From Disorder to Icosahedral Symmetry: How Conformation-Switching Subunits Enable RNA Virus Assembly

Siyu Li[1], Guillaume Tresset[2], Roya Zandi[1*]

[1]Department of Physics and Astronomy, University of California, Riverside, 92521, United States.

[2]Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides, 91405 Orsay, France.

[*]Corresponding author. Email:roya.zandi@ucr.edu

Icosahedral capsids are ubiquitous among spherical viruses, yet their assembly pathways and governing interactions remain elusive. We present a molecular dynamics model that incorporates essential physical and biological interactions, including protein diffusion, genome flexibility, and a conformational switch that mimics allostery and activates the elastic properties of proteins upon binding. This switch makes the simulations computationally feasible and enables the assembly of icosahedral capsids around a flexible genome—overcoming long-standing limitations in previous models. Using this framework, we successfully reproduce the self-assembly of subunits around a flexible genome into icosahedral shells with T numbers greater than one – most notably $T = 3$, the most common structure in nature – a feat that rigid-body models have so far failed to achieve. We systematically explore the range of morphologies formed with different genome architectures, in line with in vitro experiments using cowpea chlorotic mottle virus capsid proteins: viral RNAs with more complex structure form more complete and stable capsids than linear ones. These results provide a predictive framework for genome-guided assembly and capsid design.

Teaser Novel conformation-switching subunits enable simulations where disordered RNA–protein complexes self-organize into symmetric shells.

Single-stranded (ss) RNA viruses, which impact humans, animals, and plants, constitute the largest and most widespread genetic class of viruses (*1–6*). During their replication process, hundreds or even thousands of proteins come together to construct the protective viral shell (capsid), enclosing the genetic materials (*7–9*). Despite their profound impact on our daily lives, as exemplified by the recent Covid-19 pandemic, our understanding of the virus formation, both *in vitro* and *in vivo*, remains unususally limited.

Experimentally, characterizing assembly pathways through either the techniques that monitor individual capsids or bulk approaches is challenging because of the small size of the virus and the transient, short-lived nature of its intermediate structures (*10–14*). Due to lack of experimental resolutions and computationally very expensive atomistic simulations for the entire capsid formation (*15–17*), coarse-grained computational models have been employed to explore the role of various factors critical to the assembly process (*18–26*).

Using rigid triangular subunits, Molecular Dynamics (MD) simulations have enhanced our understanding of the formation of the smallest icosahedral capsid, $T = 1$, around the genome (*27–34*). Note that the total number of proteins in an icosahedral structure is equal to $60T$ where T is the triangulation number assuming certain integers $(1, 3, 4, 7\ldots)$ (*35*). Because a $T = 1$ structure requires only 20 triangles to form a closed icosahedral shell, its assembly pathway is relatively straightforward. In contrast, the assembly mechanisms and intermediate structures leading to $T = 3$ and $T = 4$ capsids—the most prevalent structures in nature—remain largely unexplored.

To our knowledge, rigid-body models have failed to date to achieve the spontaneous assembly of capsid proteins around long viral genomes into highly symmetric $T = 3$ structures, while also resolving the intermediate stages of assembly. Previous studies of $T = 3$ capsid assembly have either focused on empty shells or employed simulations with spherical cargoes, which artificially simplify the assembly pathway by imposing the capsid's curvature (*19, 26, 36*). However, the mechanisms underlying the assembly and packaging of RNA or a flexible genome remain largely unexplored. It is remarkable that so many subunits can be absorbed onto a flexible genome, initially forming a disordered, irregular complex, and ultimately assembling into a stable, closed shell with icosahedral symmetry. The absence of MD simulations of $T = 3$ or $T = 4$ viruses in the literature highlights

how the additional degrees of freedom and complexity introduced by a flexible genome make it challenging to explore the interactions and mechanisms that enable capsid proteins to overcome barriers between disordered and ordered states.

In this paper, we introduce a novel subunit design that incorporates protein flexibility through a conformational switching mechanism that mimics allosteric regulation. This enables quasi-equivalent positioning and provides a general, computationally efficient framework for simulating capsid assembly via molecular dynamics. The model accounts for protein diffusion and captures the packaging of a flexible genome by identical capsid subunits, leading to the spontaneous formation of $T = 3$ structures. While our primary focus is on $T = 3$ structures, the model also successfully assembles $T = 4$ capsids—representing the first molecular dynamics simulations to demonstrate the formation of a complete $T = 4$ shell around a flexible genome. By including only the most essential physical and biological interactions, this allostery-inspired design makes it feasible to explore full assembly kinetics and intermediate states that were previously computationally prohibitive—significantly reducing simulation cost while preserving biologically relevant behavior.

The model captures the emergence of different triangulation numbers without relying on complex, location-specific assembly rules—such as varying subunit geometries, tuning interaction strengths, or engineering proteins to adopt distinct conformations based on their positions within the shell. Instead, it employs a single type of subunit, as is the case for most $T = 3$ and $T = 4$ viruses in nature, and relies solely on universal physical principles. This is consistent with observations that similar capsid symmetries arise in viruses with highly divergent amino acid sequences.

Notably, our findings reveal a virus assembly pathway involving multiple intermediate states, in contrast to previous studies that identified a narrow pathway for $T = 3$ and $T = 4$ viruses (*14*). We find that the most efficient pathway involves the assembly of capsid fragments at a few locations along the genome, followed by the attraction of these fragments leading to genome condensation, which facilitates subunit rearrangement, see Fig. 2 and Movie. S1. Counterintuitively, we discover multiple pathways in which numerous fragments—containing varying numbers of subunits and defects—merge to form a perfectly closed icosahedral shell. We note that most viruses with triangulation number $T = 7$ or larger rely on scaffolding proteins for proper assembly. Although incorporating such features introduces additional complexity, the method presented here is sufficiently robust and flexible to address these systems; however, their exploration will be the subject

of a separate study.

To illustrate the model's capabilities, we compare the advantages of RNA packaging with those of a linear polymer using a combination of computational simulations and experimental techniques. Using small-angle X-ray scattering and cryo-transmission electron microscopy, we observe the assembly of $T = 3$ structures employing capsid proteins and viral RNA derived from a plant virus, as well as a non-viral RNA with a distinct topology. Our simulations capture trends that align with key features of the experimental data. The model's robustness enables us to use MD simulations to explore and interpret a broad range of previously inaccessible experimental phenomena, including virus disassembly and genome release, packaging signals, the influence of RNA size and secondary structure, the role of protein N-terminal domains, and the geometric diversity of building blocks and heterogeneous protein subunits, to name a few.

Results

Design principles

As in most previous simulations of virus assembly, we use triangular subunits as the basic building blocks (see Fig. 1) (*7, 40*). Mohajerani *et al.* referred to triangular units as "trimers-of-dimers" and successfully employed them to model the assembly of empty HBV capsids (*40*). Triangular subunits have also been widely used in other studies to represent coarse-grained assembly intermediates (*25, 28, 30, 31, 33*). In our model, these subunits can be interpreted either as simple trimers or, following Mohajerani *et al.*, as trimers-of-dimers; however, this distinction has no impact on the final assembly outcomes. Although most viruses initially form dimers, which subsequently assemble into trimers and then into pentamers and hexamers, at this coarse-grained level, simulating the assembly of 90 dimers or 60 trimers to form a $T = 3$ structure does not fundamentally alter the assembly pathway. In both cases, the formation process proceeds through disordered intermediate structures, and the interactions between subunits and the genome must guide the system through energy barriers to ultimately produce a stable icosahedral shell. This modeling approach captures the essential features of viral assembly while significantly reducing computational complexity.

In general, when a protein diffuses freely in solution, its elastic properties do not significantly

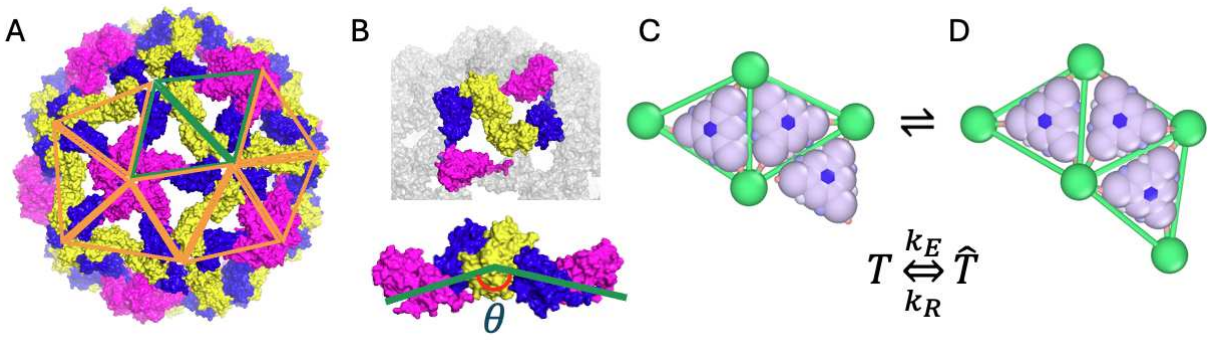

Figure 1: Schematic representation of subunits inspired by the capsid structure of Cowpea Chlorotic Mottle Virus (CCMV). A. A swollen CCMV capsid (PDB: 8CPY) at pH 7.5 (*37*) is displayed, with several triangles labeled to denote the subunits used in the simulations. The protein colors are assigned based on their symmetry group. B. Display of two trimeric subunits, highlighting the torsional angle (θ) associated with protein spontaneous curvature. Recent experiments demonstrate that CCMV capsid proteins first assemble into dimers, with the subsequent addition of dimers resulting in the formation of trimers, pentamers and hexamers. Based on this observation, we selected trimers as building blocks, primarily due to their ability to easily incorporate the proteins' elasticity and spontaneous curvature, both of which are crucial for the formation of closed shells. C. Two elastic subunits, denoted by $\hat{T}$, are positioned adjacent to a rigid subunit, denoted by T. D. Upon interacting with another subunit, a conformational change is triggered—mimicking an allosteric effect—whereby the subunit transitions into a bonding-capable state. This allosteric-like transition activates specific elastic interactions between subunits, analogous to how conformational shifts in proteins can enable or regulate binding in biological systems. At this stage, elasticity becomes relevant, and the subunit is enclosed by a green elastic triangle. The parameters k_E and k_R correspond to the effective association and dissociation rates of the subunits, respectively. Images are made using PyMOL and OVITO (*38, 39*).

impact the assembly process. To make our simulations computationally efficient, we therefore model free subunits as rigid bodies during their diffusion in solution. However, once a protein interacts with or attaches to another, elasticity becomes essential. This behavior is consistent with an allosteric mechanism, where interaction triggers a conformational switch that activates new bonding capabilities. The principle of quasi-equivalence—which states that identical proteins can adopt different conformations depending on their location within a capsid—further suggests that conformational flexibility, governed by such allosteric-like transitions, is crucial for a protein to occupy quasi-equivalent positions within the shell.

Accordingly, during simulations, when two free trimers diffuse into close proximity and begin interacting, we introduce elasticity by enclosing each free trimer within an elastic triangular frame (the green triangle shown in Fig. 1C). We refer to these conformationally flexible units as "elastic trimers" ($\hat{T}$) and the free subunits as "rigid trimers" (T). This transition enables subunits to undergo a conformational change reminiscent of an allosteric response—activating specific binding capabilities upon interaction—analogous to conformational changes observed in proteins during the assembly process.

As two elastic trimers interact, they merge at their interfaces, and elastic bonds form between them. The energetic of a growing shell then includes both stretching and bending components is,

$$\begin{aligned} E_{elasticity} &= E_s + E_b + E_s^{inner} \\ &= \frac{1}{2}k_s \sum_i (l_i - l_0)^2 + k_b \sum_i (1 - \cos(\theta_i - \theta_0)) \\ &+ \frac{1}{2}k_s \sum_{\hat{T}_i} \sum_{j=1}^{3} (l_{\hat{T}_i,j} - l_{\hat{T}0})^2, \end{aligned} \tag{1}$$

where $l_0 = 3a$ is the equilibrium size of each trimer side with $a = 3$ nm the fundamental length of the system, θ_0 the preferred dihedral angle closely related to the spontaneous radius of curvature (*20*), l_i and θ_i are the length and the dihedral angle of the bond i, and $k_s = 100k_BT/a^2$ and $k_b = 500k_BT$ are the stretching and bending moduli, respectively. Moreover, each elastic vertex, E_V (green ball) is connected to a vertex of a rigid triangle through a ligand (T_V) with equilibrium length $l_{\hat{T}0}$, which possesses an elastic energy E_s^{inner} (see Materials and Methods for more details).

Figure 1C illustrates a rigid trimer positioned near two elastic trimers. Upon interaction, the originally rigid triangle undergoes a conformational change with a certain probability, transitioning

into an elastic state. This conformational switching mimics an allosteric response, where binding activates new interaction capabilities. The probability reflects the rate at which two protein subunits associate (see below). The transition results in the formation of an elastic bond between the two trimers (see Fig. 1D).

The transformation between rigid and elastic states is reversible: trimers can dissociate and revert to rigid units. The rates of conformational change between a rigid trimer and an elastic trimer are governed by k_E and k_R, which correspond to the probabilities of transitioning from a rigid state to an elastic state (p_E) and from an elastic state back to a rigid state (p_R), respectively (see Materials and Methods for details). These rates and probabilities reflect the reaction times required for two proteins in close proximity to undergo conformational changes and interact, or alternatively, to dissociate and return to their original rigid form. With these subunits, we now have the capacity to explore stages of virus assembly that were previously inaccessible.

Fragments of capsids join to form a perfect icosahedral shell

Using the subunits depicted in Fig. 1, we simulate protein dynamics employing a Langevin integrator. The energy of the system can be written as

$$\Delta G = E_{elasticity} + E_{p-p} + E_c + E_{ele} - N_T \mu, \tag{2}$$

where the first term is provided in Eq. 1 above, the second term corresponds to the protein-protein interaction between two small ligands positioned at the edges of the trimers (for particles T_a and T_b in Fig. S1, see Materials and Methods), and the third term denotes the interaction arising from protein conformational changes induced by allosteric effects $E_c = k_B T N_L \log p_R/p_E$ with N_L the number of elastic bonds between two elastic trimers in a growing shell. The fourth term represents the electrostatic interaction between the proteins and genome and the last term corresponds to the chemical potential, $\mu = k_B T \log(c/c_0)$, of the free proteins in solution with c their concentration and c_0 a reference state.

Figure 2 illustrates the snapshots of the formation of a $T = 3$ structure around a linear chain with a length of $l = 80a$ at various time t with a protein concentration of $C_p = 100\,\mu$M. The probabilities at which the proteins undergo conformational changes are $p_E = 1.0$ and $p_R = 0.1$, which controls the reversibility of the process and is related to the strength of interaction due to the proteins

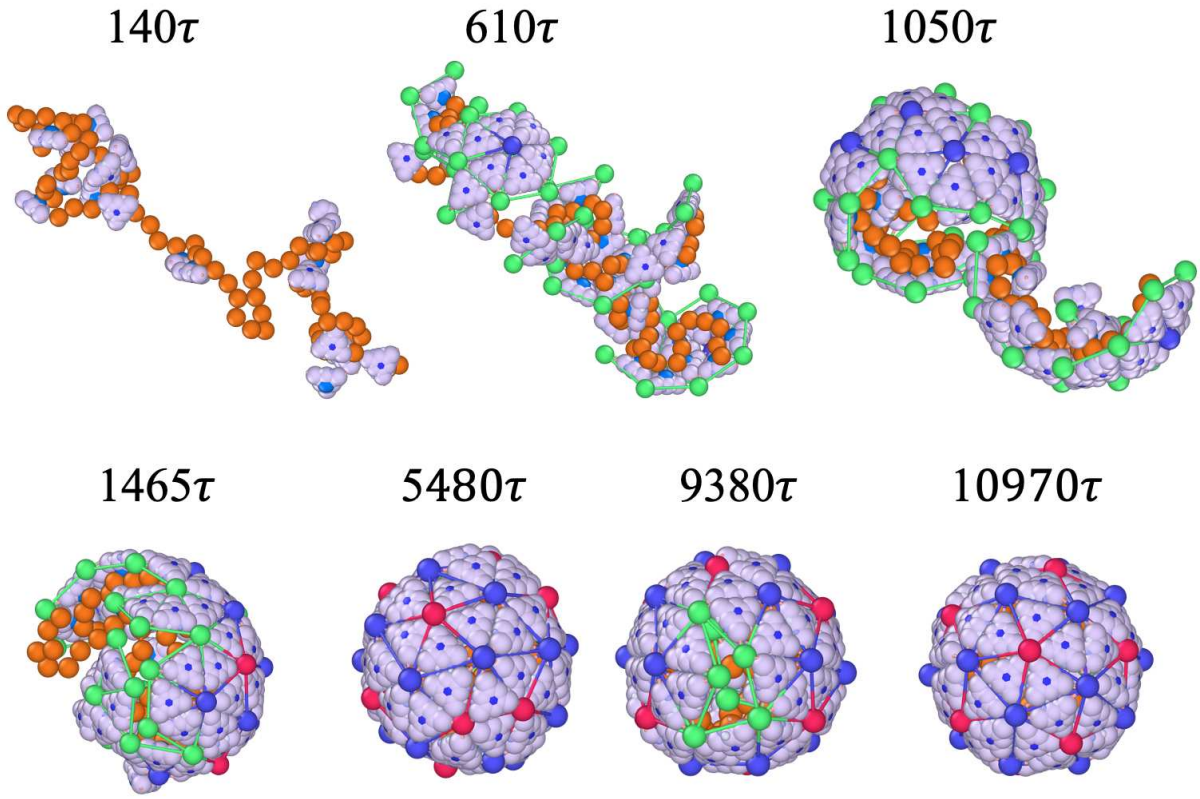

Figure 2: Snapshots of the assembly of a $T = 3$ particle at the protein concentration $C_p = 100\,\mu\text{M}$ with a chain length $80a$. Note that free proteins in solution are not shown to focus on the assembly process (see Movie S1 for simulations including background proteins). Right at the beginning the proteins aggregate around the genome due to the electrostatic interaction between positively charged proteins and the negatively charged linear chain. The proteins also attract each other with a strength of $\epsilon_{pp} = 5k_BT$. As more proteins aggregate, additional subunits transform into elastic ones capable of forming bonds, with probabilities $p_E = 1.0$ and $p_R = 0.1$, corresponding to the rates of protein–protein association and dissociation (k_E and k_R, respectively). The other parameters in the systems are stretching modulus $k_s = 100k_BT/a^2$ and $k_b = 500k_BT$.

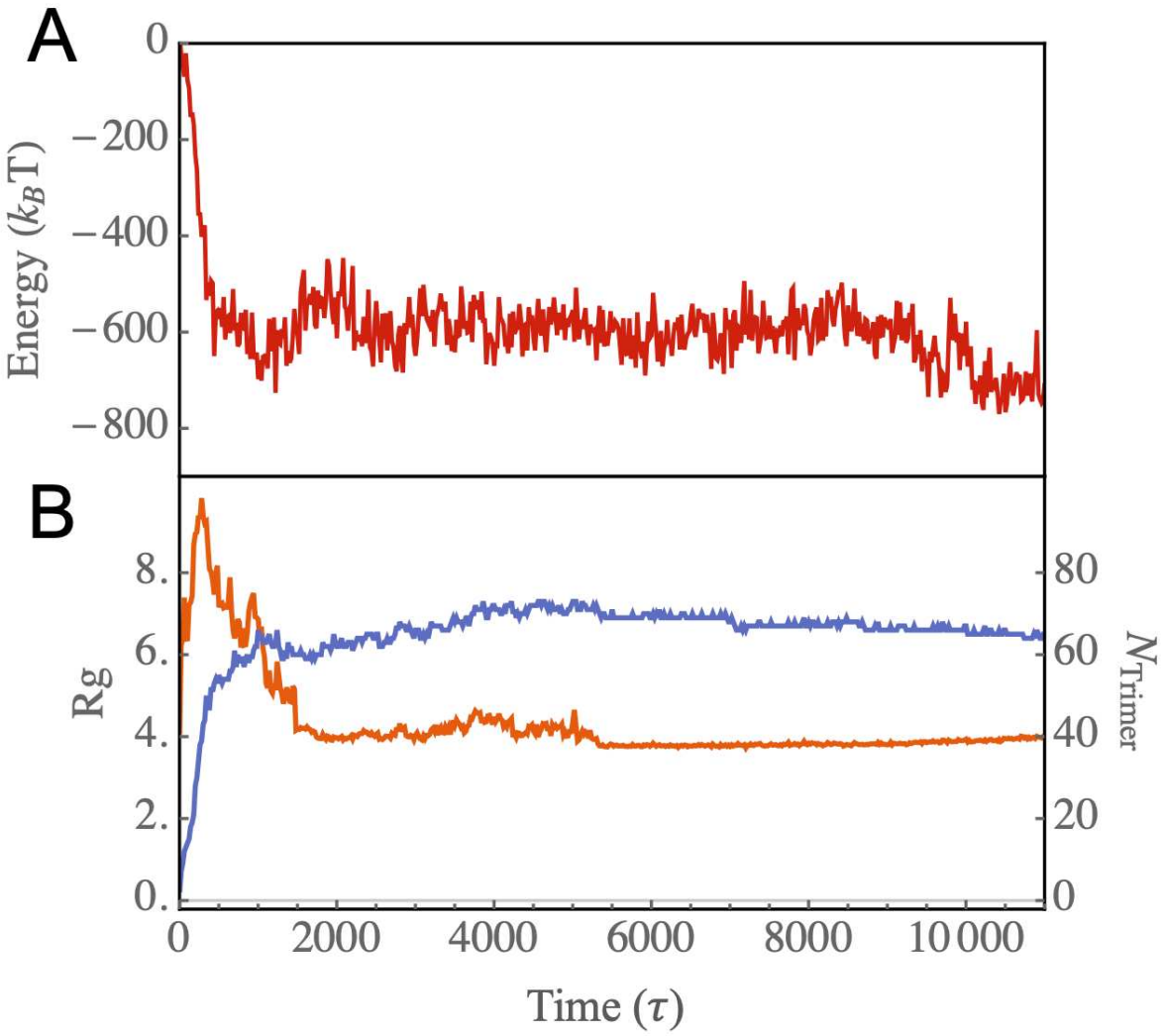

Figure 3: A. Energy of the genome-protein complex as a function of time for the assembly pathway illustrated in Fig. 2. B. The number of trimers attached to the genome (blue curve) and the radius of gyration of protein-RNA complex (R_g, orange) as a function of time. The radius of gyration includes contributions from both the genome and all proteins bound to it.

conformational changes, see Mateirals and Methods for details. As noted above, these probabilities are related to the reaction times for two proteins in close proximity to undergo conformational changes and interact, or alternatively, to dissociate and revert to their free-state forms.

Throughout the simulations, an elastic trimer in an unfavorable position may transition to a rigid state with probability p_R, allowing it to detach from the shell and diffuse back into the reservoir. Conversely, a trimer located at an energetically favorable site remains attached to the growing shell—even after transitioning to a rigid body—and may eventually revert to an elastic state. To become fully detached, a bound protein must first dissociate by converting to the rigid state, which corresponds to the removal of the elastic bond between subunits. Following dissociation, subunits can still experience hydrophobic interactions; however, the formation of an elastic bond represents a specific, conformation-dependent interaction. Thus, conformational change is required for both specific binding and detachment.

Figure 2 (see also Movie. S1) shows that around $t = 140\tau$, several free proteins become absorbed into the genome and aggregate without yet forming elastic bonds. We note that τ represents the system time unit and, when calibrated with our experiments, we find it to be on the order of milliseconds (ms) (*6, 20*). As shown in the figure, around $t = 610\tau$, multiple nucleation sites emerge around the genome, forming hexamers (blue vertices and bonds) and larger oligomers. Note that the green vertices represent the elastic vertices, located only at the edge of the growing shell.

At about $t = 1050\tau$, smaller oligomers throughout the chain begin to merge, forming larger fragments that "squeeze" and encapsulate a significant portion of the genome within them. The presence of line tension, caused by the subunits having fewer neighbors at the edge, makes the intermediate states of the capsid energetically unfavorable. To this end, around $t = 1465\tau$, the fragments join and rearrange to minimize the energy associated with the edge of the growing shell. Quite interestingly, around $t = 5480\tau$, the shell is completely closed but it has an irregular shape and the "wrong" symmetry. For the remainder of the simulations, the pentamers and hexamers that initially formed in incorrect positions associate or dissociate until they eventually assemble into a perfect icosahedral shell around $t = 10970\tau$.

The transition from an irregular shell to an icosahedral capsid takes a rather long time as the complete shell needs to partially disassemble and many subunits have to rearrange to form a perfect shell. Fig. 3A shows the total energy of a growing shell as a function of time. The figure reveals the

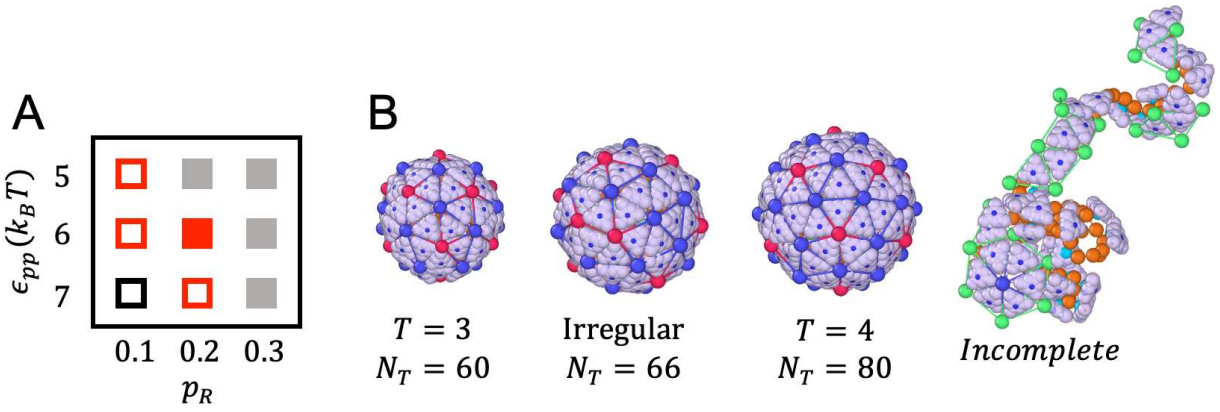

Figure 4: A. Shape "phase" diagram as a function of the strength of protein-protein interaction (ϵ_{pp}) and p_R, which corresponds to the reaction time for protein detachment. The hollow red squares denote regions where $T = 3$ structures were observed for less than 50% of the time, while the solid red square indicates when they were observed for more than 50% of the time. Black squares highlight areas where only $T = 4$ or irregular shapes appear, while the gray solid region denotes instances where only incomplete shells were observed. B. Representative structures corresponding to the phase diagram in A. The simulations were performed at the protein concentration $C_p = 100\,\mu\text{M}$, spontaneous radius of curvature $R_0 = 4.2a$, and genome length $L = 80a$.

presence of an energy barrier for this disorder to order transition, stemming from the dissociation of some subunits from the complete irregular shell, see the capsid around $t = 9380\tau$ in Fig. 2.

Figure 3B shows the radius of gyration of the complex of the genome and proteins (the orange line) and the number of trimers around the genome as a function of time (the blue line). The plot reveals an initial rise in the radius of gyration, which is due to the initial configuration of the chain. As time passes, the capsid fragments join together, consequently condensing the chain. The blue plot in Fig. 3B indicates that the number of trimers quickly reaches around 60, forming an irregular closed shell. However, the transition from disorder to order within the capsid structure is a slow process, taking an order of magnitude longer to complete. It is interesting to note that some $T = 3$ capsids contain more than 60 trimers. In these cases, we observe that some subunits are enclosed within the capsid, a behavior typically associated with linear chains. In the next section, we present the results of our simulations with branched polymers, which more closely resemble the structure of RNA.

It is important to note that a key factor influencing the formation of error-free $T = 3$ icosahedral shells is the timescale of protein binding and unbinding events, which is governed by the probabilities of conformational changes. In all simulations presented in this paper, we set the probability of transitioning from the rigid to the elastic state (p_E) to 1, while varying the probability of the reverse transition (p_R). This means that a bound elastic subunit can revert to the rigid state with probability p_R, after which it may either remain attached or detach and diffuse away (see Materials and Methods for details). Importantly, a protein can fully dissociate from the assembling shell only when it is in the rigid state.

The two-dimensional phase diagram in Fig.4A illustrates how assembly products depend on the conformational switching probability p_R and the strength of protein–protein interactions ϵ_{pp} (see Eq.2 and Materials and Methods). Two distinct regions in phase space yield predominantly icosahedral structures: (1) small p_R values combined with weak protein–protein interactions, and (2) large p_R values with strong interactions. For small p_R (i.e., lower detachment probability), the process is less reversible, necessitating weaker hydrophobic interactions to permit error correction. A low p_R reflects strong specific binding, likely mimicking stable post-binding conformational states (akin to allosteric regulation), which hinder dissociation. Conversely, increasing p_R enhances the detachment rate, but sufficiently strong hydrophobic interactions can compensate, still enabling the formation of ordered capsids.

The red squares in the phase diagram Fig. 4A indicates the regions in which we observe $T = 3$ shells as well as some irregular structures. The difference between the hollow and filled squares is that with the filled ones, we observe a specific structure for the majority of the time, whereas with the empty ones, this structure is observed for less than half of the time. For instance, when $p_R = 0.1$ and $\epsilon_{pp} = 5k_BT$, we observe $T = 3$ shells, but the majority of structures assume irregular shapes (see Fig. 4B for a representative structure). Upon further increasing ϵ_{pp} to $7k_BT$, at $p_R = 0.1$, we observe a mix of $T = 4$ structures and irregular shells (the black square), see Fig. S2.

For $\epsilon_{pp} = 5k_BT$ (weak protein-protein interaction), since the number of trimers aggregated around genome is small, a high reversible rate ($p_R = 0.2$) inhibits the long-lasting formation of elastic bonds, thereby hindering capsid formation (the gray squares). However, if we increase protein-protein interactions ($\epsilon_{pp} = 6k_BT$), for $p_R = 0.2$, each trimer gains more chances to explore the energetically more favorable locations and we predominantly observe $T = 3$ structures (the

solid red square). An elastic trimer in an unfavorable position must first undergo a conformational change to ”break” its bond, after which it can detach and diffuse back into the reservoir. In contrast, if a trimer occupies an energetically favorable position, it remains attached to the growing shell even after undergoing a conformational change, losing its elastic bond, and becoming a rigid body. Eventually, it can revert to the elastic state and continue participating in assembly. This reversibility makes the formation of perfect $T = 3$ structures easier. A representative pathway of the formation of an icosahedral shell in this regime is shown in Movie S2. At higher values of p_R, the trimers can barely keep their elastic bonds, resulting in primarily aggregating around the genome without forming shells.

At the end of this section, we note that it is possible to monitor the assembly of multiple capsids in a solution of proteins with multiple genomes. Then, an important parameter will be the stoichiometry ratio between the concentrations of protein and genome, see Fig. S7.

Secondary structures of RNA alter the assembly products

In the previous section, we focused on the assembly of capsid proteins around a linear chain. However, due to base-pairing, the structures of viral genomes deviate significantly from a linear chain. In this section, we examine the impact of RNA secondary structure on assembly outcomes and evaluate the extent to which the model reproduces experimental observations. We begin by presenting the results of our simulations, followed by highlighting the findings of our experiments.

To investigate the impact of RNA secondary structure on assembly products, we designed three polymers with distinct topologies. Representative segments for each case are shown in Fig. 5A. The linear polymer, in the figure, contains 20 monomers without any branches. The chain denoted as branchA also comprises a total of 20 monomers, with four branching points highlighted in purple. The length of each branch is 2 monomers. Lastly, the branchB polymer also consists of a total of 20 monomers. Here, each branch has one monomer, with four branches extending from each branch point. Due to its morphology, a highly branched polymer like branchB has a smaller radius of gyration than the less branched branchA polymer, assuming both have the same length. It’s evident that both branched polymers have smaller radii of gyration than linear polymers, indicating their greater compactness.

The impact of polymer structures for different lengths at a fixed protein concentration $C_p = 100\,\mu$M is depicted in the two-dimensional phase diagram presented in Fig. 5B. As before, the empty red squares represent the regime in which $T = 3$ particles are observed before $t = 5000\tau$ in at least one out of the eight simulation runs, see Fig. S3. For the case of the linear polymer, if the genome length is relatively short, $L = 80a$, the icosahedral structures predominate. However, if we increase the length of genome to $L = 100a$, we obtain irregular closed shells with no specific symmetry (Fig. 5B (black squares) and Fig. S3). Upon increasing the length further to $120a$ or $140a$, we observe doublets (blue boxes in Fig. 5B), where two icosahedral structures, each missing one pentamer, share the chain as shown in Fig. 5D.

While $T = 3$ shells cannot package a linear chain of length $L = 100a$, they can assemble around branchA polymers of the same length. However, for longer chains, such as $L = 120a$ or 140, capsid proteins adopt larger, irregular structures to encapsidate the BranchA-type genomes, see black boxes in the phase diagram in Fig. 5B and Fig. S3. Quite unexpectedly, we find that capsid subunits form incomplete shells around shorter branchB-type chains, $L = 80a$ and $100a$, if the protein concentration is $C_p = 100\,\mu$M. This is basically due to the fact that at these lengths, the radius of gyration of the chains is comparable with the radius of a $T = 3$ capsid. Fig. 5B also shows that upon increasing the length of genome to $L = 120a$ or $140a$, for branchB-type chain we obtain $T = 3$ structures again.

To explore the interplay of genome structure and protein concentration for a fixed chain length of $L = 100a$, we vary the protein concentration and observe how the structures depicted in the second column of Fig. 5B (enclosed within a narrow blue rectangle) change. The resulting phase diagram is displayed in the Fig. 5C. We find that at a lower protein concentrations, $C_p = 50\,\mu$M, the capsids remain incomplete even at $t = 5000\tau$. As before, the filled rectangles mark the regions where a particular structure was observed at least 50% of the time. Upon increasing the protein concentration, the shells enclosing linear chains remain irregular; however, we do not observe any changes for both branchA and branchB polymers. Interestingly, the productivity of $T = 3$ shells increases dramatically when packaging branchB (highly branched) at high protein concentrations, as illustrated by filled red squares in the phase diagram, Fig. 5C (see also Fig. S4). It is important to note that the structures formed around branchB genomes are robust and insensitive to other parameters in the system including the probability of protein conformational changes p_R and the

spontaneous radius of curvature of proteins R_0, see Fig. S5.

Several features given by our model are consistent with experimental observations. *In vitro* assembly experiments with cowpea chlorotic mottle virus (CCMV) – a non-enveloped, single-stranded (ss)RNA plant virus that forms $T = 3$ structures – are carried out by mixing purified capsid proteins with two types of RNA. RNA C2 is the second genomic RNA segment of CCMV and has 2,767 nucleotides. RNA RF2 is a non-viral RNA segment with a similar length of 2,687 nucleotides. Fig. 6A shows the secondary structures of RF2 and C2, which were obtained employing the ViennaRNA package (*43*). The figure clearly shows that C2 has a branched conformation akin to the branchB chain of our simulations, whereas RF2 is more extended and can be assimilated to branchA or the linear chain. Small-angle X-ray scattering (SAXS) measurements of structures obtained with RF2 and C2 (Fig. 6B) reveal close morphologies from medium to high wavenumbers ($q > 0.02\,\text{Å}^{-1}$), *i.e.*, across length scales smaller than the size of a $T = 3$ capsid ($\simeq$30 nm). The SAXS patterns at small wavenumbers indicate that structures formed with RF2 display slight aggregation, resembling multiplets, as evidenced by a q^{-2} scaling (Fig. 6B). This scaling is reminiscent of a freely jointed chain structure. Cryotransmission electron microscopy (cryoTEM) images confirm the presence of a significant number of doublets and even triplets (Fig. 6C) with RF2, as predicted in our simulations for linear chains (see Fig. 5D). By contrast, we observe well dispersed $T = 3$ structures when C2 is used (see Fig. S6), although they are coexisting with a number of incomplete capsids (Fig. 6C). This is again in line with our simulations wherein 120-long branchB chains yield a majority of icosahedral structures (Fig. 5B). While additional experiments are needed for a quantitative comparison, it is notable that varying degrees of genome branchiness gives rise to distinct capsid morphologies.

Based on both experiments and simulations, we find that the secondary structure of RNA influences the resulting assembly products. Specifically, there is qualitative agreement between the two: when the RNA has fewer branch points—that is, when it is less branched—the assembly yields more incomplete shells, doublets, or closed irregular shells than icosahedral shells.

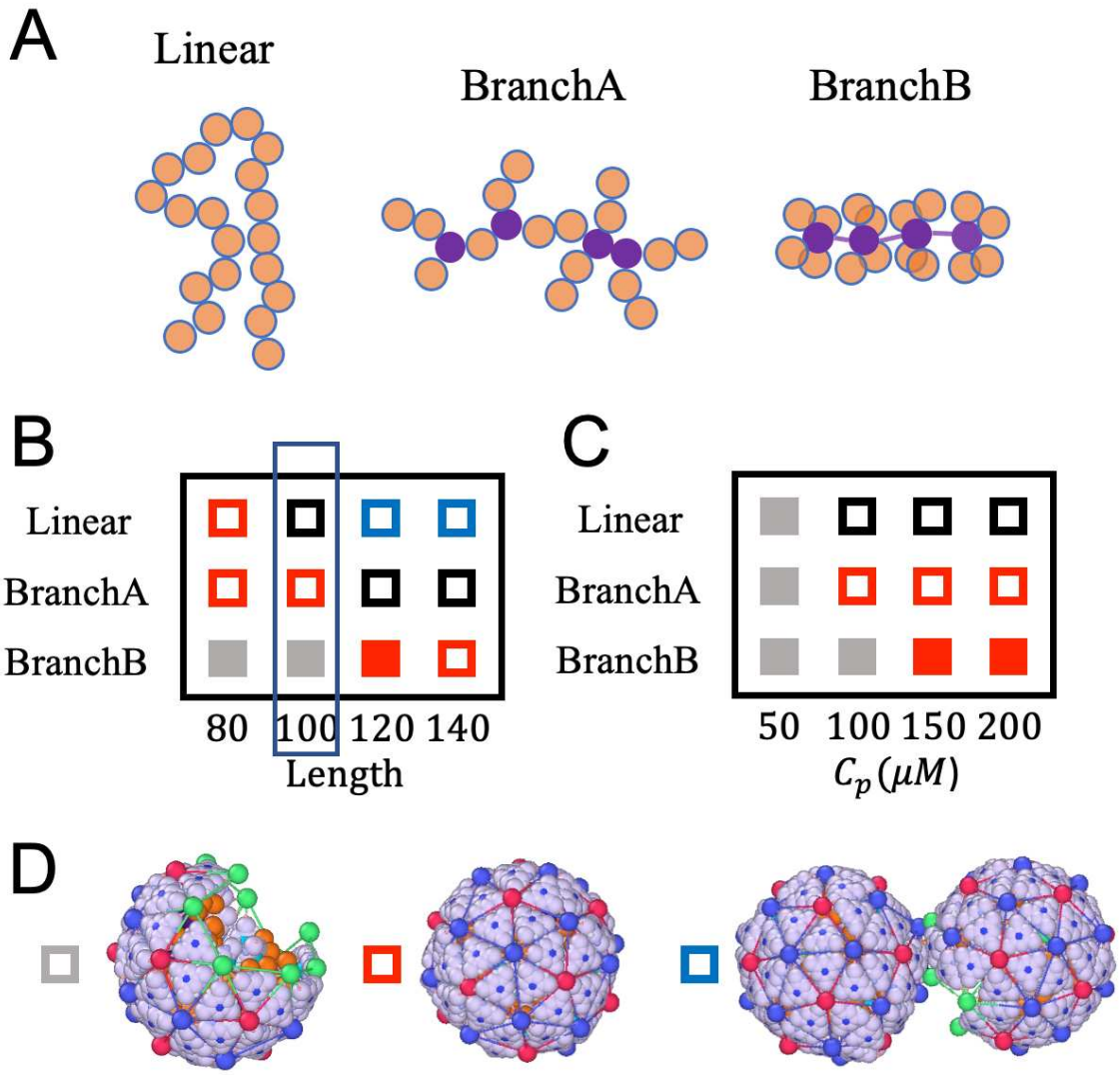

Figure 5: A. Schematic representations of RNA with various secondary structures, with branched points highlighted in purple. While in branchA-type polymer, branches are randomly distributed along the genome, in branchB type, each monomer serves as a branch point connected to four branches. Each branch consists of one monomer. B. Phase diagram with different secondary structures of RNA and various genome length at a protein concentration of $C_p = 100\,\mu$M. C. Phase diagram with different secondary structures of RNA and various protein concentrations. The genome length was set to $L = 100a$. D. Representation of an incomplete, a $T = 3$ and a doublet structures, corresponding to gray, red, and blue squares, respectively. The black squares in the phase diagram correspond to irregular structures; a representative shape is shown in Fig. 4B.

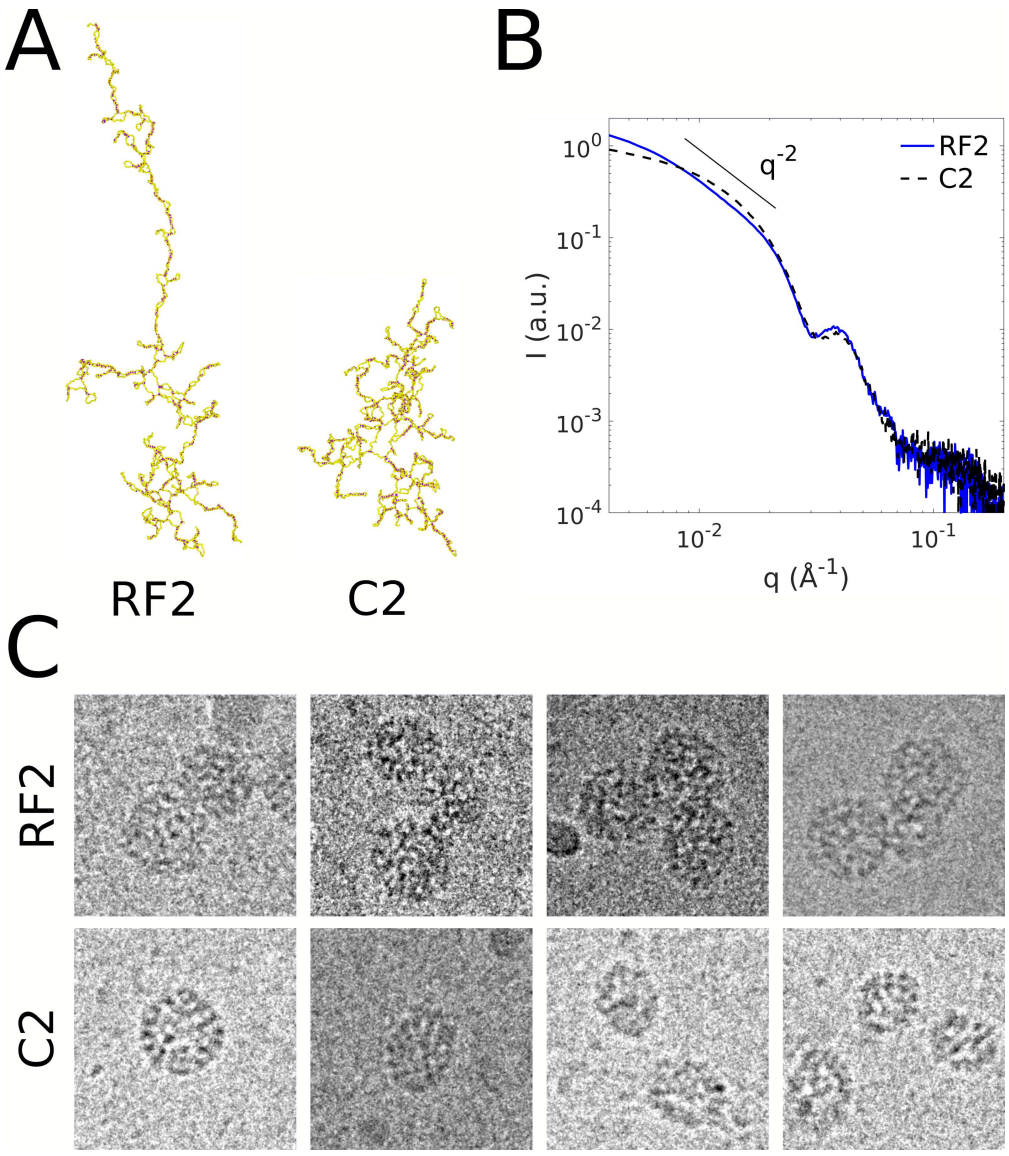

Figure 6: A. Secondary structures of RNA RF2 and C2. The structures are obtained using the ViennaRNA package (*41,42*) and visualized using OVITO. B. Rescaled SAXS intensity of structures obtained with CCMV capsid proteins at 25 μM in the one hand, and RF2 (blue solid line) or C2 (black dashed line) in the other hand, with a protein-to-RNA mass ratio of 4.5. Note that a mass ratio of 4.0 corresponds to 180 capsid proteins (T = 3) for one RNA chain. C. CryoTEM images of structures assembled with CCMV capsid proteins at 75 μM, and either RF2 or C2 with a protein-to-RNA mass ratio of 6.0. The leftmost image with C2 is a native-like T = 3 nucleocapsid. Images are all 80×80 nm.

Conclusion

While ongoing experiments continue to shed light on viral assembly, the transient nature of intermediate states and the nanometer scale of the structures present persistent challenges for direct observation. Despite the prevalence of $T = 3$ structures among single-stranded RNA viruses—and decades of experimental and computational work—no published simulations have demonstrated the spontaneous formation of a complete $T = 3$ capsid around a flexible genome. This long-standing gap has hindered our understanding of how symmetry and stability emerge during genome-guided self-assembly.

To address this, we developed a model featuring novel subunits that diffuse freely in solution and undergo conformational changes upon interaction, forming bonds and acquiring elastic properties. These elastic properties activate only after binding—an approach that significantly reduces computational cost and enables access to full assembly kinetics previously inaccessible. Using this model, we simulate the spontaneous assembly of subunits around a flexible polymer, leading to the formation of well-ordered, stable $T = 3$ and $T = 4$ capsids.

Our simulations reveal important characteristics of the assembly process: capsid proteins rapidly accumulate around the genome, driven by electrostatic interactions, often forming multiple partial capsid fragments. Although it may seem unlikely that irregular intermediates—differing in size and structure—could self-organize into a complete shell, we observe that they frequently merge into a single, closed capsid. Even under high protein concentrations, where malformed intermediates are more common, the shell often undergoes self-correction, with misassembled pentamers rearranging to restore symmetry. This transition from disorder to order unfolds over extended timescales and may require local disassembly to overcome energetic barriers.

Our findings demonstrate that conformational flexibility—introduced via an allosteric-like switching mechanism—and minimal physical interactions are sufficient to reproduce the geometric complexity of native viral capsids. The ability of capsid proteins to adopt quasi-equivalent positions and assemble into shells with different symmetries under various in vitro conditions (*21*) underscores the importance of conformational switching in mediating their interactions. The widespread ability of capsid proteins—with diverse sequences—to form icosahedral shells and package genomes under varied *in vitro* and *in vivo* conditions further suggests that virus assembly

is governed by universal physical principles. These results also highlight key limitations of earlier models, which often relied on empty shells, rigid subunits, or spherical cargos (*20, 27, 36*), and therefore did not fully capture the assembly pathways and outcomes of $T = 3$ capsids packaging a flexible genome.

In this study, we also performed in vitro experiments to examine the assembly products of CCMV coat proteins mixed with two RNAs exhibiting distinct secondary structures. As shown in Fig.6A, one RNA (C2) is highly branched, while the other (RF2) adopts a more linear configuration. The experimentally observed assembly products are shown in Fig.6C, and the corresponding simulation results are presented in Fig. 5D. Both highlight the influence of RNA secondary structure on the morphology of the final capsid structures. Specifically, when the RNA has fewer branch points, the assembly tends to yield more incomplete shells, doublets, or closed irregular structures, rather than well-formed icosahedral capsids. Our simulations qualitatively reproduce these outcomes, suggesting that the model captures key aspects of how genome architecture affects assembly. This agreement supports the relevance of our coarse-grained, dynamic framework in describing genome-guided capsid formation. Further experimental work will be needed to support a detailed quantitative comparison, which will be the focus of a future study.

Moreover, our phase diagrams (Figs. 5B and C) provide insight into how viruses may selectively package their native RNA within the crowded cytoplasmic environment, where many competing RNAs and other anionic biomolecules are present. We find that short, linear RNAs favor the formation of incomplete or defective structures, whereas long, highly branched RNAs promote the formation of regular icosahedral shells. This may explain the apparent evolutionary preference for highly branched viral genomes in many RNA viruses: the resulting assembly products tend to be statistically less defective and energetically more stable.

Armed with this model, we are now positioned to explore several previously inaccessible questions, including virus disassembly and genome release dynamics, head-to-head competition among RNAs with distinct structures (*44,45*), the role of protein N-terminal domains (*22,31,46*), the influence of alternative building block geometries (*47, 48*), and the formation of larger icosahedral shells in the presence of scaffolding proteins or heterogeneous subunits (*49–54*). It is worth noting that the presence of packaging signals would likely localize nucleation to specific sites—an effect we intend to investigate in future work.

Understanding how nucleic acid and capsid protein properties promote virus assembly could inspire the development of antiviral drugs aimed at disrupting viral replication, and it is also crucial for packaging peptides or other macromolecules for drug and gene delivery applications.

References and Notes

1. E. Hiebert, J. Bancroft, C. Bracker, The assembly in vitro of some small spherical viruses, hybrid viruses, and other nucleoproteins. Virology **34** (3), 492–508 (1968).

2. Y. Zhang, et al., Synthesis, insertion, and characterization of SARS-CoV-2 membrane protein within lipid bilayers. Science Advances **10** (9), eadm7030 (2024).

3. M. Ruszkowski, et al., Cryo-EM reconstructions of BMV-derived virus-like particles reveal assembly defects in the icosahedral lattice structure. Nanoscale **14** (8), 3224–3233 (2022).

4. L. A. Williams, A. Neophytou, R. F. Garmann, D. Chakrabarti, V. N. Manoharan, Effect of coat-protein concentration on the self-assembly of bacteriophage MS2 capsids around RNA. Nanoscale **16** (6), 3121–3132 (2024).

5. L. Lavelle, et al., Phase Diagram of Self-Assembled Viral Capsid Protein Polymorphs. J. Phys. Chem. B **113** (12), 3813–3819 (2009), doi:10.1021/jp8079765.

6. K. Kra, et al., Energetics and kinetic assembly pathways of Hepatitis B virus capsids in the presence of antivirals. ACS nano **17** (13), 12723–12733 (2023).

7. R. Zandi, B. Dragnea, A. Travesset, R. Podgornik, On virus growth and form. Physics Reports **847**, 1–102 (2020).

8. R. F. Bruinsma, G. J. Wuite, W. H. Roos, Physics of viral dynamics. Nature Reviews Physics **3** (2), 76–91 (2021).

9. K. Bond, I. B. Tsvetkova, J. C.-Y. Wang, M. F. Jarrold, B. Dragnea, Virus assembly pathways: straying away but not too far. Small **16** (51), 2004475 (2020).

10. M. Chevreuil, et al., Nonequilibrium Self-Assembly Dynamics of Icosahedral Viral Capsids Packaging Genome or Polyelectrolyte. Nat. Commun. **9** (1), 3071 (2018).

11. R. F. Garmann, M. Comas-Garcia, C. M. Knobler, W. M. Gelbart, Physical principles in the self-assembly of a simple spherical virus. Accounts of chemical research **49** (1), 48–55 (2016).

12. R. F. Garmann, et al., Single-particle studies of the effects of RNA–protein interactions on the self-assembly of RNA virus particles. Proceedings of the National Academy of Sciences **119** (39), e2206292119 (2022).

13. M. Medrano, et al., Imaging and quantitation of a succession of transient intermediates reveal the reversible self-assembly pathway of a simple icosahedral virus capsid. Journal of the American Chemical Society **138** (47), 15385–15396 (2016).

14. R. Asor, et al., Assembly reactions of hepatitis B capsid protein into capsid nanoparticles follow a narrow path through a complex reaction landscape. ACS nano **13** (7), 7610–7626 (2019).

15. J. A. Hadden, et al., All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. Elife **7**, e32478 (2018).

16. B. Qiao, M. Olvera de la Cruz, Enhanced binding of SARS-CoV-2 spike protein to receptor by distal polybasic cleavage sites. ACS nano **14** (8), 10616–10623 (2020).

17. Y.-C. Hsieh, M. Delarue, H. Orland, P. Koehl, Analyzing the geometry and dynamics of viral structures: a review of computational approaches based on alpha shape theory, normal mode analysis, and Poisson–Boltzmann theories. Viruses **15** (6), 1366 (2023).

18. J. Wagner, R. Zandi, The Robust Assembly of Small Symmetric Nanoshells. Biophys. J. **109** (5), 956–965 (2015).

19. S. Li, P. Roy, A. Travesset, R. Zandi, Why large icosahedral viruses need scaffolding proteins. Proceedings of the National Academy of Sciences **115** (43), 10971–10976 (2018).

20. S. Panahandeh, et al., How a virus circumvents energy barriers to form symmetric shells. ACS nano **14** (3), 3170–3180 (2020).

21. S. Panahandeh, S. Li, B. Dragnea, R. Zandi, Virus assembly pathways inside a host cell. ACS nano **16** (1), 317–327 (2022).

22. C. Waltmann, R. Asor, U. Raviv, M. Olvera de la Cruz, Assembly and stability of simian virus 40 polymorphs. ACS nano **14** (4), 4430–4443 (2020).

23. R. Twarock, A. Luque, Structural puzzles in virology solved with an overarching icosahedral design principle. Nature communications **10** (1), 4414 (2019).

24. F. Fatehi, R. Twarock, An interaction network approach predicts protein cage architectures in bionanotechnology. Proceedings of the National Academy of Sciences **120** (50), e2303580120 (2023).

25. J. D. Perlmutter, M. F. Hagan, Mechanisms of virus assembly. Annual review of physical chemistry **66**, 217–239 (2015).

26. W.-S. Wei, et al., Hierarchical assembly is more robust than egalitarian assembly in synthetic capsids. Proceedings of the National Academy of Sciences **121** (7), e2312775121 (2024).

27. O. M. Elrad, M. F. Hagan, Encapsulation of a polymer by an icosahedral virus. Physical biology **7** (4), 045003 (2010).

28. D. G. Angelescu, Assembled viral-like nanoparticles from elastic capsomers and polyion. The Journal of Chemical Physics **146** (13) (2017).

29. J. Mahalik, M. Muthukumar, Langevin dynamics simulation of polymer-assisted virus-like assembly. The Journal of chemical physics **136** (13) (2012).

30. J. D. Perlmutter, M. R. Perkett, M. F. Hagan, Pathways for virus assembly around nucleic acids. Journal of molecular biology **426** (18), 3148–3165 (2014).

31. J. D. Perlmutter, C. Qiao, M. F. Hagan, Viral genome structures are optimal for capsid assembly. elife **2**, e00632 (2013).

32. R. Zhang, P. Linse, Topological effects on capsomer–polyion co-assembly. The Journal of chemical physics **140** (24) (2014).

33. R. Zhang, E. Wernersson, P. Linse, Icosahedral capsid formation by capsomer subunits and a semiflexible polyion. Rsc Advances **3** (47), 25258–25267 (2013).

34. R. Zhang, P. Linse, Icosahedral capsid formation by capsomers and short polyions. The Journal of chemical physics **138** (15) (2013).

35. D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses, in Cold Spring Harbor symposia on quantitative biology (Cold Spring Harbor Laboratory Press), vol. 27 (1962), pp. 1–24.

36. O. M. Elrad, M. F. Hagan, Mechanisms of size control and polymorphism in viral capsid assembly. Nano letters **8** (11), 3850–3857 (2008).

37. O. F. Harder, S. V. Barrass, M. Drabbels, U. J. Lorenz, Fast viral dynamics revealed by microsecond time-resolved cryo-EM. Nature Communications **14** (1), 5649 (2023).

38. Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8 (2015).

39. A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. MODELLING AND SIMULATION IN MATERIALS SCIENCE AND ENGINEERING **18** (1) (2010), doi:10.1088/0965-0393/18/1/015012.

40. F. Mohajerani, et al., Multiscale Modeling of Hepatitis B Virus Capsid Assembly and Its Dimorphism. ACS Nano **16** (9), 13845–13859 (2022), doi:10.1021/acsnano.2c02119.

41. I. L. Hofacker, et al., Fast folding and comparison of RNA secondary structures. Monatshefte fur chemie **125**, 167–167 (1994).

42. S. Li, G. Erdemci-Tandogan, P. van der Schoot, R. Zandi, The effect of RNA stiffness on the self-assembly of virus particles. Journal of Physics: Condensed Matter **30** (4), 044002 (2017).

43. R. Lorenz, et al., ViennaRNA Package 2.0. Algorithms Mol. Biol. **6** (1), 26 (2011), doi: 10.1186/1748-7188-6-26.

44. M. Comas-Garcia, R. D. Cadena-Nava, A. Rao, C. M. Knobler, W. M. Gelbart, In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. Journal of virology **86** (22), 12271–12282 (2012).

45. C. Beren, L. L. Dreesens, K. N. Liu, C. M. Knobler, W. M. Gelbart, The effect of RNA secondary structure on the self-assembly of viral capsids. Biophysical journal **113** (2), 339–347 (2017).

46. S. Li, G. Erdemci-Tandogan, J. Wagner, P. Van Der Schoot, R. Zandi, Impact of a nonuniform charge distribution on virus assembly. Physical Review E **96** (2), 022401 (2017).

47. T. Chen, Z. Zhang, S. C. Glotzer, A precise packing sequence for self-assembled convex structures. Proceedings of the National Academy of Sciences **104** (3), 717–722 (2007).

48. C. M. Duque, et al., Limits of economy and fidelity for programmable assembly of size-controlled triply periodic polyhedra. Proceedings of the National Academy of Sciences **121** (18), e2315648121 (2024).

49. E. C. Dykeman, P. G. Stockley, R. Twarock, Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. Journal of molecular biology **425** (17), 3235–3249 (2013).

50. S. Tetter, et al., Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. Science **372** (6547), 1220–1224 (2021).

51. J. D. Farrell, J. Dobnikar, R. Podgornik, Role of genome topology in the stability of viral capsids. Physical Review Research **5** (1), L012040 (2023).

52. Y. Li, et al., Computational and experimental approaches to controlling bacterial microcompartment assembly. ACS Central Science **7** (4), 658–670 (2021).

53. S. Li, D. A. Matoz-Fernandez, M. Olvera de la Cruz, Effect of mechanical properties on multicomponent shell patterning. ACS nano **15** (9), 14804–14812 (2021).

54. C. Waltmann, A. Shrestha, M. Olvera de la Cruz, Patterning of multicomponent elastic shells by gaussian curvature. Physical Review E **109** (5), 054409 (2024).

55. J. A. Anderson, J. Glaser, S. C. Glotzer, HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. Computational Materials Science **173**, 109363 (2020).

56. L. Marichal, et al., Relationships between RNA topology and nucleocapsid structure in a model icosahedral virus. Biophys. J. **120** (18), 3925–3936 (2021), doi:10.1016/j.bpj.2021.08.021.

57. T. Narayanan, et al., Performance of the time-resolved ultra-small-angle X-ray scattering beamline with the Extremely Brilliant Source. J. Appl. Cryst. **55** (1), 98–111 (2022), doi: 10.1107/S1600576721012693.

Acknowledgments

Funding: S.L. and R.Z. acknowledge support from NSF DMR-2131963 and the University of California Multicampus Research Programs and Initiatives (Grant No. M21PR3267). S.L. acknowledges support from NSFC No.12204335. G.T. is grateful to L. Gargowitsch, A. Leforestier, J. Degrouard and L. Matthews, and acknowledges the European Synchrotron Radiation Facility (Grenoble, France) for allocating beamtime. The electron microscopy imaging is supported by "Investissements d'Avenir" LabEx PALM (ANR-10-LABX-0039-PALM).

Author contributions: Conceptualization: S.L., G.T., and R.Z. Methodology: S.L., G.T., Investigation: R.Z.; Data Analysis S.L., G.T., and R.Z.; Writing—original draft: S.L., G.T., R.Z.; Writing—review and editing: S.L., G.T., and R.Z.; Funding acquisition: S.L., G.T., R.Z.

Competing interests: There are no competing interests to declare.

Data and materials availability: All data supporting the conclusions of this study are included in the main text and/or Supplementary Materials. Additional data are available upon request.

Supplementary materials

Materials and Methods

Supplementary Text

Figs. S1 to S8

Movie S1-S4

Supplementary Materials for

From Disorder to Icosahedral Symmetry: How Conformation-Switching Subunits Enable RNA Virus Assembly

Siyu Li, Guillaume Tresset, Roya Zandi*

*Corresponding author. Email: royaz@ucr.edu

This PDF file includes:

Materials and Methods

Figures S1 to S8

Captions for Movies S1 to S4

Other Supplementary Materials for this manuscript:

Movies S1 to S4

Materials and Methods

Numerical simulations

To study the virus assembly pathways, we perform Molecular Dynamics (MD) simulations using the HOOMD-blue package (*55*). The system is initialized with N trimers in the rigid state (T) and one genome with a length of L. Figure S1 shows the structure of each trimer.

The interaction between the trimers is such that the ligand T_a of one trimer attracts the ligands T_b of another trimer. The interaction between ligands T_a and T_b can then be written as

$$U_{pp}^{LJ}(r) = \begin{cases} 0.5 * \epsilon_r(\sigma - r)^2 - \epsilon_{pp}, & \text{if } r < \sigma; \\ \epsilon_{pp}((\frac{\sigma}{r})^{12} - 2(\frac{\sigma}{r})^6), & \text{if } r \geq \sigma, \end{cases} \tag{S1}$$

where the repulsive part of the Lennard Jones (LJ) potential is replaced by a soft harmonic repulsive potential. Here, ϵ_{pp} represents the interaction strength, while ϵ_r stands for the repulsive strength, which we keep constant, $\epsilon_r = 100 k_B T$ in all simulations. $\sigma = R_i + R_j$ is the optimal distance between two particles, with R_i and R_j representing the radii of particles indexed i and j, respectively, and r indicating the distance between any two particles. The interactions are subject to a cutoff distance $r_{cut} = 2.4a$, with $a = 3$ nm the fundamental length of the system. All other particles in the system interact through the soft harmonic repulsive potential,

$$U^{rep}(r) = 0.5 * \epsilon_r(\sigma - r)^2, \tag{S2}$$

which is subject to a cutoff distance $r_{cut} = R_i + R_j$.

The electrostatic interaction between two particles i ans j is modeled by Debye–Hückle potential

$$U_{ij}^{DH}(r) = k_B T \; Z_i Z_j \; l_b \; \frac{e^{\kappa(R_i + R_j)}}{(1 + \kappa R_i)(1 + \kappa R_j)} \; \frac{e^{-\kappa r}}{r}, \tag{S3}$$

where the Bjerrum length $l_b = e^2\beta/4\pi\epsilon_0\epsilon$ is a measure of the dielectric constant ϵ of the solvent and is about 0.7 nm for water at room temperature. Z_i and Z_j represent the number of charges of particle i and j, respectively, while $\kappa^{-1} = 0.5a$ is the Debye screening length. In all simulations, we considered that the number of charges on trimers located on T_N partices is $Z_{T_N} = +9$, see Fig. S1.

We model RNA as a negatively charged linear chain composed of L beads whose diameter is $1.0a$ and are connected by harmonic bonds with a stretching modulus of $500 k_B T/a^2$ and equilibrium

length of $1.0a$. We assumed that the number of charges on each bead is $Z_C = -9$. The chain is the so-called Gaussian chain if we do not consider the electrostatic repulsion. We also took into account the influence of RNA's secondary structure by modeling it as a branched polymer, see Fig. 5,

We simulate the protein dynamics through Langevin integrator, with a time step $dt = 0.0005\tau$. At every 200 steps, we check the distance between any pairs of ligands T_a and T_b. When the distance between two rigid trimers is less than a cutoff distance $D_{cutI} = 0.45a$, we randomly choose one of the trimers and transforms it to an elastic subunit with a probability $p_I = 0.1$. In a different scenario where a rigid trimer moves close to an elastic trimer, we calculate the distance between the ligand T_v of the rigid trimer and the vertex (E_V) of the elastic trimer. If the distance between them is less than a cutoff distance $D_{cutA} = 1.0a$, we transfer the rigid trimer to an elastic one with a probability $P = p_E$. We consider that $p_I < p_E$, under the assumption that elastic protein subunits facilitates the transition of neighboring subunits from a rigid to an elastic state.

In situations in which an elastic trimer moves away and is not in the vicinity of any other trimers, it can switch back from elastic to a rigid one with a probablity of p_R. Figure 4 shows the different values of p_R that we have used in our simulations. There is also another mechanism for elastic trimers to become rigid subunits: we randomly choose one of the elastic trimers and transfer it back to a rigid trimer with a probability $P = p_R^n$, with n is the number of edges that the selected elastic trimer share with its neighbors. For example, for an elastic trimer on the edge sharing a bond with another trimer, the probability of transforming to a rigid trimer is p_R, while if the elastic trimer is buried inside the growing shell having three neighbors, the probability will be p_R^3.

Another crucial step in our simulations involves the merging of subunits. If any two elastic vertices E_V are close to each other, or more specifically, if the distance between them is less than $D_{cutM} = 1.0a$, we merge the two elastic vertices, resulting in the formation of a pentamer or hexamer, or the merging of partial shells into one piece.

For all simulations presented in this paper, we set the probability p_E to 1. This choice is made under the assumption that each rigid subunit gains energy of approximately Δg ranging between -1 to $-3k_BT$ when transitioning from a rigid state to an elastic state and associating to an elastic subunit. Given that the Boltzmann factor associated with this change is proportional to the exponential term $e^{-\beta\Delta g} > 1$, we consistently set p_E to 1. Correspondingly, the probability p_R is proportional to $e^{\beta\Delta g}$, which we confine within a range of $[0.1, 0.3]$. The energy gain Δg resulting from protein

conformational changes can arise from various factors. For instance, such changes may expose more hydrophobic regions, leading to an additional attraction between proteins. Alternatively, conformational changes can create the specific interactions due to morphological alterations in proteins, thereby promoting interactions between them.

We note that based on the probabilities given above, the rate constants at which the proteins undergo conformational changes can be expressed as $k_E = \nu p_E N_T$ and $k_R = \nu p_R^n N_{\hat{T}}$, see Fig. 1. Here, N_T and $N_{\hat{T}}$ represent the number of rigid and elastic trimers available for transition, respectively, and $\nu = 10\tau^{-1}$ denotes the transition frequency, with τ being the system time unit.

It is worth mentioning that all simulations presented in the paper were conducted in a protein solution with only one genome present. To replicate *in vitro* experimental conditions, we placed multiple chains within the protein solution. Since monitoring the formation of several $T = 3$ particles is computationally expensive, we focused instead on the formation of several $T = 1$ capsids. To this end, we chose the preferred dihedral angle and the size of the genome to be commensurate with $T = 1$ capsids. Using a protein concentration of $C_p = 100\mu M$, the stoichiometric ratio of protein to RNA is $200/8$, a genome length of $l = 20a$ and 200 trimers, we monitored the formation of several $T = 1$ structures, which containes 20 trimers.

As time progressed, we observed a rapid absorption of proteins onto each genome, which brought trimers closer together and initiated their transformational change into elastic structures, see Movie. S3 for the dynamics. Figure S7A records the number of proteins around each genome as a function of time, where we define each genome-trimers complex as a cluster. We observe that the formation pathway of each cluster is different, despite the fact that the shell comprises only twenty triangles. For instance, four clusters grew much slower, displaying a plateau around 10-15 trimers. The simulation snapshots at $t = 805\tau$ capture two different pathways. As shown in Fig. S7B, four clusters form closed capsids whereas the other four clusters form only half shells. One important quantity in these simulations is the stoichiometry ratio of genome to protein concentrations. If we increase the proteins concentrations, all clusters will eventually form $T = 1$ structures. These simulations confirm that the kinetic pathways of multi-shell assembly closely resemble those of a single shell.

Experimental methods

Cowpea chlorotic mottle virus (CCMV) is purified from infected cowpea leaves (*Vigna unguiculata*) (*56*). Purified virions are disassembled through overnight dialysis against 0.5 M $CaCl_2$, 1 mM EDTA, 1 mM dithiolthreitol, 0.5 mM phenylmethylsulfonyl fluoride, 50 mM Tris-HCl pH 7.5. Capsid proteins are then pelleted by ultracentrifugation at 150,000×g for 18 h and stored at 4°C in 0.5 M NaCl, 50 mM Tris-HCl pH 7.5 until use. RNA transcription is performed with a MEGAscript T7 Transcription Kit (Thermo Fisher Scientific). Freshly synthesized RNAs are purified with a MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific) and redispersed in ultraPure DNase/RNase-free distilled water (Invitrogen, Carlsbad, CA). Assembly is carried out by dialyzing a mixture of CCMV capsid proteins and RNA against 0.1 M NaCl, 50 mM Tris-HCl pH 7.5 overnight.

Small-angle X-ray scattering (SAXS) measurements are carried out at the ID02 beamline of the European Synchrotron Radiation Facilities (ESRF; Grenoble, France). The sample-to-detector distance is set to 2 m, which provides q-values ranging from 3.4×10^{-3} to 0.38 $\mathring{A}^{-1}$. The two-dimensional scattering images are radially averaged and further processed with the SAXSutilities package (*57*).

For cryotransmission electron microscopy (cryoTEM), 4 μL of solution is deposited on a glow-discharged Quantifoil R2/2 grid prior to being plunged into liquid ethane using an FEI Vitrobot. The cryofixed samples are imaged with a JEOL JEM-2010 microscope equipped with a 200-kV field emission gun and a Gatan Ultrascan 4K CCD camera.

Supplementary Figure

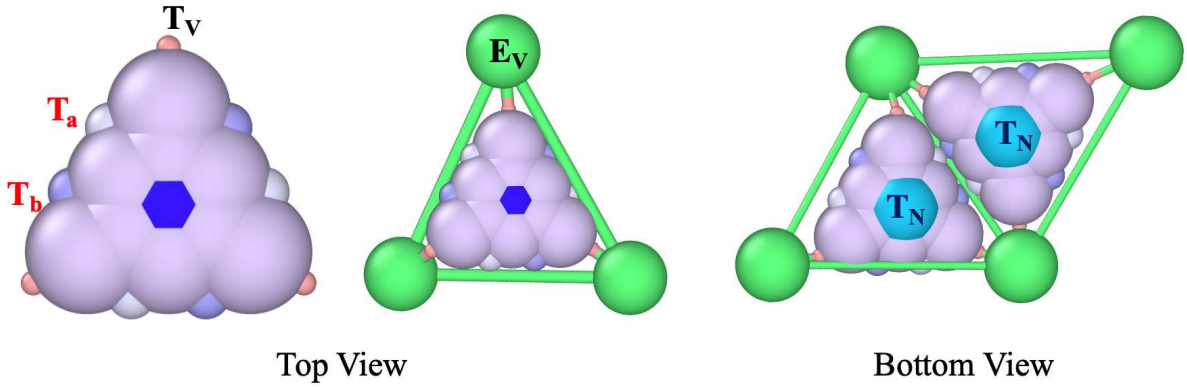

Figure S1: Illustration of the top view of the outer surface (epitope) of a rigid trimer (left) and an elastic trimer enclosed in a green triangle (middle). The structure on the right shows the inner (hypotope) surface of two elastic trimers. The rigid trimers consist of a particle denoted T_M, represented by a dark blue hexagon positioned at the trimer's center of mass, along with nine constituent particles (T_C). Additionally, the triangular subunits include three pairs of ligands (T_a, T_b) located along the trimers' edges and three vertex ligands (T_V). As a rigid subunit transitions to an elastic state, a triangle with three elastic sides and vertices (E_V) wrap around the rigid subunit. Each elastic vertex binds to the ligand T_V. The particle T_N (light blue) is located on the the inner (hypotope) surface of all trimers. The T_N particles carry nine positive charges, which interact with the negative charges on genome monomers. The radius of particles T_V, T_a, T_b, T_C, T_N, T_M and E_V are $0.1a$, $0.2a$, $0.2a$, $0.5a$, $0.5a$, $0.5a$ and $0.5a$, respectively.

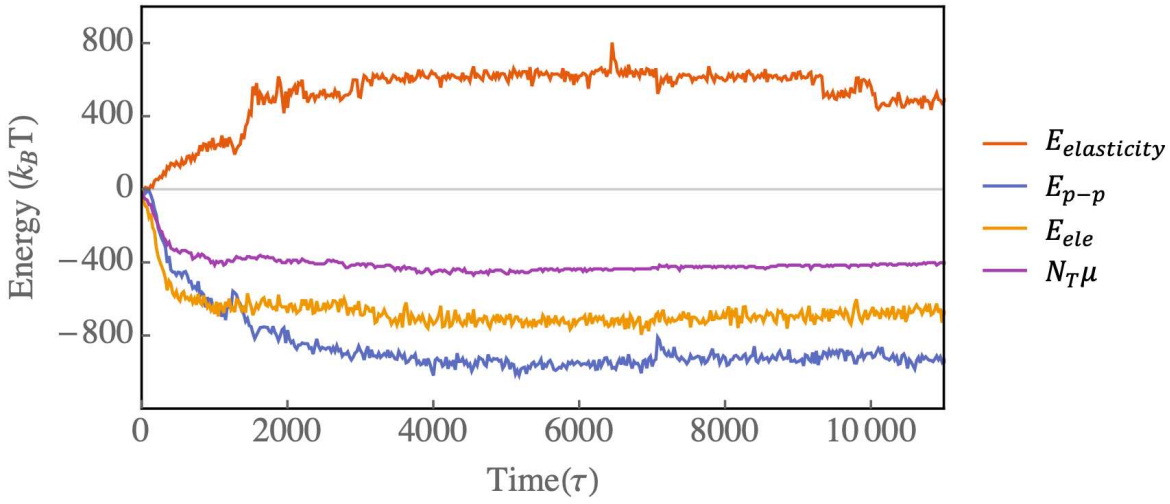

Figure S2: Plots of various energies as a function of time for the assembly pathway shown in Fig. 2. The plots correspond to the elastic energy ($E_{elasticity}$), the protein-protein attractive interaction energy (E_{p-p}), the electrostatic interaction energy (E_{ele}), and the chemical potential ($N_T\mu$). The total energy is plotted in Fig. 3.

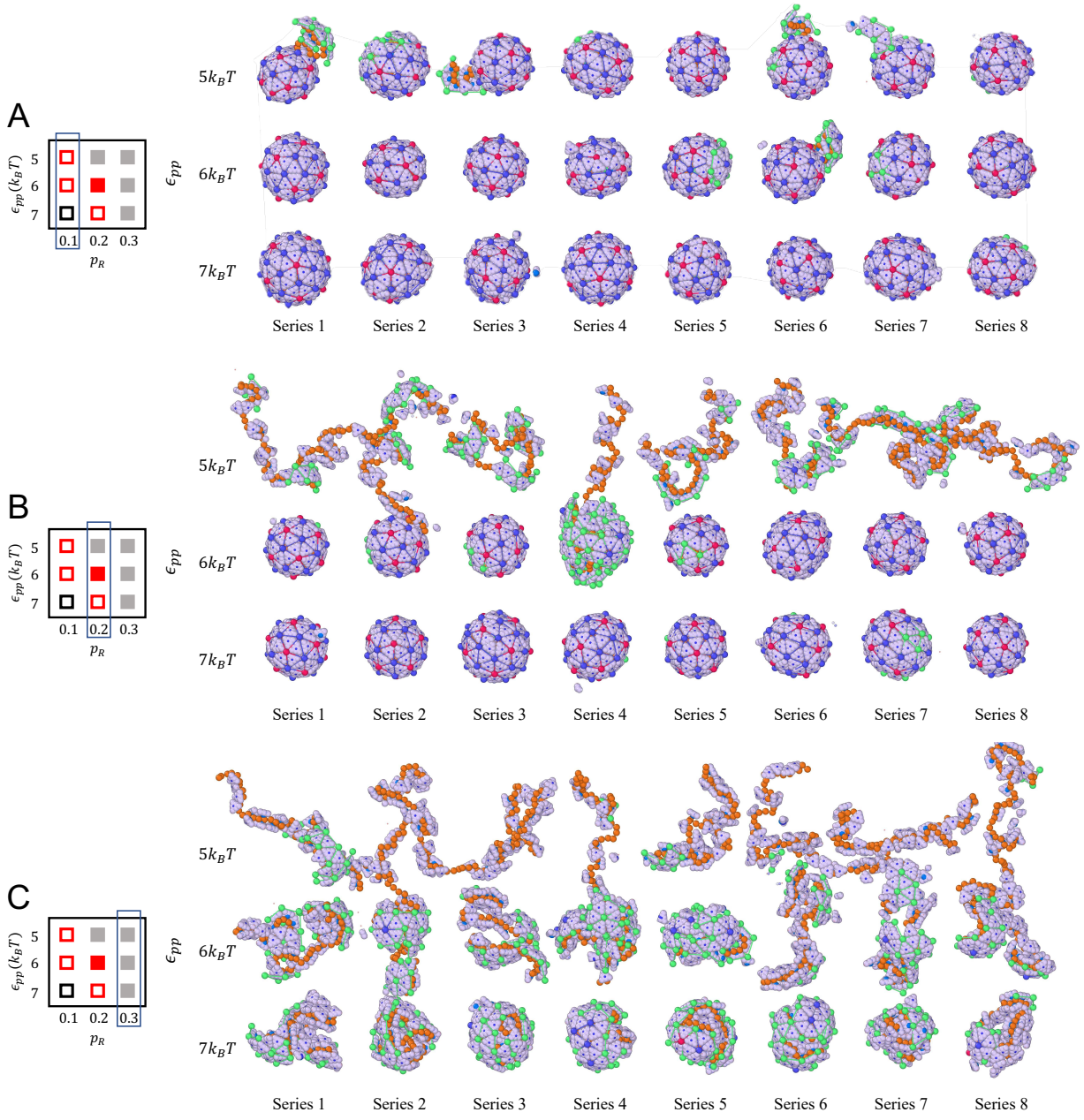

Figure S3: Snapshots of eight simulation runs observed at $t = 5000\tau$. The snapshots correspond to (A) the first column, (B) the second column, and (C) the third column of the phase diagram shown in the figure (the same as Fig. 4) with various protein-protein interactions for $p_R = 0.1$, 0.2, and 0.3. The protein concentration used is $C_p = 100\mu M$, and chain length is $80a$. The red squares in the phase diagram indicates the regions in which we observe at least one $T = 3$ before $t = 5000\tau$.

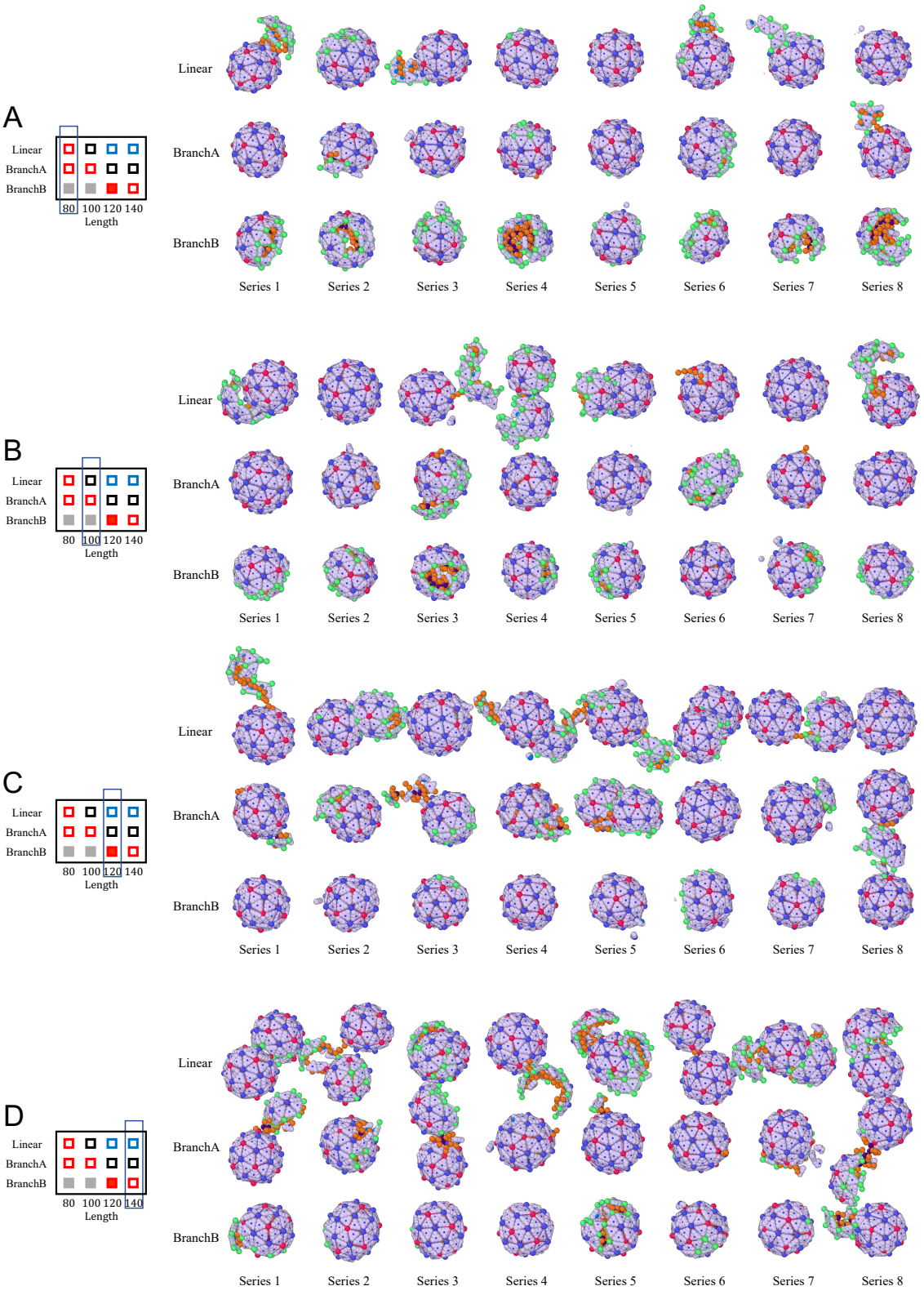

Figure S4: Snapshots of eight simulation runs observed at $t = 5000\tau$. The snapshots correspond to (A) the first column (B) the second column, (C) the third column, and (D) the forth column of the phase diagram given in Fig. 5B, with three different secondary structures of RNA. The genome lengths are $80a$, $100a$, $120a$, and $140a$. The protein concentration used is $C_p = 100\mu M$, the protein-protein interaction strength is $\epsilon_{pp} = 5k_BT$ and $p_R = 0.1$.

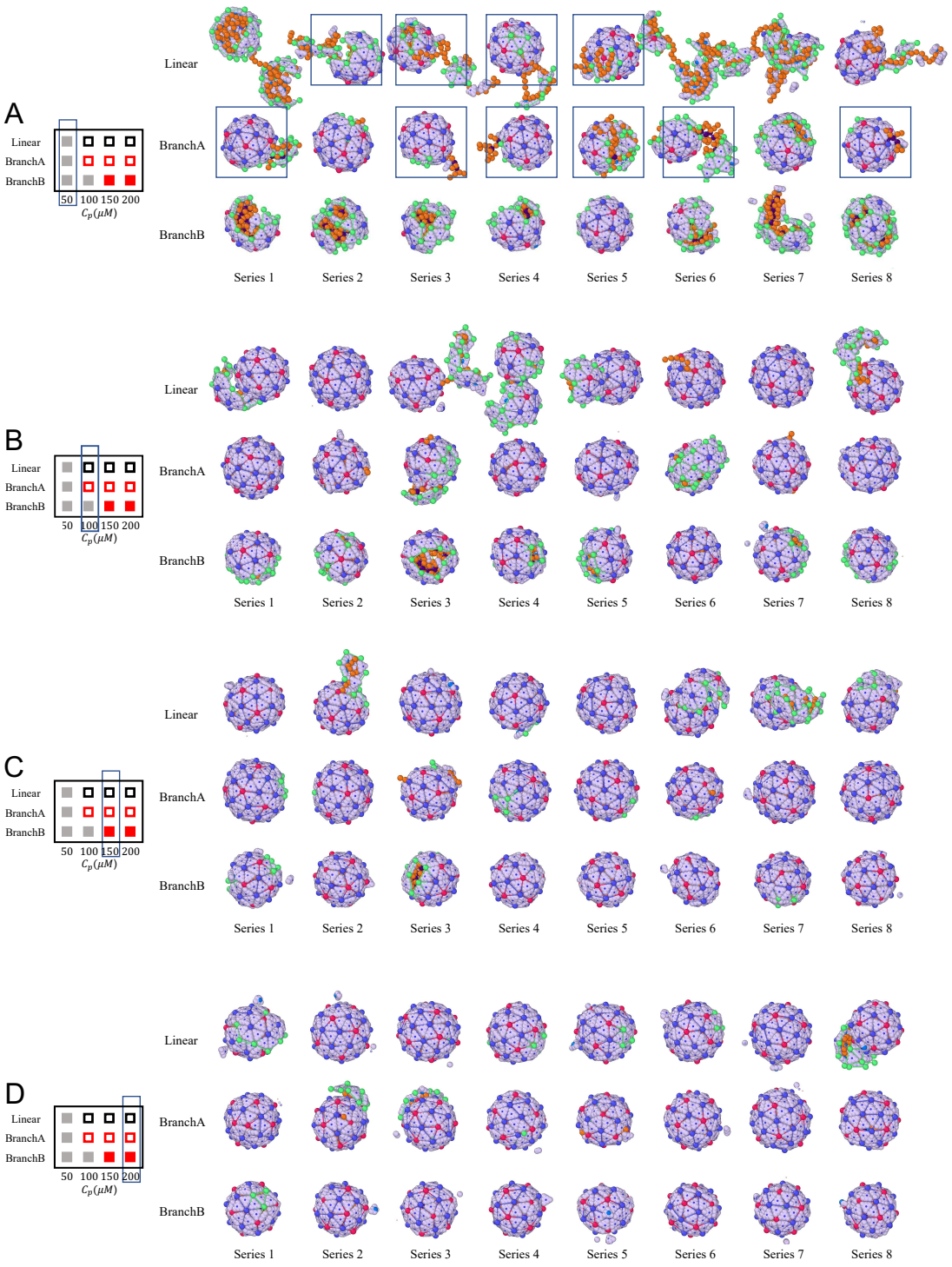

Figure S5: Snapshots of eight simulation runs observed at $t = 5000\tau$. The snapshots correspond to (A) the first column (B) the second column, (C) the third column, and (D) the forth column of phase diagram given in Fig. 5C, with three different secondary structures of RNA and various protein concentrations of $C_p = 50\mu M$, $100\mu M$, $150\mu M$, and $200\mu M$. The genome length is $L = 100a$ and the protein-protein interaction strength is $\epsilon_{pp} = 5k_BT$ and $p_R = 0.1$. Note that ten snapshots in (A) are marked by square boxes. These incomplete shells will later form closed $T = 3$ particles if we increase the simulation time. In certain cases that even if the shells are closed, the genomes are not fully packaged; they remain partially exposed outside of the shells.

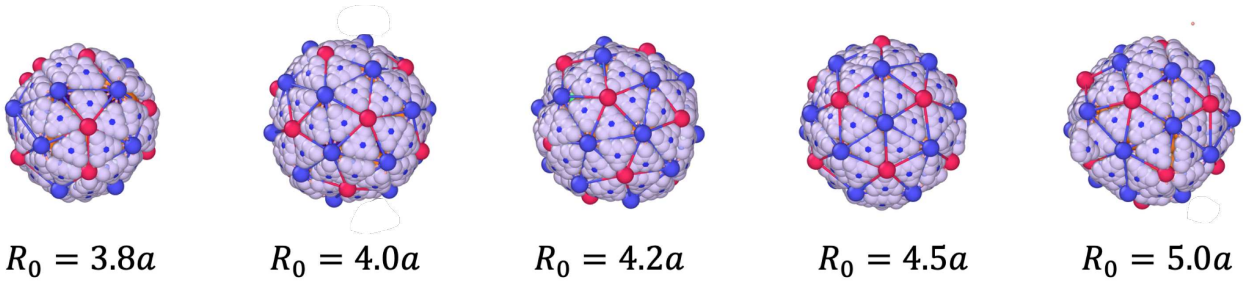

Figure S6: The final assembly products of simulations using branchB polymer at the protein concentration $C_p = 200\mu M$ for various spontaneous radius of curvature $R_0 = 3.8a$, $4.0a$, $4.2a$, $4.5a$, and $5.0a$. Except for $R_0 = 3.8a$, all the other structures have icosahedral symmetry regardless of the preferred radius of curvature. The polymer length is $100a$, the protein-protein interaction strength is $\epsilon_{pp} = 5k_BT$ and the probability of transition from the elastic to rigid state is $p_R = 0.1$.

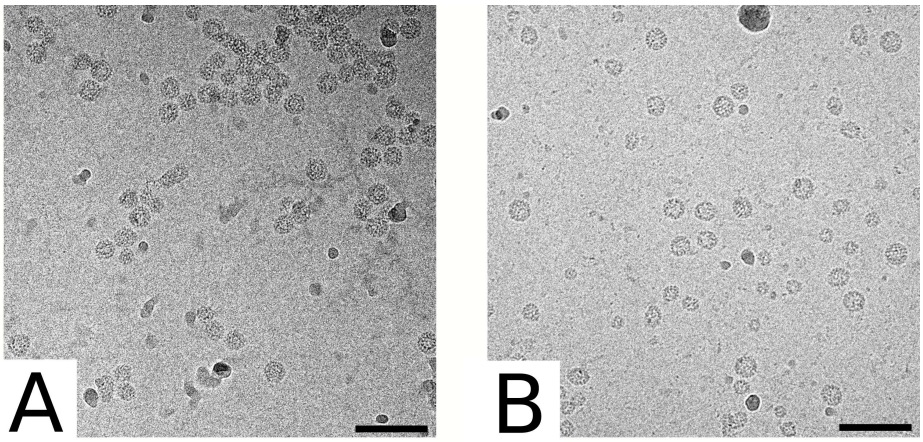

Figure S7: CryoTEM images of structures obtained with RF2 (A) and C2 (B). The concentration of CCMV capsid proteins is 75 μM and the protein-to-RNA mass ratio is 6.0. Structures with RF2 are slightly more aggregated than those with C2 and many of them are doublets or even multiplets. Note also that empty nanotubes are also encountered in both cases, possibly due to the excess of capsid proteins. Scale bar is 100 nm.

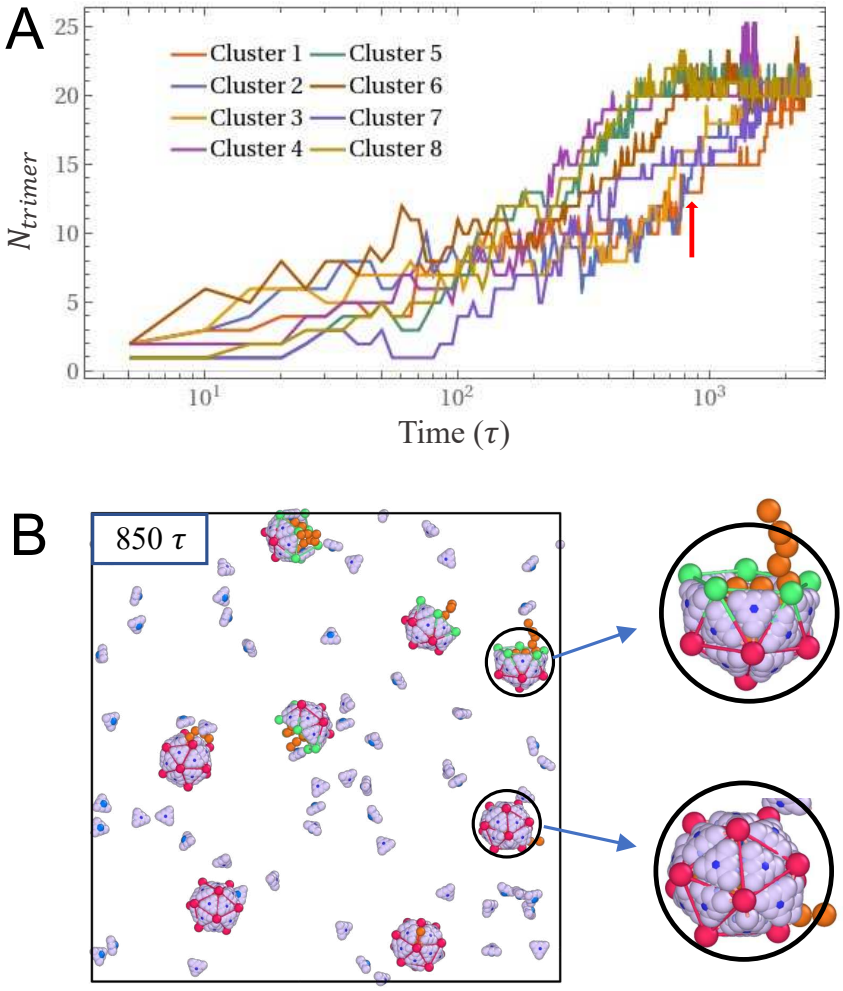

Figure S8: Assembly of several $T = 1$ structures in a mixture of short chains and trimers with a concentration of $C_p = 100\mu M$. (A) The number of trimers aggregated around each genome over time. (B) Simulation snapshot at $t = 850\tau$, where four particles form complete shells and four others form incomplete shells. The plots reveal that the simultaneous assembly of multiple $T = 1$ particles closely resembles that of a single one. It's worth noting that tracking the assembly pathways of $T = 3$ structures is challenging due to the prevalence of metastable structures. Consequently, we focused on monitoring the assembly of several $T = 1$ viruses. An important consideration is the stoichiometric ratio of protein subunits per genome, which dictates the formation of incomplete or overgrown shells.

Caption for Movie S1. Assembly pathway of a $T = 3$ capsid. Protein trimers diffuse around a linear genome of length $80a$ and self-organize from an initially disordered complex into a complete icosahedral shell. The conformational switch (allosteric) probabilities between rigid (free) and elastic (bound) trimers are $p_E = 1.0$ and $p_R = 0.1$. Protein concentration is $C_p = 100\mu$M, the spontaneous radius of curvature of elastic trimers is $R_0 = 4.2a$, and the protein–protein interaction strength is $\epsilon_{pp} = 5.0k_BT$. The transition from disorder to order occurs over an extended timescale, as many subunits must dissociate and re-associate—even after the shell appears nearly complete—for the capsid to reach its final symmetric form. These transitions involve repeated switching between free subunits and those connected to the capsid through elastic bonds formed via allosteric conformational changes.

Caption for Movie S2. Assembly pathway of a $T = 3$ capsid under faster subunit detachment conditions. Protein trimers diffuse around a linear genome of length $80a$ and self-organize into a complete icosahedral shell. Compared to Movie 1, the detachment rate of subunits from the growing shell is higher due to an increased probability of allosteric switching from the elastic (bound) to the rigid (free) state ($p_R = 0.2$). To compensate for this increased detachment, the hydrophobic protein–protein interaction strength is raised to $\epsilon_{pp} = 6.0k_BT$. Together, these changes accelerate the assembly process and increase its efficiency. The probability of transitioning from rigid to elastic remains $p_E = 1.0$, and the spontaneous radius of curvature of elastic trimers is $R_0 = 4.2a$. Protein concentration is $C_p = 100\mu$M. As in Movie 1, the capsid assembles through repeated association and dissociation events, mediated by allosteric conformational changes and the formation of elastic bonds.

Caption for Movie S3. Assembly pathway of a $T = 4$ capsid. Protein trimers diffuse around a linear genome of length $80a$ and self-organize into a complete $T = 4$ icosahedral shell. The conformational switch (allosteric) probabilities between rigid (free) and elastic (bound) trimers are $p_E = 1.0$ and $p_R = 0.1$. Protein concentration is $C_p = 100\mu$M, the spontaneous radius of curvature of elastic trimers is $R_0 = 4.2a$, and the protein–protein interaction strength is $\epsilon_{pp} = 7.0k_BT$. Assembling a $T = 4$ shell is more challenging due to its larger size, and the system requires a longer timescale to transition from a disordered complex to a fully

symmetric structure. The capsid forms through a dynamic process of subunit association and dissociation, mediated by allosteric conformational switching and elastic bond formation.

Caption for Movie S4. Simultaneous formation of multiple $T = 1$ capsids. Protein trimers assemble around eight linear genomes, each of length $20a$, forming eight complete $T = 1$ icosahedral shells. The conformational switch (allosteric) probabilities between rigid (free) and elastic (bound) trimers are $p_E = 1.0$ and $p_R = 0.1$. Protein concentration is $C_p = 100\mu$M, the spontaneous radius of curvature of elastic trimers is $R_0 = 3.0a$, and the protein–protein interaction strength is $\epsilon_{pp} = 5.0k_BT$. The smaller genome length and higher curvature facilitate the rapid and parallel assembly of multiple capsids through dynamic association and allosterically driven conformational switching.