the energy difference associated with two isolated empty sites. This is because a defect of this type removes the interactions among only seven edges rather than eight (see Fig. S3). Thus the correction to Ξ needed to account for defects of this type is given by

$$\begin{aligned}\frac{\Xi}{\exp[-G_{\rm perf}/k_BT]} &= (1+z_v+z_m)^{120} - 240z_v^2(1+z_v+z_m)^{113} + \\ &\quad 240\exp[-(2\mu-7\Delta G_{\rm ideal})](1+z_v+z_m)^{113} \\ &= (1+z_v+z_m)^{120} - \\ &\quad 240z_v^2[\exp(-\Delta G_{\rm ideal})-1](1+z_v+z_m)^{113}\end{aligned}$$

where the second term in the first line subtracts out the incorrect terms in Ξ and the third term replaces them with the correct terms. The factor of $(1+z_v+z_m)^{113}$ accounts for the possibility of isolated defects among the 113 sites not occupied by or touching the two empty sites that are together. In addition to this, two other correction terms in z_vz_m and z_m^2 would be needed to account for defects in which one of the touching sites was partially attached and the other empty, or both partially attached. Additional terms in higher powers of z_v or z_m would be needed for defects involving three or more sites touching each other.

This approach is very reminiscent of the well-known virial expansion,[1, 2], where the effect of intermolecular forces in a real gas can be expressed as correction terms in powers of the density, relative to the partition function of an ideal gas. These correction terms can be determined using cluster diagrams. For viruses, the correction terms depend on the details of the contact graph and the particular defects that are possible, and can be derived in a similar manner to the example given above.

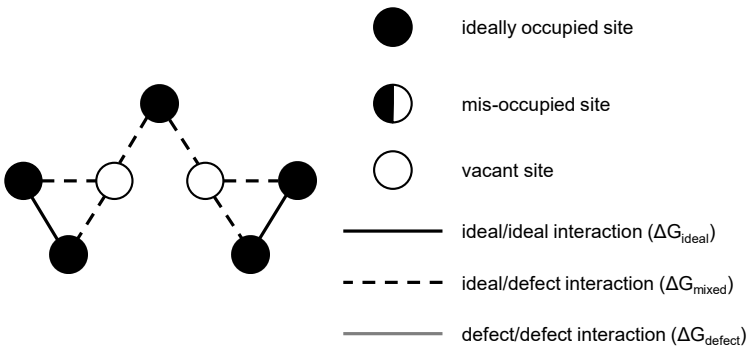


Figure S3: Example of a defect in the HBV capsid that consists of two neighboring vacant sites surrounded by occupied sites and therefore violates the isolated-defects approximation used in the analytical treatment of the lattice model described in the main text. The analytical treatment may be extended to cover this case by using a virial-like expansion.

3 Additional results from Monte Carlo simulations of the lattice model

In this section, we provide additional, more detailed results from both simulations and analytical treatment of our lattice model. Fig. S4 shows the dependence of the average connected capsid size on μ, ΔG_{ideal}, and $\Delta\Delta G$. The plots in Fig. S4 indicate that capsid formation is more favorable for a higher μ (implying a higher concentration of free subunits in equilibrium with the capsids) and for higher ΔG_{ideal} (stronger interactions between subunits) for all values of $\Delta\Delta G$. Since plots with different values of $\Delta\Delta G$ are very similar to one another, it is also the case that the average number of occupied sites does not vary much with $\Delta\Delta G$. Fig. S5 and Fig. S6 show the differences in $p_{\text{perfect capsid}}$ between the analytical treatment and simulations of the lattice model.

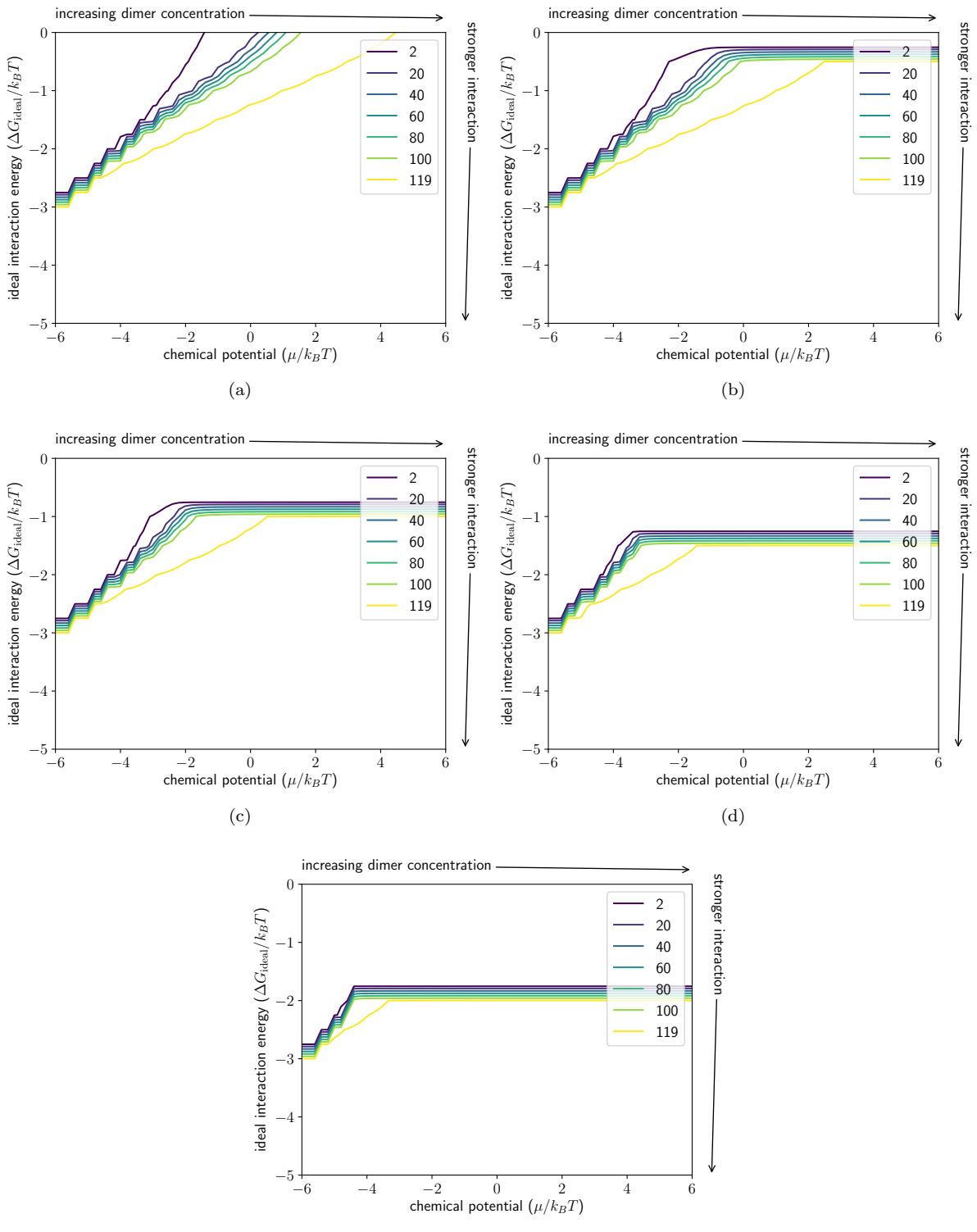


Figure S4: Dependence of capsid formation on chemical potential μ and interaction strength ΔG_{ideal}. Contour plot of the average connected capsid size (including mis-occupied sites) in Monte Carlo simulations of our model on the hepatitis B virus contact graph in the $(\mu, \Delta G_{\text{ideal}})$ plane for (a) $\Delta\Delta G = 0$; (b) $\Delta\Delta G = 0.5k_BT$; (c) $\Delta\Delta G = 1.0k_BT$; (d) $\Delta\Delta G = 1.5k_BT$; (e) $\Delta\Delta G = 2.0k_BT$. Note that for ideal solutions $\Delta\mu \sim -k_BT \ln[\text{dimer}]$. No simulations were performed where $|\Delta G_{\text{mixed}}| \leq |\Delta G_{\text{ideal}}|$. Higher concentrations and stronger interactions lead to more fully formed capsids.

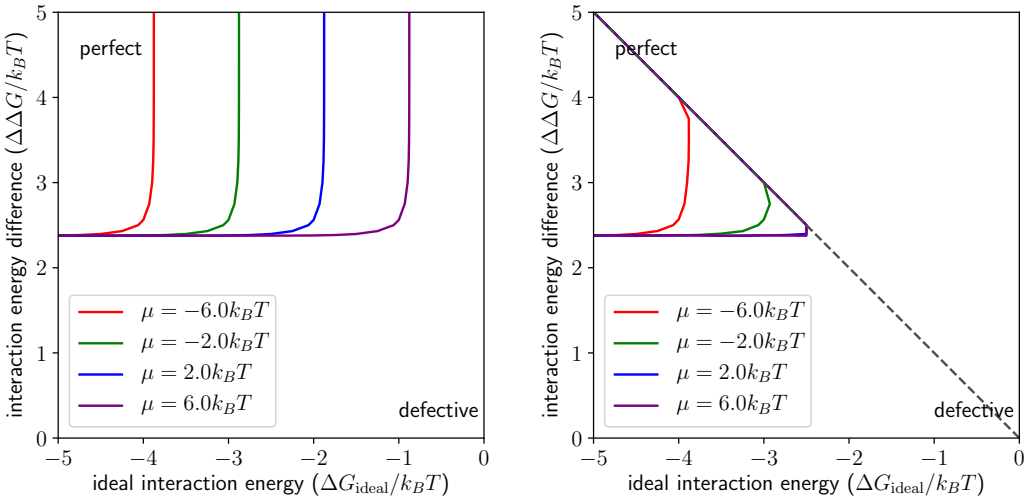
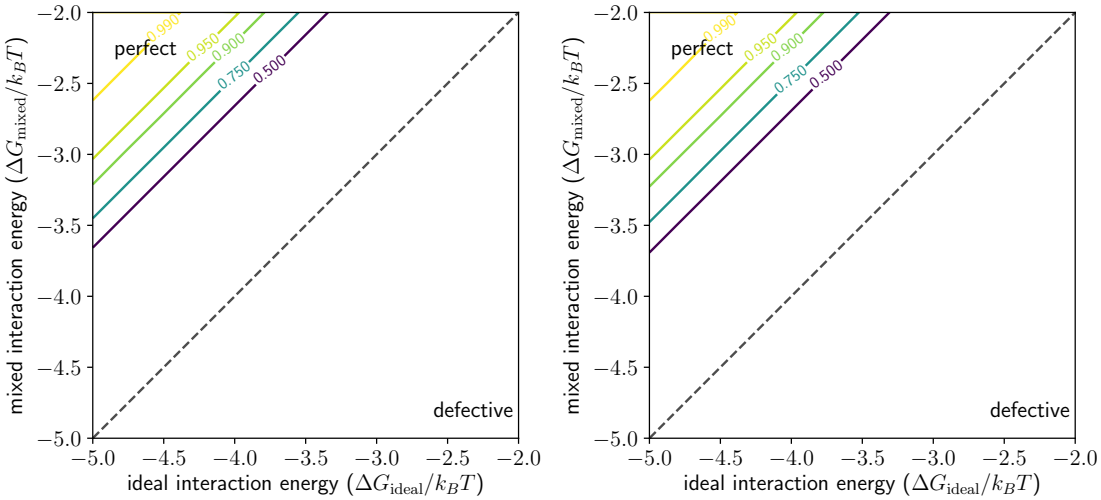
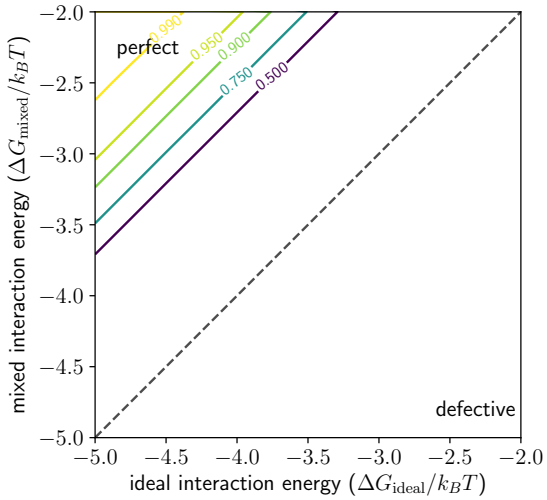


Figure S5: When are perfect capsids expected? (a) Plot of $p_{\text{perfect capsid}}$ as determined from eq. 9 and (b) $p_{\text{perfectcapsid}}$, as determined from MC simulations, both showing the position of the 0.99 contour in the $(\Delta G_{\text{ideal}}, \Delta\Delta G)$ space (chemical potential and interaction energy). No simulations were performed in the region above the gray dotted line, where $|\Delta\Delta G| \geq |\Delta G_{\text{ideal}}|$. Different curves show this contour for different values of the chemical potential μ.



(a) (b)



(c)

Figure S6: Effect of $\Delta G_{\mathrm{defect}}$ on the probability of perfect capsids. Probability of perfect capsids in the $(\Delta G_{\mathrm{ideal}}, \Delta G_{\mathrm{mixed}})$ plane as determined from MC simulations with different values of $\Delta G_{\mathrm{defect}}$. (a) for $\Delta G_{\mathrm{defect}} = \Delta G_{\mathrm{ideal}} + 0.1 k_B T$; (b) for $\Delta G_{\mathrm{defect}} = \Delta G_{\mathrm{ideal}} + 1.0 k_B T$; (c) for $\Delta G_{\mathrm{defect}} = \Delta G_{\mathrm{ideal}} + 10.0 k_B T$.

Supporting References

[1] Chandler, D. *Introduction to Modern Statistical Mechanics* (Oxford University Press, USA, 1987).

[2] Pathria, R. K. *Statistical Mechanics* (Elsevier, Amsterdam, 1996).