



**HAL**  
open science

## How a Virus Circumvents Energy Barriers to Form Symmetric Shells

Sanaz Panahandeh, Siyu Li, Laurent Marichal, Rafael Leite Rubim,  
Guillaume Tresset, Roya Zandi

► **To cite this version:**

Sanaz Panahandeh, Siyu Li, Laurent Marichal, Rafael Leite Rubim, Guillaume Tresset, et al.. How a Virus Circumvents Energy Barriers to Form Symmetric Shells. ACS Nano, 2020, 14 (3), pp.3170-3180. 10.1021/acsnano.9b08354 . hal-03035711

**HAL Id: hal-03035711**

**<https://hal.science/hal-03035711v1>**

Submitted on 4 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# How a virus circumvents energy barriers to form symmetric shells

Sanaz Panahandeh,<sup>†</sup> Siyu Li,<sup>†</sup> Laurent Marichal,<sup>‡</sup> Rafael Leite Rubim,<sup>‡</sup> Guillaume Tresset,<sup>‡</sup> and Roya Zandi\*,<sup>†</sup>

<sup>†</sup>*Department of Physics and Astronomy, University of California, Riverside, California 92521, USA*

<sup>‡</sup>*Laboratoire de Physique des Solides, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay Cedex, France*

E-mail: spana002@ucr.edu

## Abstract

Previous self-assembly experiments on a model icosahedral plant virus have shown that, under physiological conditions, capsid proteins initially binds to the genome through an en masse mechanism and form nucleoprotein complexes in a disordered state, which raises the questions as to how virions are assembled into a highly ordered structure in the host cell. Using small-angle X-ray scattering, we find out that a disorder-order transition occurs under physiological conditions upon an increase in capsid protein concentrations. Our cryo-transmission electron microscopy reveals closed spherical shells containing *in vitro* transcribed viral RNA even at pH 7.5, in marked contrast with the previous observations. We use Monte Carlo simulations to explain this disorder-order transition and find that as the shell grows, the structures of disordered intermediates in which the distribution of pentamers does not belong to the icosahedral subgroups becomes energetically so unfavorable that the caps can easily

dissociate and reassemble overcoming the energy barriers to the formation of perfect icosahedral shells. In addition, we monitor the growth of capsids under the condition that the nucleation and growth is the dominant pathway and show that the key for the disorder-order transition in both en masse and nucleation-growth pathways lies in the strength of elastic energy compared to the other forces in the system including protein-protein interactions and the chemical potential of free subunits. Our findings explain at least in part why perfect virions with icosahedral order form under different conditions including physiological ones.

## Introduction

The process of formation of virus particles in which the protein subunits encapsidate genome (RNA or DNA) to form a stable, protective shell called the capsid, is an essential step in the viral life cycle.<sup>1</sup> The capsid proteins of many small single-stranded (ss)RNA viruses spontaneously package their wild type (wt) and other negatively charged polyelectrolytes, a process basically driven by the electrostatic interaction between positively charged protein subunits and negatively charged RNA.<sup>2-6</sup> Understanding the phenomena of formation of viral particles is of great interest due to their potential applications in nanomedicine and biomaterials. Capsids can be employed as nanocontainers, biosensors, nanoreactors and drug or gene delivery vehicles to name a few.<sup>5,7-10</sup>

Regardless of the virion size and assembly procedures, most spherical viruses adopt structures with icosahedral symmetry.<sup>11,12</sup> The total number of proteins in an icosahedral capsid is equal to  $60T$  where  $T$  is the triangulation number assuming only certain integers ( $T=1,3,4,7..$ )<sup>13,14</sup> Independent of the number of proteins, there are always 12 pentamers in a viral shell sitting on the vertices of an icosahedron to preserve the symmetry of the capsid.

How exactly capsid proteins (CPs) assemble to assume a specific size and symmetry have been investigated for over half a century now.<sup>15,16</sup> Since the self-assembly of virus particles involves a wide range of thermodynamics parameters, different time scales and

an extraordinary number of possible pathways, the kinetics of assembly has been remained elusive, linked to Levinthal’s paradox for protein folding.<sup>17,18</sup> The role of genome on the assembly pathways and structure of capsid is even more intriguing.<sup>3,19–24</sup> The kinetics of virus growth in the presence of RNA is at least 3 orders of magnitude faster than that of empty capsid assembly, indicating that the mechanism of assembly of CPs around RNA might be quite different. Some questions then naturally arise: What is the role of RNA in the assembly process and by what means then, does RNA preserve assembly accuracy at fast assembly speed?

Two different mechanisms for the role of genome have been proposed: (i) en masse assembly, and (ii) nucleation and growth.<sup>25–27</sup> Several years ago, McPherson suggested the “en masse” model in which the nucleic acid attracts CPs in solution to its surface through long-range electrostatic interactions. Note that the assembly interfaces in many CPs are principally short-ranged hydrophobic in character, while there is a strong electrostatic, non-specific long-ranged interaction between RNA and CPs. To this end, the positively charged domains of CPs associate with the negatively charged RNA quite fast and form an amorphous complex. Hydrophobic interfaces then start to associate, which leads to the assembly of a perfect icosahedral shell. Based on the en masse mechanism, the assembly pathways correspond to situations in which intermediates are predominantly disordered.

More recently Chevreuil et al.<sup>28</sup> studied en masse assembly by carrying out time-resolved small-angle X-ray scattering experiments on cowpea chlorotic mottle virus (CCMV), a  $T = 3$  single-stranded RNA plant virus. They found that at neutral pH, a considerable number of CPs were rapidly ( $\sim 28$  ms) adsorbed to the genome, which more slowly ( $\sim 48$  s) self-organized into compact but amorphous nucleoprotein complexes (NPC) (see also<sup>29</sup>). By lowering the pH, they observed a disorder-order transition as the protein-protein interaction became strong enough to close up the capsid and to overcome the high energy barrier ( $\sim 20k_B T$  with  $k_B$  the Boltzmann constant and  $T$  the temperature) separating NPCs from virions. It is important to note that a marked difference between *in vitro* and *in vivo* assemblies is that

capsid proteins form ordered icosahedral structure at physiological pH (*in vivo* conditions) but till now ordered structures are observed only at acidic pH in the *in vitro* self-assembly studies.<sup>30-33</sup>

While because of the electrostatic interaction between the genome and CPs, the en masse assembly is expected to be the dominant assembly pathway, a set of more recent experiments, however, point to the nucleation-elongation mechanism. Garmann et al. have employed interferometric scattering microscopy to measure the assembly kinetics of single MS2 virus particles around MS2 RNA strands tethered to the surface of a coverslip.<sup>32,34</sup> Comparisons of individual assembly pathways indicated that most trajectories exhibit a sigmoidal time-dependence with plateau scattering intensity values similar to those acquired from complete particles. According to their experimental data, each trajectory is characterized by a lag time after which the assembly takes off, a signature of the nucleation-elongation mechanism. The measurements of Garmann *et al.* reveal that the lag time is not due to the diffusion of the CPs but it is directly related to the nucleation-and-growth mechanism.<sup>34</sup>

While the experiments of Chevreuil et al. point to the en masse assembly mechanism, it is hard to explain how a transition from a disordered amorphous RNA-CPs complex to a highly ordered icosahedral capsid can take place where the location of pentamers is very precise and CP-CP interaction is around  $7 k_B T$ . Further, we emphasize that the assembly through nucleation-and-growth mechanism is not bound to follow an ordered pathway at all the time *i.e.*, a disordered-ordered transition might also occur as the shell grows. The molecular dynamic simulations of Elrad and Hagan have shed light on the assembly pathways of virus particles in the presence of genome,<sup>26,35</sup> however the simulations were designed such that only  $T = 1$  structures could form while the kinetic pathways to the formation of slightly larger viruses such as  $T = 3, 4$  are still elusive.

In this paper, we study the kinetic pathways of virus assembly using MC simulations and specifically elucidate the role of elastic energy in the disordered-ordered transition in  $T = 3$  viruses formed through either en masse or nucleation-growth mechanism. We in particular

focus on the impact of protein concentration on the disordered-ordered transition. Through MC simulations and a set of new experiments employing small-angle X-ray scattering and cryo-transmission electron microscopy, we explore the role of protein concentrations in the formation of  $T = 3$  structures of CCMV particles. Our findings shed light on the *in vivo* assembly of icosahedral structures and can explain to some extent why virion forms at neutral pH *in vivo* while the acidic pH was so far required in the *in vitro* self-assembly studies of  $T = 3$  structures.

The paper is organized as follows. In the next section we introduce the simulation and experimental methods. In Sect. III we present the results of numerical simulations indicating the importance of elastic energy in the disorder-order transition. We also study the impact of protein concentration on the assembly pathway and compare it with our SAX experiments. Finally, we present our conclusion and summarize our findings.

## Method

### Simulations

#### Viral shell

To study the growth of viral shells, we use a coarse-grained model in which we focus on the fact that viral capsids are formed basically from hexamers and pentamers. The simplest building block that have been extensively used to model viruses and other protein nanocages are trimers.<sup>7,36,37</sup> The model allows us to discretize the growing elastic shell with triangular subunits and associate a monomer with each triangle vertex.<sup>38</sup>

The total energy of a growing shell is the sum of the attractive interaction between

subunits  $E_{hp}$ , and the stretching  $E_s$  and bending  $E_b$  energies as follows,

$$E_s = \sum_{l_i} \frac{1}{2} k_s (b_{l_i} - b_0)^2 \quad (1)$$

$$E_b = \sum_{\langle t_i, t_j \rangle} k_b [1 - \cos(\theta_{\langle t_i, t_j \rangle} - \theta_0)] \quad (2)$$

$$E_{hp} = \sum_{v_i} \epsilon_{hp} [nt_{v_i} \cdot (nt_{v_i} - 1)], \quad (3)$$

where the stretching energy sums over all edges  $l_i$  with  $b_0$  the equilibrium length and  $k_s$  the stretching modulus. The bending energy involves the dihedral angle between connected trimers  $\langle t_i, t_j \rangle$ , with  $k_b$  the bending rigidity and  $\theta_0 = 2 \arcsin(\frac{b_0}{\sqrt{12R_0^2 - 3b_0^2}})$  the preferred dihedral angle related to the spontaneous radius  $R_0$ .<sup>39</sup> The attractive energy due to the hydrophobic interaction sums over all vertices  $v_i$  with  $\epsilon_{hp}$  the strength of monomer-monomer interaction. The range of hydrophobic interaction is considered to be small so that each vertex only interacts with its nearest neighbor vertices. The number of connected trimers is denoted by  $nt_{v_i}$  for any vertex  $v_i$ . We note that the interaction between subunits is due to both the hydrophobic interactions and electrostatic repulsion between subunits.<sup>40</sup> Since the resultant force is attractive for the viruses to assemble, for brevity, we call it, the hydrophobic interaction, in the rest of the paper.

Using Monte Carlo (MC) simulations, we then monitor the growth of a viral shell through reversible steps.<sup>37</sup> We consider three different MC moves: 1. Diffusion, 2. Attachment or detachment and 3. Merging or disjointing.

During the diffusion process, the number of capsid subunits remains fixed, corresponding to a canonical system. At each MC step, either a trimer or a vertex is chosen randomly, see Figs. 1 and 2. The move is accepted with the probability  $p_m = \min(1, e^{-\beta E})$  with  $\beta = 1/k_B T$  and  $E$  the change in the total energy of the system after relaxation. The diffusion of trimers in Fig. 1 could be considered as a combination of detachment and growth for trimers, the result of which should be qualitatively the same as diffusion in canonical ensemble. The diffusion stops based on the convergence of the energy of the system. The arrows in Fig. 2

indicate the reversibility of each action.

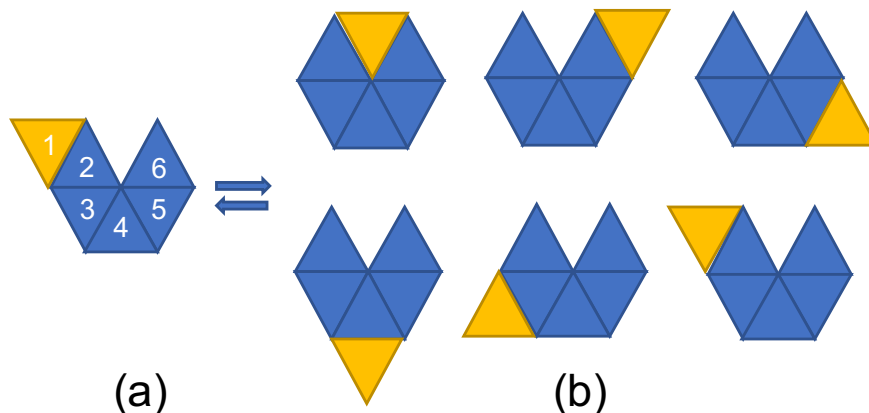


Figure 1: Schematic presentation of diffusion of a trimer around the shell edge. (a) Trimer 1 (yellow) is chosen randomly to diffuse around the edge. In this case, only trimers 1 and 6 are allowed to move during the diffusion mode. Trimers 2-5 are connected such that a few bonds need to be broken before they can diffuse. (b) Possible locations for trimer 1 to diffuse.

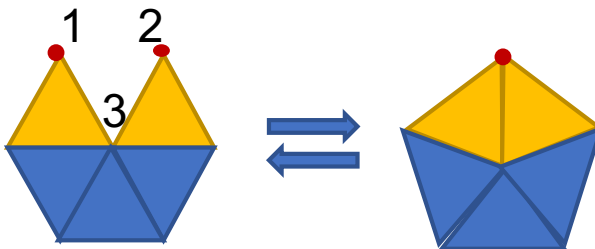


Figure 2: Schematic of vertex diffusion: Vertices 1 and 2 merge to form a pentamer. The action is reversible in that both merging (left to right arrow) and disjointing (right to left arrow) are possible in the Monte-Carlo moves.

After the diffusion step, a protein will be added or removed with the probability  $p_g$  and  $p_d$ , respectively. The probability of growth follows the detailed balance of the grand canonical ensemble with  $p_g = \min(1, \frac{T_g}{T_d} e^{\beta\mu - \beta\Delta E})$  and the probability of removal is  $p_d = \min(1, \frac{T_d}{T_g} e^{-\beta\mu - \beta\Delta E})$ .<sup>41</sup> The possible growth and detachment sites are denoted by  $T_g$  and  $T_d$  respectively.

The above MC moves can be simply illustrated by the following equations,

$$C \rightleftharpoons C' \quad (4)$$

$$C + T \rightleftharpoons TC \quad (5)$$



where Eq. 4 involves the trimer diffusion, vertex moves, two trimers merging or disjointing while the total number of triangles are kept constant. Equation 5 corresponds to the shell growth or dissociation where a trimer joins to or detaches from the capsid.

We note that in most experiments monitoring the assembly of proteins around the genome till now, RNA condensates into its final size much faster than the other time scales involved in the process.<sup>28,32,42</sup> Thus we replace the genome with a spherical core as the goal of this paper is basically on how the shell overcomes many energy barriers to form symmetric structures and not on the genome configurations. Replacing genome by a spherical core will decrease many degrees of freedoms in the system allowing us to focus carefully on the energetics involved in the assembly of viral shells. Thus, we consider a spherical core interacting through Lennard-Jones (LJ) with the capsid proteins  $E_{lj} = \sum_{vi} \epsilon_{lj} [(\frac{\sigma_{lj}}{r_{<vi,g>}})^{12} - 2(\frac{\sigma_{lj}}{r_{<vi,g>}})^6]$  where  $\epsilon_{lj}$  is the potential strength and  $r_{<vi,g>}$  is the distance between the core center and the triangle vertex  $v_i$ .

## Experiments

### Sample preparation

CCMV virions were purified from infected blackeye cowpea leaves (*Vigna unguiculata*) after the protocol developed by Ali and Roossinck.<sup>43</sup> Briefly, infected leaves were blended with 0.15 M sodium acetate pH 4.8 and ice-cold chloroform prior to precipitation of the virions with 8% (v.w<sup>-1</sup>) poly(ethylene glycol). The virions were further purified through a 20% (v.w<sup>-1</sup>) sucrose cushion and stored at -80°C until use. Proteins were purified from virions via RNA precipitation with 0.5M CaCl<sub>2</sub> and stored at 4°C until use. Full RNA genome was extracted from virions by using TRIzol reagent (Life Technologies, France) according to the protocol recommended by the manufacturer. CCMV RNA 2 was transcribed *in vitro* with a linearized plasmid coding for the RNA and containing a T7 promoter. All the RNA were stored in RNase-free water at -80 °C. Nucleoprotein complexes (NPCs) were assembled by mixing capsid proteins and RNA in the desired ratio and in some cases, dialyzed against the

final buffer. The detailed protocols for sample preparation are available in the Supporting Information (SI).

### **Small-angle X-ray scattering (SAXS)**

Scattering patterns were recorded at the ID02 and SWING beamlines of the European Synchrotron Radiation Facility (Grenoble, France) and the SOLEIL synchrotron (Saint-Aubin, France), respectively. Between 10 and 100 frames were collected for each sample with a beam exposure time set to 10 ms. *Ab initio* shape reconstructions were performed by using the programs DAMMIF<sup>44</sup> and GNOM.<sup>45</sup> For each shape reconstruction, 20 models were averaged then filtered, and the final structures were rendered with Chimera.<sup>46</sup> Sphericity indexes were computed through a principal component analysis. Let  $a$ ,  $b$  and  $c$  be the standard deviations in descending order along the principal axes of a reconstructed structure, the sphericity index is defined as  $[c^2/(ab)]^{1/3}$ . The sphericity index is positive and tends to one for a sphere.

### **Cryo-transmission electron microscopy**

The sample was frozen on a holey carbon grid (Quantifoil R2/2) by using a FEI Vitrobot and observed with a JEOL JEM-2010 electron microscope equipped with a 200-kV field emission gun. Images were collected with a Gatan Ultrascan 4K CCD camera with a  $\times 50,000$ -magnification using a minimal dose system. Defocus was set to  $2.5 \mu\text{m}$ .

## **Result and Discussion**

As noted in the introduction two different mechanisms are proposed for the role of genome in the assembly of viral shells: (1) Nucleation and growth, and (2) en masse assembly. We will decipher each in the following sections. Because of the extraordinary number of degrees of freedom involved in the assembly process, we focus only on understanding a number of

recent experiments in which genome has assumed different roles. We study the experiments of Garmann et al.<sup>34</sup> to elucidate the process of nucleation and growth in the presence of genome and then investigate the experiments of Chevreuil et al. to explore how genome promotes the assembly through en masse or multi-nucleation mechanism.<sup>28</sup> In order to obtain a better understanding of the disordered-ordered transition, we perform a new set of x-ray scattering experiments to study the role of protein concentration in the disorder-order transition and compare them with the results of our simulations. In what follows we first present the results of our simulations for the single nucleation assembly and then show our experimental data and simulations corresponding to the multi-nucleation or en masse mechanism.

## Nucleation and growth

We first investigate the assembly of empty capsids where the nucleation and growth is the dominant mechanism and then present our results on the single nucleation mechanism in the presence of genome.

### Empty capsids

Following Eqs. 1-2, the total energy of an empty capsid can be written as

$$E_{tot} = E_{el} + E_{hp} - N_T \mu = (\epsilon_{el} + \epsilon - \mu)N_T, \quad (6)$$

where  $N_T$  is the total number of trimers,  $E_{el} = E_s + E_b$  is the total elastic energy of the cap,  $E_{hp}$  is the protein-protein interaction (basically hydrophobic) and  $\mu$  is connected to the chemical potential of free proteins in solution. For a dilute system,  $\mu = k_B T \log \frac{\rho}{\rho_0}$  with  $\rho$  the density of free proteins in solution and  $\rho_0$  the reference density. The average elastic energy and attractive protein-protein interaction per subunit are denoted by  $\epsilon_{el}$  and  $\epsilon$ , respectively.

Figure 3a.III illustrates the snapshots of simulations for the growth of an empty shell through the reversible pathway described in the previous section. As the cap grows, at

some point, the formation of pentamers becomes energetically unavoidable because of the spherical geometry of the shell. The first pentamer always forms in the vicinity of the cap edge, and then the shell grows around the pentamer such that at the end, the pentamer is in the middle of the cap. The following pentamers appear as illustrated in the figure and the shell grows around them. Quite interestingly the position of pentamers follows very well the prediction of continuum elasticity theory where the ground state energy of a spherical cap was calculated to obtain the optimal position of disclinations,<sup>47,48</sup> see Fig. S1. This confirms that the system is completely reversible, equilibrated, and is able to find the minimum energy structures along the pathway.

The fact that most spherical viruses adopt structures with icosahedral symmetry reveals the important role of elasticity in the energetics of viral shells.<sup>38,48,49</sup> Nevertheless, it has remained a mystery how an error-free shell formed out of 90 dimers or 60 trimers grows with perfect icosahedral symmetry under many different *in vitro* assembly conditions. One expects that when a considerable part of a shell is formed, the pentamers formed well inside the cap far from the edge will become more or less frozen at their locations. Thus if a pentamer is formed in a “wrong” position, the icosahedral symmetry should permanently be broken. However, the reversible simulations show that as a shell grows, the elastic energy could become very strong to easily break the bonds for a wide range of hydrophobic interactions and fix the positions of pentamers formed in the locations not consistent with icosahedral subgroup symmetries.

Figure 4 shows the snapshots of simulations of a cap in which two pentamers are formed next to each other at the beginning of the assembly. The reversible growth allows the bonds to break and a pentamer to change to a hexamer as the assembled shell grows to have about 25 trimers. Note that the dynamic of system specifically depends on the stretching modulus of proteins,  $k_s$ . Lower stretching rigidity slows down the pentamer-hexamer transition, emphasizing the role of elasticity in the kinetic pathway of assembly. Figure S2 shows that as  $k_s$  decreases from  $600k_B T$  to  $200k_B T$ , it takes much longer for the system to “fix” the

position of the pentamer.

Figure 4 clearly illustrates that the elastic energy of the shell drops (indicated with an arrow) when the bonds of a “wrong” pentamer are broken. During this time the hydrophobic energy increases, which is the penalty for breaking the bonds, resulting into an energy barrier to the formation of icosahedral shells. However in this case the elastic energy of the shell is strong enough to overcome the local energy barriers and assume the symmetric structure. It is important to emphasize that the elastic energy diverges if a shell grows without pentamers.

If the elastic energy is too small or too large compared to the attractive interaction between subunits, the symmetry of the shell will also be broken. To explore the interplay of elasticity and the hydrophobic interaction between subunits in the final structure of shells, we construct a phase diagram for the spring modulus  $k_s$  ranging from  $20k_B T$  to  $1200k_B T$  and  $\epsilon_{hp} \in [1k_B T, 2.2k_B T]$  as illustrated in Fig. 3b. In region I of the phase diagram, the hydrophobic interaction is so weak that no capsid could nucleate. In region II the elastic energy is weak compared to the hydrophobic interaction, and the shell easily gets kinetically trapped and irregular structures form with total number of subunits less than 60 (a  $T = 3$  structure)(Fig. 3a). One can see that for a large region of the phase diagram (region III), the competition between elastic and hydrophobic is such that the final product is always a  $T = 3$  structure. As the elastic energy becomes stronger in region IV, even for very strong hydrophobic interactions, the shell is not be able to close but assumes a structure with cylindrical symmetry and grows “branchy” way, forming a messy structure (Fig. 3a). While  $T = 3$  constitutes the minimum energy structure for the entire phase diagram, messy structures with no symmetry can readily form in different regions as shown in Fig. 3. It is not always possible to obtain  $T = 3$  structures without using different MC techniques to avoid local traps. We calculate the energy barrier to the formation of  $T = 3$  structures in the next section to explain why in some regions the shells can easily get trapped in a local minimum structure.

At the end of this section, we emphasize that even though during the simulations one

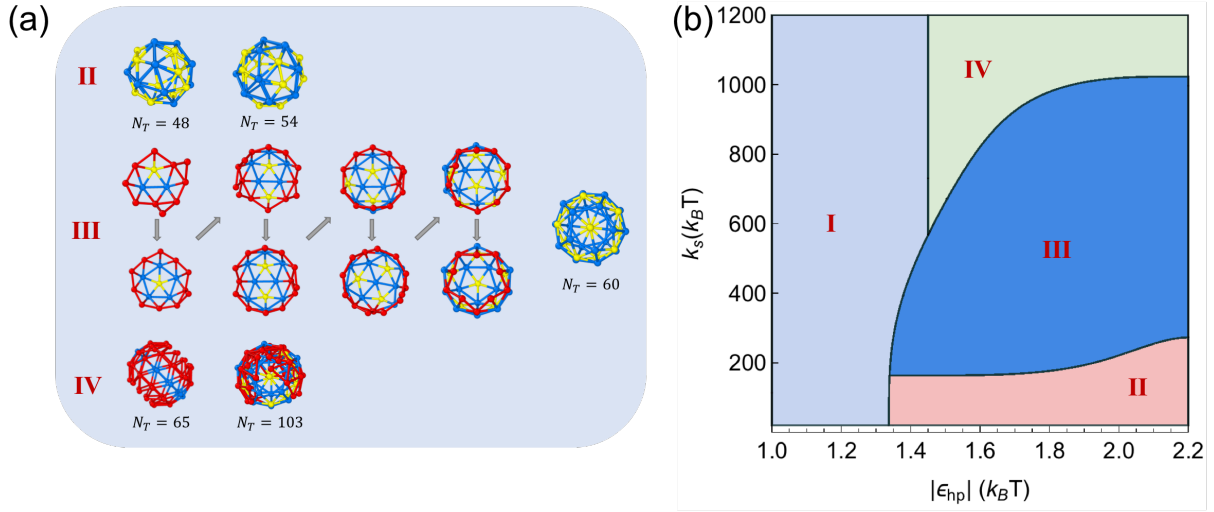


Figure 3: Phase diagram for the reversible growth model and snapshots of assembled structures as a function of stretching modulus  $k_s$  and monomer-monomer hydrophobic interaction  $\epsilon_{hp}$ . In region I no shell forms due to lack of nucleation; in region II capsids close to form irregular shapes with lower symmetries and total trimer number  $N_T$  is less than 60; in region III, only  $T = 3$  structures form. Snapshots of a representative pathway is illustrated on the panel (a) where pentamers first form at the boundary and then the shell continues to grow around them. Images in panel (a) are made with OVITO.<sup>50</sup> The pathway is consistent and matches very well the ground state calculation corresponding to the locations of pentamers as a shell grows, see SI Fig. S1. In region IV the elasticity is so high that the structures cannot close but form messy shells. Other parameters of simulations are spontaneous radius  $R_0 = 1.5$ , bending rigidity  $k_b = 200k_B T$  and chemical potential  $\mu = -14.6k_B T$ .

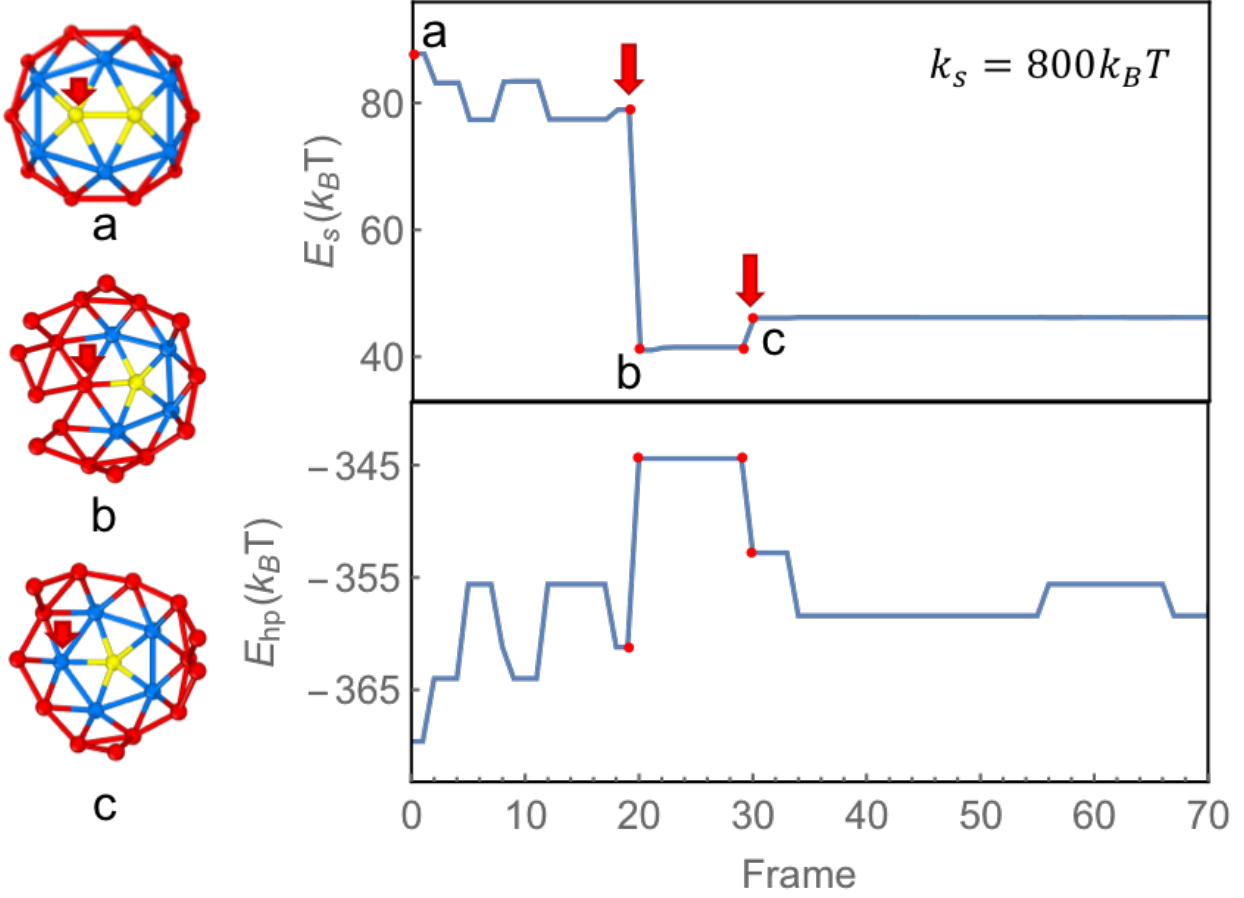


Figure 4: Snapshots of simulations for a partially formed capsid with 24 subunits. The right column shows the elastic and hydrophobic energies vs. Frame (number of MC moves) or time. (a) Two pentamers are formed in the vicinity of each other. (b) Around 20 MC steps due to high elastic energy (marked with a red arrow) a bond breaks to decrease the elastic energy while increasing the hydrophobic energy (point b in the energy plots). (c) A hexamer forms and slightly increases the elastic energy while decreasing the hydrophobic energy. The protein subunits stretching rigidity is  $k_s = 800k_B T$  and bending rigidity  $k_b = 200k_B T$ . Other parameter used are hydrophobic interaction  $\epsilon_{hp} = -1.4k_B T$  and chemical potential  $\mu = -14.6k_B T$ .

protein was added at a time, our conclusion remains the same if several subunits join the growing shell simultaneously. The strength of elastic energy as the shell assembles makes the process reversible and the results robust, independent of the number of subunits joining the growing shell .

## classical nucleation and energy barrier

To understand why closed structures without any particular symmetry appear in the phase diagram of Fig. 3 where  $T = 3$  structure has the lowest energy, we calculate the energy barriers for the formation of  $T = 3$  in different regions of the diagram.

For a complete shell, the total hydrophobic interaction is  $E_{hp} = 15(Q - 4)\epsilon_{hp}$  with  $Q$  the total number of subunits and  $\epsilon_{hp}$  the monomer-monomer interaction as noted before. For a growing shell or a cap, the hydrophobic interaction, Eq. 3, becomes

$$\begin{aligned}
 E_{hp} &= \left( 15N_T - 60\frac{N_T}{Q} - 9\sqrt{\sqrt{3}\pi}\sqrt{\frac{N_T}{Q}(Q - N_T)} \right) \epsilon_{hp} \\
 &= 15\epsilon_{hp}\frac{Q - 4}{Q}N_T - 15\epsilon_{hp}\sqrt{\alpha}\sqrt{\frac{N_T}{Q}(Q - N_T)} \\
 &= \epsilon N_T + \epsilon_l,
 \end{aligned} \tag{7}$$

where  $\alpha = \frac{9\sqrt{3}\pi}{25}$  (see Section. III in SI). The average hydrophobic interaction per subunit is  $\epsilon = 15\epsilon_{hp}\frac{Q-4}{Q}$ , which converges to  $15\epsilon_{hp}$  as  $Q \rightarrow +\infty$ .

Combining Eq. 6 and Eq. 7, we obtain the free energy of the growing capsid as

$$\begin{aligned}
 \frac{E_{tot}}{|\epsilon|} &= \frac{\epsilon_{el} + \epsilon - \mu}{|\epsilon|}N_T + \sqrt{\alpha}\frac{Q}{Q - 4}\sqrt{\frac{N_T}{Q}(Q - N_T)} \\
 &= AN_T + a\sqrt{N_T(Q - N_T)},
 \end{aligned} \tag{8}$$

where  $a = \frac{\sqrt{\alpha Q}}{Q-4}$  is a geometric factor and  $A = \frac{\epsilon_{el} + \epsilon - \mu}{|\epsilon|}$  measures the difference between the



chemical potentials of free subunits in solution and in full capsids. To make it dimensionless, we have divided it by the hydrophobic energy per subunit  $\epsilon$ . For a  $T = 3$  structure

$$A = \frac{\epsilon_{el} + 14\epsilon_{hp} - \mu}{|14\epsilon_{hp}|}. \quad (9)$$

The height of energy barrier and the nucleation size can then be calculated through the maximum of the free energy as follows,

$$\begin{aligned} N_T^* &= \frac{1}{2}Q \left( 1 + \frac{A}{\sqrt{A^2 + a^2}} \right) \\ E_{tot}^* &= \frac{1}{2}|\epsilon|Q \left( A + \sqrt{A^2 + a^2} \right) \end{aligned} \quad (10)$$

Or

$$\begin{aligned} N_T^* &= \frac{1}{2}Q \left( 1 + \frac{\Gamma}{\sqrt{\Gamma^2 + 1}} \right) \\ E_{tot}^* &= \frac{1}{2}|\epsilon|aQ \left( \Gamma + \sqrt{\Gamma^2 + 1} \right), \end{aligned} \quad (11)$$

where  $\Gamma = \frac{A}{a}$  is the supersaturation.

Figure 5 shows the total free energy (Eq. 8) vs. number of subunits for  $A \in (-0.5, 0.5)$  (Eq. 9) at three different values of the monomer-monomer hydrophobic interaction,  $\epsilon_{hp} = -1.0, -1.4, -1.8k_B T$ . The figure reveals the impact of different thermodynamics parameters on the height of energy barriers. As illustrated in the figure, for a fixed  $\mu = -14.6k_B T$  and a given hydrophobic interaction  $\epsilon_{hp}$ , the energy barrier increases when the elastic energy per subunit  $\epsilon_{el}$  increases, see Eq. 9.

The results shown in Fig. 5 is consistent with the phase diagram presented in Fig. 3 in that for  $\mu = -14.6k_B T$  if factor  $A$  is large because of the strong elastic energy and the small hydrophobic  $\epsilon_{hp} = -1k_B T$ , then the energy barrier is too high for the formation of capsids and no structure nucleates. In fact, for  $\epsilon_{hp} = -1.0k_B T$  regardless of stretching modulus, no capsid nucleates unless the chemical potential is at least  $\mu = -10k_B T$ . On the other hand,

when hydrophobic interaction is too strong compared to the elastic energy,  $A$  is small and the nucleation size is less than five subunits resulting in the formation of kinetically trapped smaller structures than  $T = 3$ , corresponding to region II in the phase diagram.

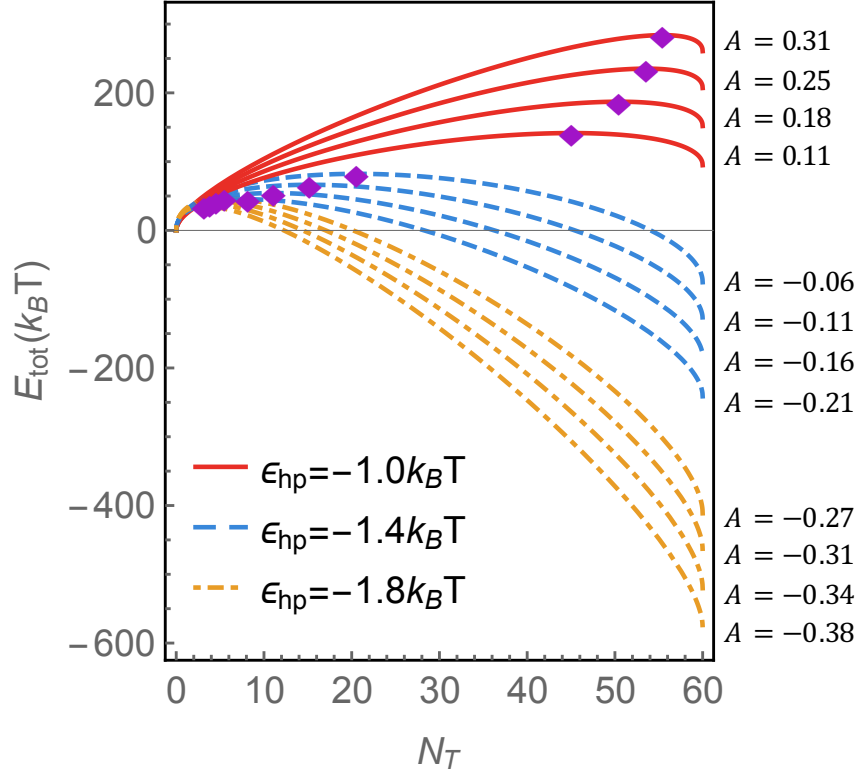


Figure 5: Total energy  $E_{tot}$  as a function of subunits. The energy barriers are indicated with diamonds. The value of coefficient  $A$  (Eq. 9) is indicated next to each curve. The stretching moduli used are  $k_s = 800, 600, 400, 200 k_B T$  (from top to bottom for each color). The chemical potential is  $\mu = -14.6 k_B T$  and the hydrophobic interaction  $\epsilon_{hp} = -1.0 k_B T$  (solid),  $-1.4 k_B T$  (dashed) and  $-1.8 k_B T$  (dotted dashed).

In the next section we present the growth of capsids in the presence of genome through nucleation and growth mechanism.

## Assembly in the presence of genome

The self-assembly experiments of Garmann et al. indicate the assembly proceeds through nucleation and growth mechanism with genome present.<sup>34</sup> The process of nucleation and growth on the surface of a gold nanoparticle or condensed genome is very similar to that of

an empty capsid. In the former case, we start the simulations assuming that one trimer is sitting on the core interacting attractively with it.

Mimicking the experiments of Garmann et al. we assume that the concentration of proteins is such that the process is not diffusion limited, *i.e.* there are enough proteins in the vicinity of genomes to form capsids. More specifically, we consider that the subunits diffuse from the solution to the core but the duration of attachment of a single protein to the core is smaller than the nucleation time. Thus the delay in the assembly is just because of nucleation. To this end, we only focus on the assembly of proteins on the core and do not study the diffusion of proteins to the core. The total energy of the complexes of core and proteins can be written as,

$$E_{tot} = E_{lj} + E_{el} + E_{hp} - N_T \mu = (-\epsilon_{lj} + \epsilon_{el} + \epsilon - \mu)N_T, \quad (12)$$

with  $\epsilon_{lj}$  the depth of the LJ potential representing the core-protein interaction on the core surface. For simplicity we always set  $R_{core} = R_0$ .

The results of simulations in the presence of core are presented in Fig. 8. For these sets of simulations we have used experiments of Garmann et al. to calibrate our MC steps to second.<sup>34</sup> During the simulations, we observe that after about four or five subunits are assembled, the growth becomes very fast. As shown in the figure, it takes about 70s on average for  $\mu = -10.5k_B T$  (Orange lines) the capsid to nucleate, which is followed by a rapid growth phase of 33s. As the protein concentration goes higher ( $\mu = -10k_B T$ , blue lines), the nucleation time becomes shorter, about 24s, and the rapid growth phase becomes about 25s during which the capsid assembles to almost its final size. Note that in these cases, in the absence of genome, no capsids could form. Changing  $\mu$  from  $-10.5k_B T$  to  $-10k_B T$ , we increase the protein concentration by 64% and observe that the nucleation time decreases about 65%. This is consistent at least qualitatively with the results of Gramann et al.,<sup>34</sup> which according to their experiments, as the protein concentration changes from  $1.5\mu\text{M}$  to

2 $\mu$ M, the nucleation time decreases about from about 160s to about 95s. The results of our simulations also shows that the growth time decreases with increasing protein concentration, but with much slower rate than the nucleation time as seen in the experiments.

If we increase the protein concentration even further to  $\mu = -8k_B T$ , the chance of formation of another nucleus will be high and the process might become multi-nucleus or en masse assembly, as will be explained in the next section. All these simulations show that the interaction of capsid proteins with genome facilitates the formation of capsids, *i.e.* lowers the energy barrier.

The above results show that in the presence of genome, even if the protein-protein interaction is not strong enough for the formation of empty capsids and the protein-genome interaction is weak such that the rate of detachment is larger than the rate of capsid nucleation, the combined effect of both interactions can give rise to the assembly of filled capsids. Note that if the presence of genome decreases the energy barrier only by one or two  $k_B T$ , because of the associated Boltzmann factor the rate of formation of capsid will increase by a factor of three or four.

In the next section, we will study the en masse or multi-nucleation mechanism. First we present our new experiments and then the results of our simulations assuming that the protein-genome interaction is so strong that the proteins can easily get adsorbed to the genome and then nucleation occurs in different positions on the core.

## En masse assembly

### Experiments

The kinetic studies of Chevreuril et al.<sup>28</sup> performed by mixing CCMV subunits (dimers of CPs) and the full (single-stranded RNA) genome at neutral pH, showed that the assembly proceeded through an en masse rather than through a single nucleation-growth process, and the final objects were amorphous nucleoprotein complexes (NPCs). Quite intriguingly,

when genome was replaced by a flexible, linear polyelectrolyte, namely, poly(styrene sulfonic acid),<sup>51,52</sup> small but closed structures were formed, again via the en masse pathway.<sup>28</sup> Why then cannot subunits form ordered structures in the presence of genomic RNA at neutral pH, which is yet physiologically relevant?

Figure 6 illustrates the effect of an increase in the concentration of subunits and genomic RNA on the structure of NPCs at neutral pH. SAXS patterns of NPCs (Fig. 6a) made with the full CCMV genome exhibit an increasingly pronounced interference lobe in the medium  $q$ -values as the concentrations are raised. The lobe is even sharper when the full genome is replaced solely by *in vitro* transcribed RNA 2. The reconstructed shapes and their associated sphericity indexes (Fig. 6b) suggest that the structures become more ordered and gradually acquire a spherical symmetry at high concentrations. Notice that the reconstructed shape of sample **IV** displays a hollow core (see inset of Fig. 6b) like native virions.<sup>53</sup> The reconstructed shapes must be considered with caution though, as they only provide ensemble-averaged, low-resolution structures and cannot reflect the actual variability between individual objects. Note that the analysis of the SAXS patterns by a polydisperse vesicle model (see SI Fig. S4) also supports a higher degree of order at high concentrations.

In order to get a better insight into the morphology of the assembled objects, we perform cryo-transmission electron microscopy on a highly concentrated sample of NPCs made with *in vitro* transcribed RNA 2 because the SAXS patterns reveal objects with the highest sphericity index (Fig. 6b). Quite consistently, the imaged objects (Fig. 7) are mostly ordered and morphologically similar to native virions in these conditions, with a diameter around 30 nm. Some objects are still slightly disordered or aggregated. No empty spherical capsids are found but we can see a few hollow nanotubes (see SI Fig. S5) in agreement with the phase diagram of CCMV CPs at neutral pH.<sup>54</sup>

In summary, our experimental studies show that a disorder-order transition occurs within NPCs upon an increase in subunits and genomic RNA concentrations, and that closed spherical shells packaging viral RNA can be assembled at neutral pH. These findings motivate

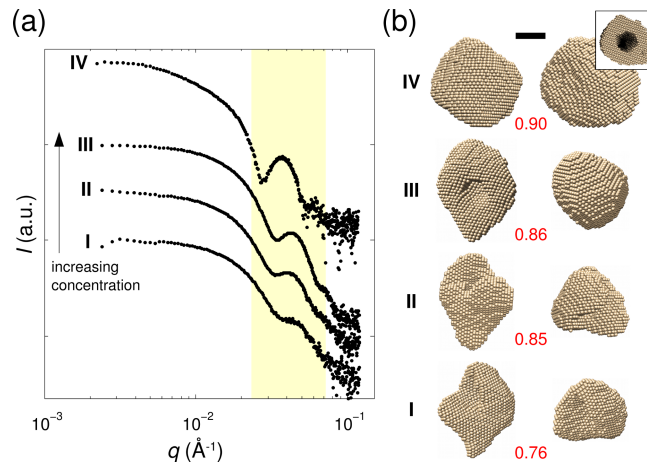


Figure 6: Equilibrium structure of NPCs at neutral pH and increasing concentrations in subunits and genomic RNA. (a) SAXS patterns obtained at a fixed subunit-to-RNA mass ratio of 6. Samples **I**, **II** and **III** are made with the full CCMV genome at subunit concentrations of 0.5, 1.0 and 2.1 g.L<sup>-1</sup>, respectively, whereas sample **IV** contains only *in vitro* transcribed CCMV RNA 2 for a subunit concentration of 2.1 g.L<sup>-1</sup>. The yellow area highlights the growth of an interference lobe in the curves due to an increasingly well-defined length scale in the scattering objects. The scattering curves are shifted for clarity. (b) *Ab initio* shape reconstructions carried out with the scattering curves in (a). The numbers in red are the sphericity indexes calculated for the corresponding structures. The inset is a cross-sectional view of the structure obtained for the sample **IV**. Scale bar is 10 nm.

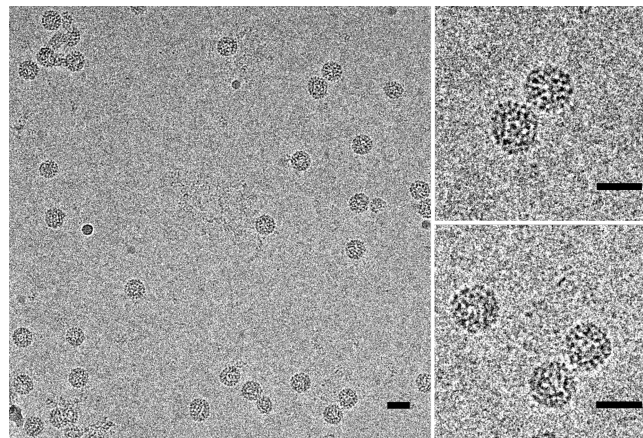


Figure 7: Cryo-transmission electron microscopy images of NPCs prepared at neutral pH. CCMV subunits are at a concentration of 3.9 g.L<sup>-1</sup> and *in vitro* transcribed CCMV RNA 2 is mixed at a subunit-to-RNA mass ratio of 6. Scale bars are 30 nm.

us now to investigate theoretically the ability of the chemical potential to lower the energy barrier between NPCs and assembled virions.

## Simulations

The snapshots of the growth of a capsid through en masse is illustrated in Fig. 9. Right at the beginning of simulations many proteins get adsorbed to the core and a messy structure forms (Fig. 9a), similar to the NPCs observed in previous experiments<sup>28,55</sup> at pH 7.5 and the new ones presented in the last section. Special care is taken that there is no overlap between subunits. We note that as for the simulations in which the nucleation and growth is the dominant mechanism, we do not study the diffusion of proteins into the genome in case of en masse (or multi-nucleations) assembly either. More specifically, we consider the assembly is reaction limited and thus the important time scale in the problem involves the addition of subunits to each other in order to form larger patches. Due to strong protein-genome interaction, we start simulations with about 10 to 15 subunits attached to the genome and consider that the growth proceeds basically through joining of the new subunits to other subunits already attached to the genome. Figure 9b shows a snapshot of the simulations when the interaction is weak ( $\epsilon_{hp} = -0.7 k_B T$ ), and the number of trimers is 40 with multiple nuclei already formed. We continue adding the subunits randomly to the edge of one of the existing subunits and accept the move as explained in the Method section. The structure is still disordered.

The previous self-assembly experiments<sup>28,55</sup> showed that when the pH is lowered, the attractive interaction between the protein subunits becomes stronger. Following the experimental steps, we also decrease  $\epsilon_{hp}$  from  $-0.7 k_B T$  to  $-1 k_B T$ . When there are 45 trimers on the genome, we decrease  $\epsilon_{hp}$  to  $-1 k_B T$ . Around 5000s, 54 trimers are relaxed to their minimum energy positions in Fig. 9c. The shell will form a  $T = 3$  around 14000s (Fig. 9d). The kinetics of assembly through en masse pathway is illustrated in Fig. 10 in the form of the number of subunits vs. time. The behavior of the growth curves obtained in the en

masse assembly is very similar to those of Chevreuil et al. in that upon decreasing pH, a disordered to ordered transition occurs. We have employed the results of Fig. 3a in Ref.<sup>28</sup> to calibrate the MC steps with time.

In addition to the pH change, the new experiments presented in the previous section reveal the important role of protein concentrations in the disordered to ordered transition. To explore the impact of protein concentration and protein-protein interaction and to explain the experiments presented in Fig. 6, we build a phase diagram for the formation of closed shells as a function of chemical potential and hydrophobic (attractive) interaction, see Fig. 11.

Figure 11 shows that for fixed  $k_s = 600k_B T$  and  $k_b = 200k_B T$ , at high protein concentrations and strong protein-protein attractive interaction, only  $T = 3$  capsids form (region III). Note that if we increase both protein concentration and the strength of protein interaction significantly, the structures will be stuck in a local minimum energy and aberrant particles form, not shown in the phase diagram. In contrast at low hydrophobic interactions and low protein concentration no structures nucleate (region I). In the purple region (VI) of the diagram, only messy/amorphous structures form as observed in the experiments at pH=7.5. In this regime the protein-protein interaction is not strong enough to form an ordered structure. Nevertheless, the figure shows that for a fixed hydrophobic interaction, upon increasing protein concentrations, a disordered-ordered transition can occur, see for example the region around  $\mu = -14k_B T$  and  $\epsilon_{hp} = -0.8k_B T$ . As illustrated in the figure, if for  $\epsilon_{hp} = -0.8k_B T$ , the chemical potential increases from -15 to -13  $k_B T$ , one can observe the formation of  $T = 3$  structures, consistent with the new experimental results shown in Fig. 6.

Quite interestingly, when we change the size of core to commensurate with a capsid whose diameter is 23nm consistent with the assembly experiments in the presence of PSS, we obtain a structure with the symmetry of a tennis ball (see SI Fig. S6), which has been previously observed in self-assembly studies of clathrin shells.<sup>38</sup> The phase diagram for PSS is very similar to that for RNA but the boundary lines between ordered-disordered configurations have moved, see Fig. S7 in SI. We find that capsid proteins assemble to form closed shells at



lower protein concentration that are needed for assembly with RNA. This behavior supports at least qualitatively the experimental results with PSS.<sup>28</sup>

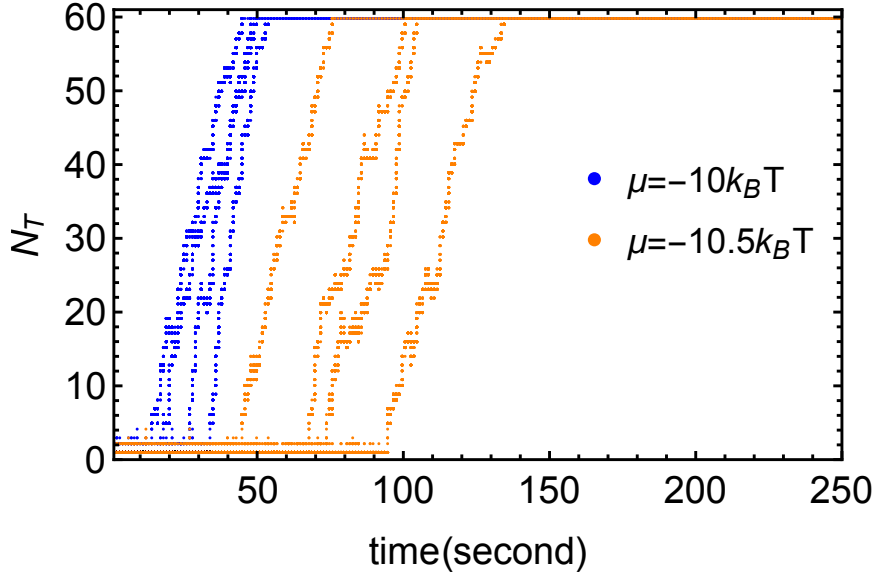


Figure 8: Several pathways for the formation of  $T = 3$  structures. The plot shows the number of subunits in the assembled caps as a function of time (second) for two different of chemical potentials  $\mu = -10 k_B T$  (Blue color) and  $\mu = -10.5 k_B T$  (Orange color). For each color four different runs are shown revealing fluctuations in the pathways of formation of  $T = 3$  icosahedral structures. The other parameters in simulations are  $k_s = 600 k_B T$ ,  $k_b = 200 k_B T$ ,  $\epsilon_{hp} = -1 k_B T$  and the genome-protein interaction  $\epsilon_{lj} = 1 k_B T$ . For these parameters, the assembly only proceeds through the nucleation and growth mechanism.

## Conclusion

Since during the assembly of viral particles, the intermediate states are transient and as such not easily accessible experimentally, the most fundamental questions about assembly pathways have remained unresolved. In this paper, we have focused on two different mechanisms observed in recent experiments: one supporting the en masse assembly and the other the nucleation and growth pathway.

Our results show that the elastic energy plays a crucial role in the assembly of error-free symmetric shells in both cases. We have found that as a shell grows, if the pentamers form in the positions that break icosahedral order, through a reversible pathway, the pentamers can

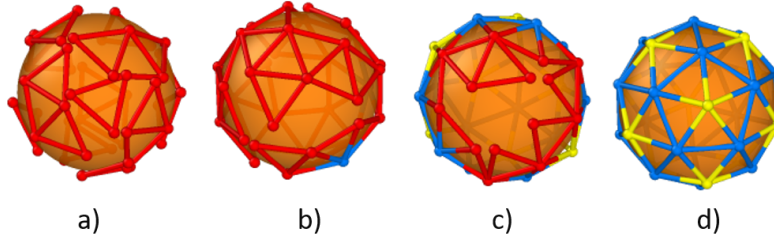


Figure 9: Snapshots of simulations in Fig. 10. The chemical potential is  $\mu = -14k_B T$ ,  $k_s=600k_B T$ ,  $k_b=200k_B T$  and  $\epsilon_{lj} = 11.2k_B T$ . (a) View of an amorphous nucleoprotein complexes after less than a second for  $\epsilon_{hp} = -0.7k_B T$ . There are around 20 subunits on the surface of the core.(b) View of disorder structure. After 96s there are 40 trimers on the genome and multiple nuclei has formed. Later when capsid has 45 trimers, we increase the hydrophobic strenght by changing  $\epsilon_{hp}$  from  $-0.7k_B T$  to  $-1k_B T$ . This is consistent with the experiments when the pH is lowered from 7.5 to 5.2. (c) View of a partially formed capsid with 54 trimers after 5000s. Subunits are relaxed to their lowest minimum energy positions.(d) The structure of a closed  $T = 3$  capsid with 60 trimers. It takes around 14000s for the capsid to complete. Red colors correspond to vertices on the edge. Blue and yellow colors show the position of hexamers and pentamers respectively. The images are made with OVITO.<sup>50</sup>

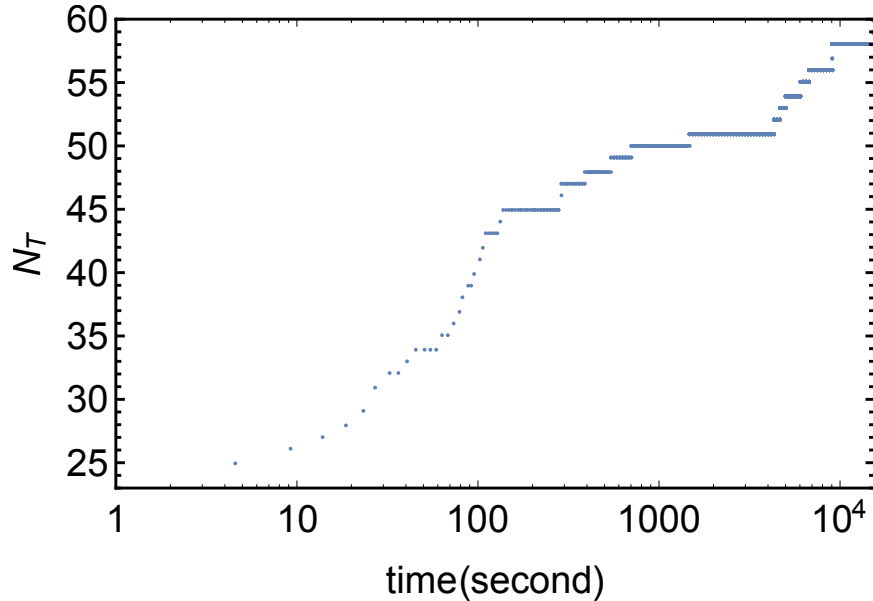


Figure 10: The number of subunits assembled around the genome vs. time (second) for multi nucleation assembly at  $\mu = -14k_B T$ ,  $k_s=600k_B T$ ,  $k_b=200k_B T$  and  $\epsilon_{lj} = 11.2k_B T$ . The number of subunits increases from 23 to around 45 trimers in the first 130s where the hydrophobic strength is weak ( $\epsilon_{hp} = -0.7k_B T$ ). At  $N_T = 45$ , Decreasing  $\epsilon_{hp}$  to  $-1k_B T$  partially formed disordered capsid starts relaxing to the final structure with  $T=3$  symmetry. The process of relaxation and reorganization to an ordered capsid takes around 14000s.

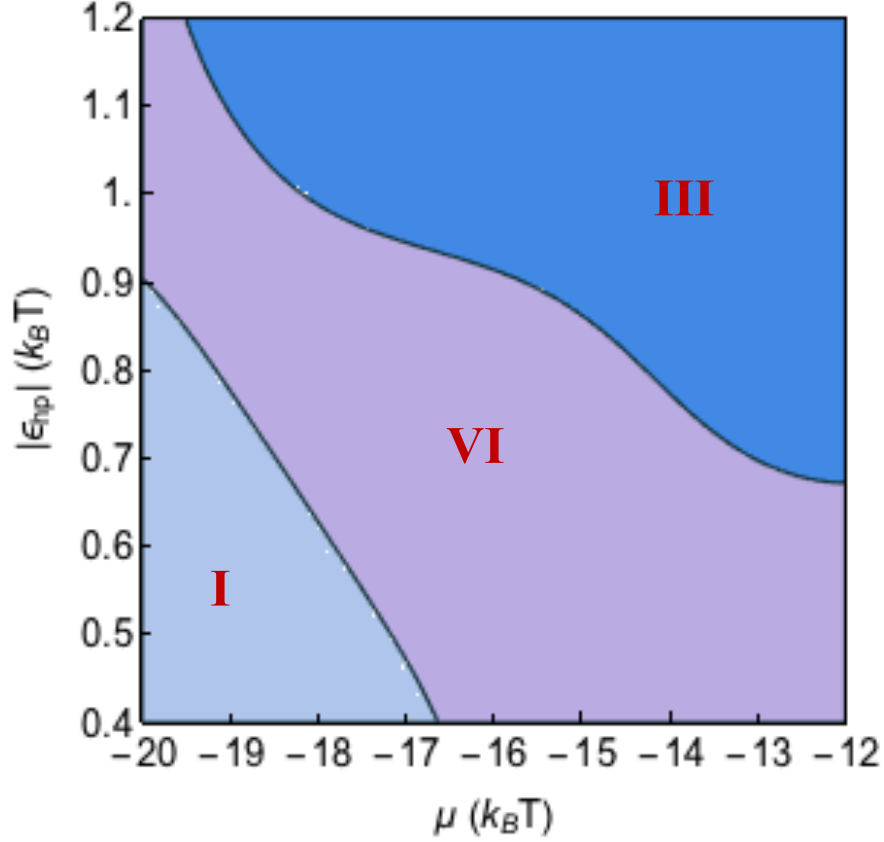


Figure 11: Phase diagram of structures obtained from en masse simulations for different values of chemical potentials and the strength of hydrophobic interactions. The strong genome-protein attractive interaction ( $\epsilon_{lj} = 11.2 k_B T$ ) makes the assembly proceed through en masse pathway. The blue shade corresponds to the region in which  $T = 3$  capsids form but the purple one represents the phase where amorphous structures assemble. No capsid nucleates in light blue region. As we move from left to right, the concentration of free subunits increases and we observe the disordered to ordered transition. The other parameters in the simulations are  $k_s = 600 k_B T$ ,  $k_b = 200 k_B T$  and the spontaneous radius  $R_0 = 1.5$ .

move to preserve the symmetry and form a  $T = 3$  shell. This is counter-intuitive as many bonds need to break for a pentamer formed far from the edge, well inside the capsid to move. However, our results show that the elastic energy becomes so strong that as many bonds as necessary can be broken to move the pentamer, see Fig. 4. Obviously, the relative strength of hydrophobic interaction and elastic energy is very important in this process, see Figs. 3 and 5. Figure 12 shows a messy shell that was initially assembled fast in the presence of high protein concentration. These sorts of shells can form either through nucleation and growth or en masse pathway. Upon decreasing the protein concentration, we find that the shell relaxes to a  $T = 3$  structure through reversible pathway. The elastic energy indeed plays a crucial role in preserving the symmetry in transferring disordered to ordered structures.

The important role of protein concentration in the assembly pathway is summarized in Fig. 13 for two different values of genome-protein interaction,  $\epsilon_{lj}$ . If the genome-protein interaction is weak  $\epsilon_{lj} = 1k_B T$ , for the low protein concentration, no shell nucleates. As the protein concentration increases, the shell grows through the nucleation-growth pathway. At higher concentrations, we will observe that proteins get adsorbed onto the spherical core and shell grows through en masse assembly into a  $T = 3$  structure. In contrast for large  $\epsilon_{lj} = 11.2k_B T$  at low protein concentration, amorphous structures form. Upon increasing protein concentration ordered structures are obtained through en masse assembly.

These results are consistent with our new SAXS measurements, which clearly show an increasing degree of order upon the increase of concentration. While even our best curves are not as regular as those obtained with native virions, in marked contrast with the previous experiments we find that some structures look spherical and fully assembled. We note that the concentration may not be high enough or other factors such as crowding effect or divalent cations may be important in the formation of icosahedral capsids as well. Nevertheless, our experimental and numerical results clearly demonstrate that the chemical potential can significantly lower the energy barrier between amorphous complexes and virions, and it must have definitively a role *in vivo* assembly as well.

Deciphering the factors that contribute to the dominant assembly pathway could enable biomedical attempts to block viral replication and infection and promote the use of capsids in bionanotechnology.

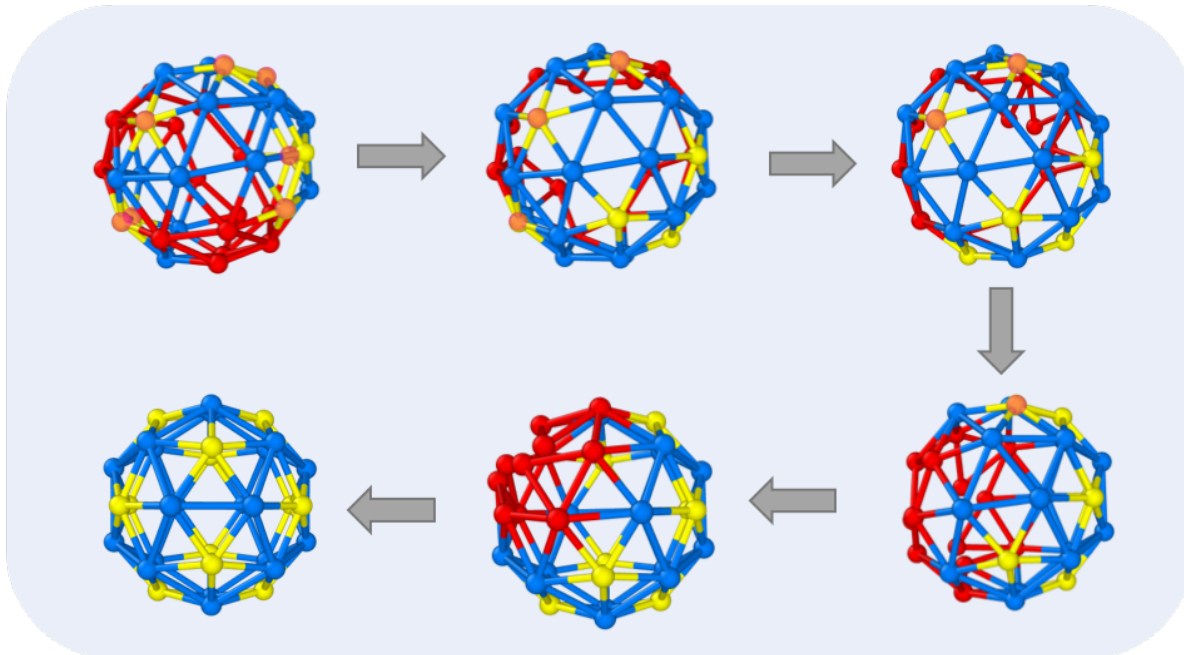


Figure 12: Relaxation of a messy shell to an icosahedral structure. At the beginning, the shell has 57 trimers with 7 pentamers formed in "wrong" positions (marked with orange color). The shell relaxes to a T=3 capsid upon a change in pH or protein concentration. The blue, yellow and red colors correspond to hexamers, pentamers and edge respectively. The parameters used are stretching rigidity  $k_s = 200k_B T$ , bending rigidity  $k_b = 200k_B T$ , hydrophobic interaction  $\epsilon_{hp} = 1.0k_B T$  and chemical potential  $\mu = -14k_B T$ . The images are made with OVITO.<sup>50</sup>

## Acknowledgement

This work was supported by the NSF through Awards DMR-1719550. The plasmid coding for the CCMV RNA 2 was kindly provided by Christian Beren from William Gelbart's group at the University of California Los Angeles. G.T. acknowledges financial support from the Agence Nationale de la Recherche (contract ANR-16-CE30-0017-01). The electron microscopy imaging is supported by "Investissements d'Avenir" LabEx PALM (contract ANR-10-LABX-0039-PALM) and we thank J ril Degrouard for his technical assistance with the

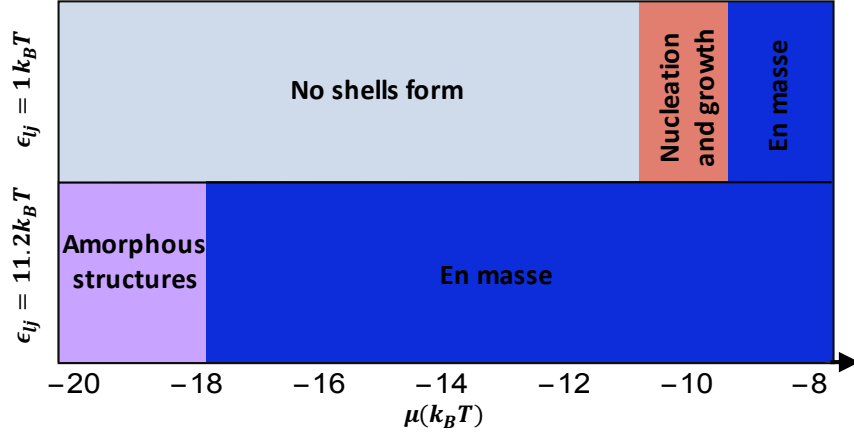


Figure 13: Capsids form with different mechanisms based on the strength of protein-genome interaction. At weak Lennard-Jones interaction ( $\epsilon_{lj} = 1.0k_B T$ ), for the chemical potentials between  $-11 k_B T$  and  $-9 k_B T$  capsids form through the nucleation and growth mechanism. For higher chemical potentials, the en masse assembly becomes the dominant pathway for the formation of  $T = 3$  structures. However, at strong Lennard-Jones interaction ( $\epsilon_{lj} = 11.2 k_B T$ ) we obtain amorphous structures for low chemical potentials but  $T=3$  structures form through en masse pathway for higher chemical potentials. Other parameters used in the simulations are  $k_s=600k_B T$ ,  $k_b=200k_B T$  and  $\epsilon_{hp} = -1 k_B T$ .

microscope. L.M., R.L.R. and G.T. also acknowledge the European Synchrotron Radiation Facility (ESRF) and the SOLEIL synchrotron for allocation of synchrotron beam time on ID02 and SWING beamlines.

## Supporting Information Available

The following files are available free of charge.

Additional seven supporting figures, a supporting table and more details regarding: The comparison of the elastic energy obtained in the simulations and continuum elasticity theory for a spherical cap; comparison of the appearance of disclinations in the simulations and continuum theory; the impact of elasticity modulus on the order-disorder transition; derivation of equation (7) in the main text; detailed protocols for sample preparation in the experiments; measurement of polydispersity index; SAXS patterns of NPCs fitted with a polydisperse vesicle model and table of fitting parame-

ters for the SAXS patterns; cryo-transmission electron microscopy images of nanotubes; the structure and phase diagram for assembly of capsid proteins around the smaller core size.

## References

- (1) Flint, S.; Enquist, L.; Krug, R.; Racaniello, V.; A.M., S. *Principles of Virology: Molecular Biology, Pathogenesis, and Control*; Press: Washington DC, 2000.
- (2) Sun, J.; DuFort, C.; Daniel, M.-C.; Murali, A.; Chen, C.; Gopinath, K.; Stein, B.; De, M.; Rotello, V. M.; Holzenburg, A.; Kao, C. C.; Dragnea, B. Core-controlled polymorphism in virus-like particles. *Proceedings of the National Academy of Sciences* **2007**, *104*, 1354–1359.
- (3) Šiber, A.; Podgornik, R. Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses. *Physical Review E* **2008**, *78*, 051915.
- (4) Erdemci-Tandogan, G.; Wagner, J.; Van Der Schoot, P.; Podgornik, R.; Zandi, R. RNA topology remodels electrostatic stabilization of viruses. *Physical Review E* **2014**, *89*, 032707.
- (5) Erdemci-Tandogan, G.; Wagner, J.; van der Schoot, P.; Podgornik, R.; Zandi, R. Effects of RNA branching on the electrostatic stabilization of viruses. *Physical Review E* **2016**, *94*, 022408.
- (6) Zlotnick, A. Theoretical aspects of virus capsid assembly. *Journal of Molecular Recognition* **2005**, *18*, 479–490.
- (7) Vernizzi, G.; Sknepnek, R.; De La Cruz, M. O. Platonic and Archimedean geometries in multicomponent elastic membranes. *Proceedings of the National Academy of Sciences* **2011**, *108*, 4292–4296.

- (8) Zlotnick, A.; Aldrich, R.; Johnson, J. M.; Ceres, P.; Young, M. J. Mechanism of Capsid Assembly for an Icosahedral Plant Virus. *Virology* **2000**, *277*, 450.
- (9) Perlmutter, J. D.; Qiao, C.; Hagan, M. F. Viral genome structures are optimal for capsid assembly. *Elife* **2013**, *2*, e00632.
- (10) Mohajerani, F.; Hagan, M. F. The role of the encapsulated cargo in microcompartment assembly. *PLoS computational biology* **2018**, *14*, e1006351.
- (11) Weiss, P. Armor-plated puzzle: Deciphering the code of viral geometry. *Science News* **2005**, *168*, 152–154.
- (12) Johnson, J. E.; Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *Journal of molecular biology* **1997**, *269*, 665–675.
- (13) Caspar, D. L.; Klug, A. Physical principles in the construction of regular viruses. Cold Spring Harbor symposia on quantitative biology. 1962; pp 1–24.
- (14) Twarock, R.; Luque, A. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nature Communications* **2019**, *10*, 1–9.
- (15) Bancroft, J. B. The self-assembly of spherical plant viruses. *Adv. Virus Res.* **1970**, *16*, 99.
- (16) Chen, T.; Zhang, Z.; Glotzer, S. C. A precise packing sequence for self-assembled convex structures. *Proceedings of the National Academy of Sciences* **2007**, *104*, 717.
- (17) 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**,
- (18) Levinthal, C. How to fold graciously. *Mossbauer spectroscopy in biological systems* **1969**, *67*, 22–24.



- (19) Kindt, J.; Tzlil, S.; Ben-Shaul, A.; Gelbart, W. M. DNA packaging and ejection forces in bacteriophage. *Proceedings of the National Academy of Sciences* **2001**, *98*, 13671–13674.
- (20) Purohit, P. K.; Kondev, J.; Phillips, R. Mechanics of DNA packaging in viruses. *Proceedings of the National Academy of Sciences* **2003**, *100*, 3173–3178.
- (21) May, E. R. Recent developments in molecular simulation approaches to study spherical virus capsids. *Molecular Simulation* **2014**, *40*, 878–888.
- (22) Singaram, S. W.; Garmann, R. F.; Knobler, C. M.; Gelbart, W. M.; Ben-Shaul, A. Role of RNA Branchedness in the Competition for Viral Capsid Proteins. *The Journal of Physical Chemistry B* **2015**, *119*, 13991.
- (23) Tubiana, L.; Bozic, A. L.; Micheletti, C.; Podgornik, R. Synonymous Mutations Reduce Genome Compactness in Icosahedral ssRNA Viruses. *Biophysical J.* **2015**, *108*, 194.
- (24) Erdemci-Tandogan, G.; Orland, H.; Zandi, R. RNA Base Pairing Determines the Conformations of RNA Inside Spherical Viruses. *Physical review letters* **2017**, *119*, 188102.
- (25) Larson, S. B.; McPherson, A. Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struc. Biol.* **2001**, *11*, 59.
- (26) Hagan, M. F. Modeling Viral Capsid Assembly. *Adv. Chem. Phys.* **2013**, *155*, (in press).
- (27) Perlmutter, J. D.; Perkett, M. R.; Hagan, M. F. Pathways for virus assembly around nucleic acids. *Journal of molecular biology* **2014**, *426*, 3148–3165.
- (28) Chevreuril, M.; Law-Hine, D.; Chen, J.; Bressanelli, S.; Combet, S.; Constantin, D.; Degrouard, J.; Möller, J.; Zeghal, M.; Tresset, G. Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nat. Commun.* **2018**, *9*, 3071.

- (29) Garmann, R. F.; Comas-Garcia, M.; Knobler, C. M.; Gelbart, W. M. Physical principles in the self-assembly of a simple spherical virus. *Acc. Chem. Res.* **2016**, *49*, 48–55.
- (30) Garmann, R. F.; Comas-Garcia, M.; Gopal, A.; Knobler, C. M.; Gelbart, W. M. The assembly pathway of an icosahedral single-stranded RNA virus depends on the strength of inter-subunit attractions. *Journal of molecular biology* **2014**, *426*, 1050–1060.
- (31) Cadena-Nava, R. D.; Comas-Garcia, M.; Garmann, R. F.; Rao, A.; Knobler, C. M.; Gelbart, W. M. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *Journal of virology* **2012**, *86*, 3318–3326.
- (32) Daniel, M. C.; Tsvetkova, I. B.; Quinkert, Z. T.; Murali, A.; De, M.; Rotello, V. M.; Kao, C. C.; Dragnea, B. Role of Surface Charge Density in Nanoparticle-Templated Assembly of Bromovirus Protein Cages. *ACS Nano* **2010**, *4*, 3853.
- (33) Lin, H.-K.; van der Schoot, P.; Zandi, R. Impact of Charge Variation on the Encapsulation of Nanoparticles by Virus Coat Proteins. *Phys. Biol.* **2012**, *9*, 066004.
- (34) Garmann, R. F.; Goldfain, A. M.; Manoharan, V. N. Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proceedings of the National Academy of Sciences* **2019**,
- (35) Elrad, O. M.; Hagan, M. F. Encapsulation of a Polymer by an Icosahedral Virus. *Phys. Biol.* **2010**, *7*, 045003.
- (36) Siber, A.; Zandi, R.; Podgornik, R. Thermodynamics of nanospheres encapsulated in virus capsids. *Phys. Rev. E* **2010**, *81*, 051919.
- (37) Rotskoff, G. M.; Geissler, P. L. Robust nonequilibrium pathways to microcompartment assembly. *Proceedings of the National Academy of Sciences* **2018**, *115*, 6341–6346.

- (38) Wagner, J.; Zandi, R. The robust assembly of small symmetric nanoshells. *Biophysical journal* **2015**, *109*, 956–965.
- (39) Hicks, S. D.; Henley, C. Irreversible growth model for virus capsid assembly. *Physical Review E* **2006**, *74*, 031912.
- (40) Kegel, W. K.; van der Schoot, P. Competing Hydrophobic and Screened-Coulomb Interactions in Hepatitis B Virus Capsid Assembly. *Biophysical Journal* **2004**, *86*, 3905–3913.
- (41) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier, 2001; Vol. 1.
- (42) Borodavka, A.; Tuma, R.; Stockley, P. G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proceedings of the National Academy of Sciences* **2012**, *109*, 15769–15774.
- (43) Ali, A.; Roossinck, M. J. Rapid and efficient purification of cowpea chlorotic mottle virus by sucrose cushion ultracentrifugation. *J. Virol. Methods* **2007**, *141*, 84–86.
- (44) Franke, D.; Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* **2009**, *42*, 342–346.
- (45) Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503.
- (46) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (47) Li, S.; Zandi, R.; Travesset, A.; Grason, G. M. Ground States of Crystalline Caps: Generalized Jellium on Curved Space. *Phys. Rev. Lett.* **2019**, *123*, 145501.

- (48) Li, S.; Roy, P.; Travesset, A.; Zandi, R. Why large icosahedral viruses need scaffolding proteins. *Proceedings of the National Academy of Sciences* **2018**, *115*, 10971–10976.
- (49) Panahandeh, S.; Li, S.; Zandi, R. The equilibrium structure of self-assembled protein nano-cages. *Nanoscale* **2018**, *10*, 22802–22809.
- (50) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool. *MODELLING AND SIMULATION IN MATERIALS SCIENCE AND ENGINEERING* **2010**, *18*.
- (51) Hu, Y.; Zandi, R.; Anavitarte, A.; Knobler, C. M.; Gelbart, W. M. Packaging of a Polymer by a Viral Capsid: The Interplay between Polymer Length and Capsid Size. *Biophysical Journal* **2008**, *94*, 1428–1436.
- (52) Tresset, G.; Tatou, M.; Le Coeur, C.; Zeghal, M.; Bailleux, V.; Lecchi, A.; Brach, K.; Klekotko, M.; Porcar, L. Weighing polyelectrolytes packaged in viruslike particles. *Phys. Rev. Lett.* **2014**, *113*, 128305.
- (53) Fox, J. M.; Wang, G.; Speir, J. A.; Olson, N. H.; Johnson, J. E.; Baker, T. S.; Young, M. J. Comparison of the native CCMV virion with in vitro assembled CCMV virions by cryoelectron microscopy and image reconstruction. *Virology* **1998**, *244*, 212–218.
- (54) Lavelle, L.; Gingery, M.; Phillips, M.; Gelbart, W. M.; Knobler, C. M.; Cadena-Nava, R. D.; Vega-Acosta, J. R.; Pinedo-Torres, L. A.; Ruiz-Garcia, J. Phase diagram of self-assembled viral capsid protein polymorphs. *Journal of Physical Chemistry B* **2009**, *113*, 3813–3819.
- (55) Garmann, R. F.; Comas-Garcia, M.; Knobler, C. M.; Gelbart, W. M. Physical principles in the self-assembly of a simple spherical virus. *Accounts of chemical research* **2015**, *49*, 48–55.

# Graphical TOC Entry

