

Temporal control by cofactors prevents kinetic trapping in retroviral Gag lattice assembly

Yian Qian,¹ Daniel Evans,¹ Bhavya Mishra,² Yiben Fu,¹ Zixiu Hugh Liu,¹ Sikao Guo,¹ and Margaret E. Johnson^{1,*}

¹TC Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland and ²Department of Physics, and Center for Cellular and Biomolecular Machines, University of California, Merced, California

ABSTRACT For retroviruses like HIV to proliferate, they must form virions shaped by the self-assembly of Gag polyproteins into a rigid lattice. This immature Gag lattice has been structurally characterized and reconstituted *in vitro*, revealing the sensitivity of lattice assembly to multiple cofactors. Due to this sensitivity, the energetic criterion for forming stable lattices is unknown, as are their corresponding rates. Here, we use a reaction-diffusion model designed from the cryo-ET structure of the immature Gag lattice to map a phase diagram of assembly outcomes controlled by experimentally constrained rates and free energies, over experimentally relevant timescales. We find that productive assembly of complete lattices in bulk solution is extraordinarily difficult due to the large size of this 3700 monomer complex. Multiple Gag lattices nucleate before growth can complete, resulting in loss of free monomers and frequent kinetic trapping. We therefore derive a time-dependent protocol to titrate or “activate” the Gag monomers slowly within the solution volume, mimicking the biological roles of cofactors. This general strategy works remarkably well, yielding productive growth of self-assembled lattices for multiple interaction strengths and binding rates. By comparing to the *in vitro* assembly kinetics, we can estimate bounds on rates of Gag binding to Gag and the cellular cofactor IP6. Our results show that Gag binding to IP6 can provide the additional time delay necessary to support smooth growth of the immature lattice with relatively fast assembly kinetics, mostly avoiding kinetic traps. Our work provides a foundation for predicting and disrupting formation of the immature Gag lattice via targeting specific protein-protein binding interactions.

SIGNIFICANCE Self-assembly of the HIV virion within cells is controlled by interactions of the viral Gag protein with cofactors such as RNA. The kinetics of this assembly must be tuned to prevent formation of incomplete and unproductive Gag lattices. Our work here combines theory, simulations, and experimental kinetic data to build a predictive model of how activation of the Gag proteins by interactions with cofactors can ensure robust, productive assembly. We provide theoretical guidelines that can be used to both extract key kinetic parameters from bulk experimental studies, and to design experimentally testable conditions to enhance or inhibit self-assembly. These results thus provide a quantitative model for characterizing how cofactor binding can control the kinetics of viral assembly.

INTRODUCTION

All retroviruses, including HIV, must form new virions that can bud out of the plasma membrane of the infected host cell (1). These virions are comprised of the retroviral polyprotein Gag, genomic RNA, and additional cofactors (1). Gag is the primary structural component of the virions, initially assembling an immature lattice that is bound to the membrane and forms a trihexagonal organization as revealed by cryoelectron tomography (cryo-ET) (2,3). Once the

virion has budded, the Gag polyprotein is cleaved by its attached proteases (4) and reassembles into the mature viral capsid (5). The stability of the immature Gag lattice appears tuned to ensure successful assembly while also ensuring the remodeling necessary to transform from the immature lattice to the infectious mature capsid (6,7). Indeed, maturation inhibitors are a promising strategy for antiviral drugs that function by overstabilizing the immature lattice (7–9). Assembly of the immature lattice can be reconstituted *in vitro* (10,11), but because the immature lattice does not assemble from the Gag monomers alone (10), it is not known how the energetics and cooperativity of Gag-Gag contacts, Gag contacts with cofactors, or their timescales of binding drive stable lattice formation. Here, we develop a coarse-grained model of Gag assembly to quantify

Submitted February 18, 2023, and accepted for publication June 27, 2023.

*Correspondence: margaret.johnson@jhu.edu

Yian Qian and Daniel Evans contributed equally to this work.

Editor: Margaret Shun Cheung.

<https://doi.org/10.1016/j.bpj.2023.06.021>

© 2023 Biophysical Society.

assembly pathways as a function of Gag-Gag binding rates and stability, ultimately designing time-dependent protocols that allow us to robustly assemble immature lattices with varying kinetics. These models on their own and with comparison to *in vitro* kinetics (11) predict how time-dependent control of cofactor binding can induce (or disrupt) stable assembly of the immature Gag lattice.

A primary challenge in understanding and predicting the assembly of the immature Gag lattice is its dependence and sensitivity to a range of cofactors. The minimal conditions to assemble the virus-like lattice in solution *in vitro* requires Gag along with at least one negatively charged molecule such as RNA or the cellular small molecule inositol hexakisphosphate (IP6) (12–14). Gag does form stable dimers on its own, which are further strengthened with cofactors present (15). Thus, it is the formation of the higher-order contacts in the immature lattice that must be distinctly “switched on” by interactions with cofactors, at least in part by inducing conformational changes into the Gag proteins (10). Combining RNA and IP6 together works synergistically to promote and accelerate immature Gag lattice assembly, indicating that they have somewhat complementary roles in stabilizing the lattice (11). With just RNA or just IP6, assembly takes 1–2 h at 50 mM Gag, whereas with both, it can proceed in seconds (11). Indeed, IP6 promotes formation of infectious Gag virions *in vivo* by stabilizing the immature lattice (6). IP6 coordinates with the hexamer contacts in the Gag lattice in a 1:6 stoichiometry, stabilizing Gag into an assembly-competent form for the immature Gag lattice (15). The immature Gag lattice also contains another set of interfaces that form a trimeric cycle that can provide additional stability when brought together via the dimer and hexamer assembly, although it is likely that these smaller interfaces are relatively weak (16). Through our computational models, we can reach the seconds to minutes timescales of cofactor-mediated assembly, where we observe robust growth and a high yield of virions forms if large enough concentrations of RNA and IP6 are present (11). We can thus directly test how the relative strengths and speeds of the different Gag-Gag interactions control lattice assembly.

Another challenge with the immature Gag lattice is its size; a completed spherical lattice contains 4000 monomers, which is an order of magnitude higher than most computational models of self-assembly. There are two major issues that arise with this larger size. The first is simply that the models become more computationally expensive to characterize across varying parameters compared with completed assemblies that contain 12 (17,18), 60, or even 100s of subunits (19,20). For smaller systems, systems of rate equations can be constructed to efficiently characterize phase diagrams and assembly pathways as thermodynamic and kinetic parameters are varied (18). Spatial simulations using Brownian dynamics simulations can also be relatively efficient for characterizing assembly pathways in systems with 100s of subunits (19,21), but 1000s of subunits are rare (22). Nonspatial

simulations of large HIV-like assemblies must limit the types of assembly pathways followed (23). The second major issue, however, is that with more subunits required to complete the assembly the rate of nucleating new lattice structures must be correspondingly slowed to allow elongation of existing nuclei to completion, where a “nuclei” can in principle be as small as a dimer (24). Otherwise, assemblies become kinetically trapped in intermediates that cannot readily combine to form larger structures as monomers are fully depleted or starved (25). The parameter regimes that can support assembly will be significantly compressed relative to assemblies with fewer subunits. Hence, the Gag lattice is intrinsically primed to be prone to kinetic traps due to its size.

Kinetic trapping, which can dramatically slow the formation of productive and complete equilibrium assemblies, sometimes beyond experimentally relevant timescales, is a consequence of the nonequilibrium nature of the self-assembly process. The initial conditions of unbound subunits can follow pathways to long-lived metastable intermediates despite the existence of a thermodynamically favorable steady state of complete structures. Thus, unlike the quasi-equilibrium classical nucleation theory (26), these intermediates can fail (at least temporarily) to complete stable growth after formation of a critical nucleus (a free energy maxima). We note here we do not extract a precise size of a critical nucleus given that we study multiple interaction free energies (which would have different critical nuclei), or demarcate a specific “nucleation” and “growth” phase that would exhibit distinct kinetics (27), but we do define a consistent practical cutoff of 18 monomers in nucleated structures for analysis (materials and methods).

For such a large lattice, how does HIV avoid kinetic traps? *In vivo*, we identify two main strategies that can help avoid kinetic traps and productively assemble virions by effectively controlling initial conditions. First, subunits do not appear in bulk in the cytoplasm but are produced at a rate that slowly increases their concentration in time and space (28). Second, as detailed above, cofactors can “activate” the Gag monomers such that lattices only nucleate and thus grow when they have bound RNA. Both these mechanisms introduce additional timescales that can reduce multiple nucleations and instead promote growth of nascent viruses. Theory and modeling have shown how, although kinetic traps for smaller assemblies can be reduced by optimizing energetic parameters to moderately weak strengths (19,25), similar strategies to those used by the cell that introduce new timescales can significantly improve yield. Cooperativity during growth that disfavors formation of nascent structures by modulating binding sites helps avoid traps (29), as can variable addition of monomers (30) and allosteric activation (31). *In vitro* studies of HIV-1 immature lattice assembly in solution showed how the stoichiometry of the cofactors relative to the Gag monomers tune the kinetics of lattice assembly, indicating that cofactor binding can act as a rate-limiting timescale in assembly (11). In our work

here, we therefore use the theory of diffusion-influenced reactions (32–34) to derive a timescale that could effectively represent these same biologically relevant processes used to suppress trapping, by directly controlling the concentrations of monomers with time. Our simulation results illustrate how even this single additional timescale can effectively eliminate the trapped intermediates to promote successful assembly.

A key advantage of the structure-resolved reaction-diffusion modeling approach we use here is that the timescales and assembly yield are controlled by biochemically measurable rates and free energies, and thus the predicted timescales we define compare directly with experiments. Recently, coarse-grained models of the immature HIV lattice have shown how IP6 stabilizes lattice assembly and speeds up binding with more molecular details, in ways that may promote defects in the structures (35). These coarse-grained models built on earlier work characterizing the specificity of Gag contacts needed for proper assembly (36), and the role of RNA and membrane binding in stabilizing Gag assembly (37). These models successfully captured morphological variations in assembled structures and explicit cofactor interactions but are not able to map the kinetics of binding and assembly directly to experimental timescales, and the strength of protein-protein interactions can be similarly difficult to map to experimental K_D values. Our model is thus complementary, producing assembled structures that are similarly derived from experiment and built from the monomeric building blocks, while providing direct observations of tunable assembly kinetics over longer (minutes) timescales. We are thus able to estimate rates for Gag binding to the cofactor IP6 by comparison with the in vitro kinetic data (11), and set bounds on the rates of Gag-Gag binding interactions.

In this paper, we first describe the construction of our computational model of Gag self-assembly from cryo-ET structures and validate the kinetics and equilibrium of our simulations with comparisons to theory. We test a range of binding free energies and rate constants for the distinct Gag-Gag binding interactions of dimer, hexamer, and trimer contacts. We identify regimes where assembly could be eliminated, where a two-phase equilibrium of dilute monomers and partial assemblies formed, and where kinetic trapping onset. Kinetic trapping was predominant, with our structural analysis quantifying the frequent irregularity of the ensuing lattice growth. To mimic the time-dependent activation of Gag by cofactors, we used a titration rate to slowly increase the Gag concentration, finding that, when the rate was slow enough, we produced robust and regular lattice growth. We therefore derived a titration rate dependent on the underlying Gag assembly kinetics, showing that it is remarkably successful in promoting productive lattice assembly. Finally, we connected our model to in vitro light-scattering kinetics of Gag assembly driven by cofactors RNA and IP6. We were able to estimate bounds on binding rates between Gag

and IP6 and Gag-Gag using relatively simple theoretical arguments and our derived titration rate. We conclude by discussing the future extensions and applications of this integrative modeling approach.

MATERIALS AND METHODS

Model construction

Our model consists of coarse-grained representations of Gag protein monomers derived from the cryo-ET structure (16) of the immature Gag lattice (Fig. 1 A). Each Gag is a rigid body with a center of mass point and five distinct binding sites (Fig. 1 B). The center-to-center distance between two hexamer substructures is 7.7 nm, the final sphere radius $R \approx 50$ nm, and the complete lattice contains 3700 Gag proteins. There are three types of binding interactions: the homo-dimer interaction, the hexamer interaction mediated by the two distinct hexamer sites, and the trimer interaction mediated by the two distinct trimer sites. All three contacts were identified and characterized in the experimental structural study (16), and here we identify the interfaces for each of these interactions using a 3.5 Å cutoff between atoms in contact between two monomers (supporting methods). When two molecules bind via these specific interaction sites, they adopt a predefined orientation that ensures assembly into the experimental cryo-ET structure. To derive our model of Gag-Gag interactions from the cryo-ET structure, we must correct for the small variations of the 18 monomer positions within the lattice. Without corrections, our monomers would assemble spherical fragments with distinct curvatures and thus significant defects due to slight changes of orientations between hexamer, trimer, and dimer contacts. Thus, we symmetrized the Gag lattice to ensure a single curvature across all contacts. See supporting material for mathematical details.

A trihexagonal lattice cannot perfectly tile a spherical surface; a sphere requires pentagonal inclusions as well. However, since the hexagons are small relative to the size of the sphere, the defects that emerge due to some imperfectly aligned contacts are minimal. We allow imperfectly aligned proteins to still form bonds if they are within a cutoff distance 10% longer than the binding radius s . In this way, these contacts still contribute to stabilizing the lattice, which also mimics the inherent flexibility of molecules themselves that is not captured by the rigid orientations imposed here.

Reaction-diffusion simulations

All simulations are performed using the NERDSS (nonequilibrium reaction diffusion self-assembly simulator) software. NERDSS solves a particle-based and structure resolved reaction-diffusion model using the free-propagator reweighting algorithm (38). The method is derived from Smoluchowski's model for reactive collisions between diffusing particles and has been rigorously tested to produce accurate binding kinetics and equilibria in 3D solution (39) and with the addition of rotational motion and assembly (40). The software and executable models are available open source at github.com/mjohn218/NERDSS, and all analysis described below is available as Python code at github.com/mjohn218/ioNERDSS.

We briefly outline here how the simulations work. Step 1 Initialization. Copies of molecules are placed randomly inside a rectangular box, followed by a steric overlap check for all reactive binding site pairs to ensure they are not at a distance less than the binding radius s for the reaction. Molecules that are titrated into the system (zeroth-order reaction) are also placed randomly within the box without steric overlap. Step 2 Reactions. Reaction and diffusion are treated separately, and one molecule can only participate in one of them within each time step. We evaluate reactions first by calculating the reaction probability of zeroth- (creation), first- (dissociation), and second-order (bimolecular) reactions. Dissociation reactions are modeled as Poisson processes, occurring with probability $p_{\text{dissoc}} \Delta t \ll 1$

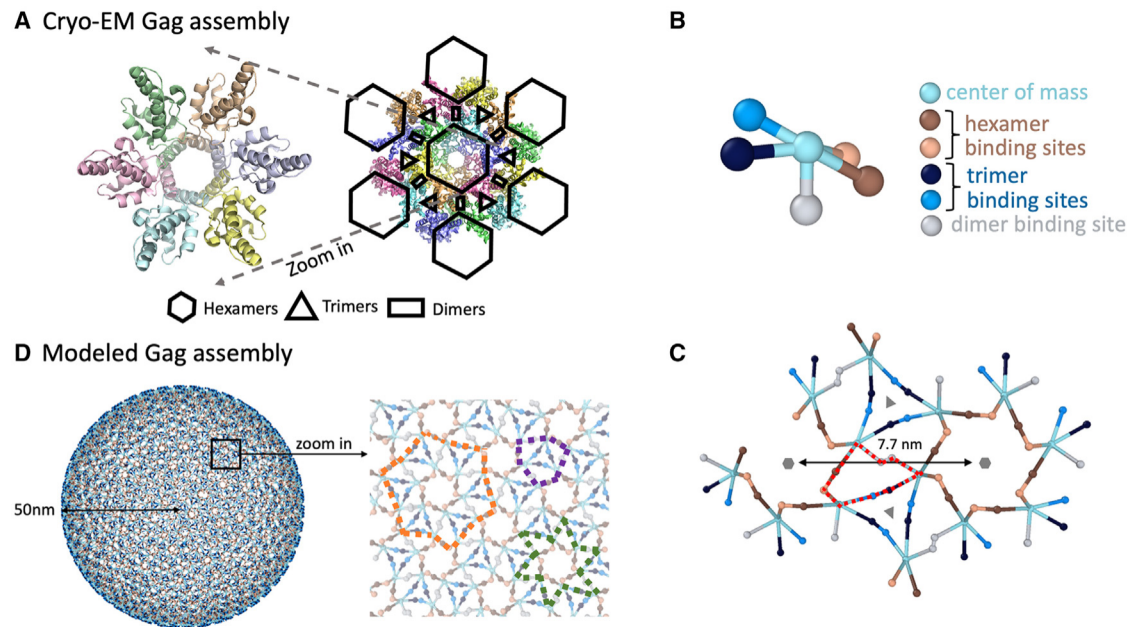


FIGURE 1 Coarse-grained model of Gag monomers and their binding interactions. (A) The cryo-ET immature Gag lattice from PDB: 5L93. On the left, six distinct Gag monomers form a hexameric ring with each Gag contacting its two neighbors. The right image includes 18 monomers demonstrating how the larger lattice assembles from hexagons through dimer contacts, with trimer contacts adding additional stabilizing interfaces. (B) Each coarse-grained Gag monomer has a center of mass site and five distinct interfaces that mediate the three types of interactions (materials and methods and supporting material). The hexamer and trimer interactions both involve a front-to-back contact (like actin polymerization) between distinct interfaces. (C) Geometry of the coarse-grained Gag monomers assembled into the lattice, showing all three types of interactions between 13 monomers. In addition to the stabilizing cycles of the hexamer and trimer, cycles are also formed involving all three types of interactions (red dashed lines). (D) The spherical Gag lattice was assembled from a population of monomers, requiring 3700 Gags to close. Some defects in the lattice are unavoidable as a rigid hexagonal lattice cannot perfectly tile a spherical surface. From the zoomed-in structure, extra cycles formed through any two interactions can be observed. The purple line follows dimer and hexamer interactions, the orange one trimer and hexamer interactions, the green one dimer and trimer interactions. To see this figure in color, go online.

1 $\exp \delta k_b D_t \Delta t$, with intrinsic off-rate k_b . All bimolecular reactions are reversible unless explicitly stated. The probability of bimolecular association events $p_{\text{assoc}} \delta D_t \Delta t$ within timestep Δt depends on the initial separation r_0 of reactant sites A and B, microscopic association rate k_a , the total diffusion coefficient of reactants $D_{\text{tot}} \propto D_A + D_B$, and binding radius s . The exact value is calculated from the Green's function solution of Smoluchowski's model, with reweighting applied due to the simple Brownian updates used for particle displacements (39). The binding probability is independent of the orientation of the two sites, it depends on only the separation. If an association reaction is accepted via comparison of the reaction probability with a uniform random number, the two molecules will be translated and rotated into the user-specified binding orientation. We reject association events between multiprotein complexes that result in steric overlap between protein monomers. We also reject association events that produce unphysically large rotational reorientations of complexes to "snap" them into the proper binding orientation. This is controlled by a scalar (scaleMaxDisplace) that multiplies the expected displacement due to translational and rotational diffusion to define a cutoff (see Table 1). Step 3 Propagation. If the molecules within a complex do not undergo a reaction, the rigid complex diffuses both translationally and rotationally along three orthogonal axes x , y , and z . Translational diffusion along any axis is simply a Brownian motion $\Delta x \propto \sqrt{D_t \Delta t}$, where the displacement Δx follows a Gaussian distribution with mean zero and standard deviation $\sqrt{2 D_t \Delta t}$. D_t is the translational diffusion constant. Similarly, rotational diffusion characterizes molecular rotation around a specific global axis, $\Delta \theta \propto \sqrt{D_r \Delta t}$, where $\Delta \theta$ follows Gaussian distribution with mean zero and standard deviation $\sqrt{2 D_r \Delta t}$. D_r is the rotational diffusion constant. We define the diffusion coefficients of Gag monomers (Table 1) using the Einstein-Stokes equations, with $D_t \propto 10 \text{ nm}^2/\text{ms}$ and $D_r \propto 0.02 \text{ rad}^2/\text{ms}$, which is true for a hydrodynamic radius of 20 nm at water viscos-

ity. We also tested faster values implying a more compact radius at the same viscosity, or a monomer $D_t \propto 10$ also mimics the smaller radius with the 3 decrease in viscosity occurring in the cytoplasm. Crowding in the cytoplasm can affect rate constants, but the effect is small when rates are not diffusion controlled, as is the case here (41). The diffusion coefficients of complexes slow as their size increases, consistent with the assumption that the hydrodynamic radius is additive (40).

Simulation conditions

We perform simulations in a rectangular volume with reflective boundary conditions at the walls. Our simulations all reach a final concentration of 50 mM of Gag monomers, consistent with experiments on immature Gag lattice assembly in solution (11). We use a box with volume 405^3 nm^3 that contains 2000 Gag proteins maximally, because it is computationally more efficient and still reports on the kinetic challenges to assembling large lattices. In mimicking activation by cofactors can ensure slow nucleation and fast growth of a single lattice we also use a larger volume of 510^3 nm^3 that contains 3700 Gag proteins and thus allows the formation of a complete spherical lattice. In addition to "bulk" simulations with all Gag monomers present at time zero, we also perform titration simulations where initially 10 monomers are present, and remaining proteins stochastically enter the system with rate k_c (see Table 1).

Energetic and kinetic parameters

All parameters used in simulations are listed in Table 1. The stability of the Gag dimerization contact is constrained by experiment (15), which found

TABLE 1 Simulation, energetic and kinetic parameters for Gag interactions

	Parameter name	Description	Value
Simulation parameter	Dt	time step	0.1; 0.5 ms
	box size	the size of the simulation system	405 ³ ; 510 ³ nm ³
Diffusion coefficients	D _t	molecule's translational diffusion constant	10; 50 mm ² =s
	D _r	molecule's rotational diffusion constant	0.01; 0.02; 0.05 rad ² =ms
Binding parameters	dimeric k _{a,dim}	3D microscopic association rate	0.06; 0.6; 6 mM ⁻¹ s ⁻¹
	hexameric k _{a,hex}		0.006; 0.06; 0.6; 6; 12; 60 mM ⁻¹ s ⁻¹
	trimeric k _{a,trim}	microscopic dissociation rate	0; 6.02 10 ⁵ mM ⁻¹ s ⁻¹ 1 s ⁻¹
	dimeric k _{b,dim}		10; 20; 100; 10 ³ ; 10 ⁴ s ⁻¹ 1; 1000 s ⁻¹
	hexameric k _{b,hex}		11.0, 13.3, 15.6 k _B T
	trimeric k _{b,trim}		1.8, 4.1, 6.4, 8.7, 11.0, 13.3, 15.6 k _B T
	dimeric DG _{dim}	Gibbs free energy DG	0, 4.1 k _B T
	hexameric DG _{hex}		30
	trimeric DG _{trim}	association events that result in shifts of binding sites on either component larger than scaleMaxDisplace	0.06; 0.3; 1; 60 mM=s
	scaleMaxDisplace		
Titration rate	k _c	molecule creation (titration) rate	

^aD_{tot} captures effective diffusion coefficient due to translational and rotational diffusion.

DG of 12 k_BT but additional stabilization occurred after cofactor interaction. However, other energetic and kinetic parameters are not known. We thus choose a biologically realistic range of rate (10³–10² mM⁻¹s⁻¹) and free energy values, keeping the trimer contact weakest as indicated by previous work (16). The rates do not change throughout the simulations, meaning there is no explicit cooperativity or size dependence of binding (27). The stability or free energy DG of each pairwise interaction is determined by the binding (k_ap and unbinding (k_bp rates using standard relationships, K_D = $\frac{k_{off}}{k_{on}}$ = $\frac{k_b}{k_a}$ = $\frac{c_0}{c_0} \exp \frac{DG}{k_B T}$ where c₀ is the standard state 1 M, DG = G_{bound} - G_{unbound}, and k_{on} is the corresponding macroscopic rate defined in 3D by

$$k_{on} \propto \frac{1}{k_a} \propto \frac{1}{4\pi p D} \quad (1)$$

(see, e.g., (39)). NERDSS allows for a free energy strain penalty upon formation of closed cycles, but here we apply no penalty. This means when two bonds form to close a cycle, the stability is the sum over the bond free energies.

Fitting of association kinetics to analytical theory

To quantify how the formation of higher-order contacts impacts the assembly kinetics, we analyzed the kinetics of pairwise bond formation of both dimers and hexamers. If the bonds formed independent of any lattice, their kinetics would match up with theory from rate equations, given that our components are initially well mixed and in 3D, which minimizes any spatial effects. These analytical solutions are known, and for completeness the kinetics of the monomer species A ↔ B undergoing the reversible reaction A ↔ B with rates k_{on} and k_{off} is given by:

$$A \leftrightarrow B \propto \frac{1}{e^{r_1 t} - r_1 e^{r_2 t}} \quad (2)$$

where $r_{1,2} = \frac{q_1 + q_2}{2} \pm \sqrt{\frac{q_1^2 + q_2^2}{4} - k_{off} A_0}$, and initially $A \propto 0$, $B \propto A_0$, $C \propto 0$. For distinct reactants, q₁ and q₂ change slightly (see, e.g., (42)).

The effect of diffusion is captured in the macroscopic rate constant using Eq. 1. Given this equation, we can then fit our simulated data on bond formation (Fig. 2), where in the case of dimer association we treat only dissociation as a fit parameter, k_{off,dim}. For slow hexamer interactions, we recover k_{off,hex} = k_{off,dim}, validating that the kinetics is correct as designed and dimer

bonds form fully independently at short enough times. As hexamers start to form more quickly, or as trimers start to form, we see slowed dissociation of dimers, such that k_{off,dim} < k_{off,hex}. We perform the same procedure for hexamer bond-forming kinetics, except that the sites are distinct (A ↔ B ↔ C). For the hexamers, both the association and dissociation kinetics are influenced by dimer contacts, and thus we must treat both k_{off,hex} and k_{on,hex} as fit parameters. We therefore then compare the ratio of these rates to their pairwise values, R_{fit,hex} = $\frac{k_{on,hex}^{fit}}{k_{on,hex}^{pairwise}}$ vs. R_{hex} = $\frac{K_{D,hex}}{K_{D,hex}}$. Without the dimer and trimer interactions turned on, R_{fit,hex} = R_{hex} as expected, but as they are turned on, R_{fit,hex} < R_{hex}, due to both faster association and slower dissociation.

Analysis of structural regularity of assembled lattices via the regularization index

NERDSS records the sizes of complexes and the positions of molecules within them. To quantify how uniform and compact the growth of the Gag lattices were, we defined a regularization index (RI). The RI compares the surface coverage of an assembled Gag complex relative to the most compact version that contains the same number of monomers and thus has the same surface area. The most compact version is defined as growth of a spherical cap with the same radius, R = 50 nm. Thus, an ideal lattice that contains N_{gag} with SA_{gag} = 8.5 nm² covers a spherical cap defined by a

polar angle q_{cap} = cos⁻¹ $\frac{N_{gag} SA_{gag}}{2\pi R^2}$. To calculate the RI, we orient our

Gag lattices to be centered relative to a polar angle of zero (centered according to their center of mass) and count what fraction of the Gag monomers are enclosed within the cap defined by the deflection angle q_{cap}.

$$RI \propto \frac{n_{gag \text{ within cap defined by } q_{cap}}}{N_{gag}} \quad (3)$$

An RI of 1 thus has n_{gag} = N_{gag} and refers to ideal, compact growth. Lower values indicate more extended, fractal-like structures. See [supporting material](#) for further details.

Derivation of an optimal titration rate that avoids kinetic traps

To avoid kinetic traps, one approach is to suppress the formation of multiple competing nucleated structures while still allowing efficient growth. We

lattice is not a reactive sphere, as monomers attach at the edge, and it is not a volume-excluding sphere as it grows, but we consider it a reasonable approximation to use fixed estimates, particularly as a R . We also assume that the binding events are irreversible. For stable enough contacts, the unbinding rate is slow compared with growth, but we do see for the weakest contacts $DG_{\text{hex}} \approx 6.4 \text{ k}_B T$; ultimately the growth speed lags behind unbinding as the lattice becomes large, and additional nuclei can form.

Analysis of light-scattering experimental data

To compare our simulation results quantitatively against experiment, we analyzed time-dependent light-scattering data kindly provided from recently published work (11). We estimated a conversion from the absorbance/intensity units to concentration based on the published quantification that >95% of Gag monomers, or nearly 50 mM, were assembled into lattices at the end of experiments. This corresponded to 0.45 absorbance units, au. To determine the initial growth rates of Gag assembly as a function of IP6 concentration, we identified the short-time regimes where growth was approximately linear following a lag time. We extract concentrations of IP6 given that they are initially present only in the “seed” volume, whereas Gag is in both the “bulk” and seed volume, with a bulk to seed solution ratio of 65:15. For IP6 $\approx 1.56 \text{ mM}$ (seed stoichiometry is Gag/RNA/IP6 $\approx 6:2:1$), we linear fit over 3–15 min, measuring a slope of 0.0053 au/min, using MATLAB (The MathWorks, Natick, MA). For IP6 $\approx 9.375 \text{ mM}$ (Gag/RNA/IP6 $\approx 6:2:6$), we linear fit over 1–10 min, giving a slope of 0.015 au/min, and for IP6 $\approx 93.75 \text{ mM}$ (Gag/RNA/IP6 $\approx 6:2:60$) we linear fit over 0.5–2.25 min, giving a slope of 0.0656 au/min. Using the conversion above, we thus extract titration rates of 0.0098, 0.028, and 0.12 mM/s with increasing IP6. Although the stoichiometry we assume will be 6Gag/1IP6, the binding event generates six activated Gag so for the rate estimation the factors cancel. We predict binding rates $k_{\text{IP-Gag}}$ of 126, 60, and $26 \text{ M}^{-1} \text{ s}^{-1}$ with increasing IP6, using Eq. 6 defined below. If the binding to IP6 was rate limiting in all cases, we would expect the exact same rate even as the concentration increases here by 60. The slowdown in the rate (by 4) instead indicates that the growth rate of Gag assembly is being limited by both the IP6 binding and the Gag-Gag assembly kinetics. When we estimate the production of activated Gag we use the fastest rate of $k_{\text{Gag}} \approx 126 \text{ M}^{-1} \text{ s}^{-1}$, solving the set of ordinary differential equations for bulk $G_{\text{agg}} = \frac{dG_{\text{agg}}}{dt} \approx k_{\text{IP-Gag}} \frac{G_{\text{agg}}}{G_{\text{agg}} + G_{\text{agg}}^{\text{off}}} - k_{\text{Gag}} G_{\text{agg}}$ and IP6 $\frac{dG_{\text{agg}}}{dt} \approx k_{\text{IP-Gag}} \frac{G_{\text{agg}}}{G_{\text{agg}} + G_{\text{agg}}^{\text{off}}} - k_{\text{Gag}} G_{\text{agg}}$. So we still assume second-order binding (not sixth order), but with $\frac{1}{G_{\text{agg}}^{\text{off}}} \approx \frac{50}{6} \text{ mM}$, and then $\frac{1}{G_{\text{agg}}^{\text{off}}} \approx 50 \text{ mM}$.

Finally, given these same titration rates, we estimate the rate of binding between Gag once activated, or $k_{\text{a,hex}}^{\text{IP6}}$. To extract this rate, we invert Eq. 8 defined below with the parameters derived from our model, solving for k_{a} given k_{tit} . We show below that these model parameters work effectively in ensuring robust lattice assembly.

RESULTS

The coarse-grained Gag model can assemble a completed spherical lattice consistent with cryo-ET structures

To simulate assembly of the immature Gag lattice consistent with the cryo-ET structure (16), we derived a coarse-grained model of Gag monomers that form three distinct types of interactions (Fig. 1). As shown in Fig. 1, the hexamer and trimer interactions involve two distinct interfaces that thus allow cycles to form in a front-to-back arrangement. Homodimer interactions link together hexamers or

trimers, and ultimately only two types of interactions are necessary to assemble a complete lattice. Below we test assembly both with and without trimer interactions “turned on.” We note that additional stabilizing cycles can form within the Gag lattice beyond just hexamers and trimers, due to the variety of intermolecular contacts between Gag, including a cycle stabilized by a trimer, hexamer, and dimer bond (red dashed lines in Fig. 1 C). Further, at larger length scales, additional stabilizing cycles form via a combination of any two interactions (Fig. 1 D). To assemble the sphere in Fig. 1 D from initially unbound monomers, we promoted nucleation of a single lattice by combining slow titration of monomers into a small volume with fast and irreversible binding rates. Below we characterize assembly kinetics and yield for reversible binding interactions.

Kinetics of fast dimer bond formation agrees with simple theory despite higher-order assembly

By tracking the formation of dimer bonds versus time in our lattice assembly simulations (Fig. 2), we find that the short-time kinetics of dimer association is in close agreement with the analytical theory of populations that can only form dimers ($A + A \rightleftharpoons C$), even when the hexamer association rate is as fast as the dimer rate (Fig. 2 A). This is because the dimer is much more stable than the short-lived hexamer bond; if we make the hexamer contacts more stable, the bonds will start to assemble more synchronously. Thus, as long as dimerization is faster and/or more stable than hexamer contacts, the formation of higher-order contacts does not significantly impact the faster kinetics of monomers forming dimers. Only at times approaching 1 s do we see that the lattice system forms more dimers than expected for the simple bimolecular theory, showing that the formation of hexamers helps to stabilize dimer bonds against disassembly and slow the dissociation process. The comparison between the true dimer dissociation rate ($k_{\text{dim,off}}$), and an effective one ($k_{\text{off,dim}}^{\text{fit}}$), which is fit to the simulation data (see materials and methods), shows how this stabilization effect is stronger with faster hexamer association rate over the first second of assembly (Fig. 2 C). If the trimer bonds can also form, the dimers are stabilized further against dissociation over these short times, even with a slow trimerization rate of $k_{\text{a,trim}} \approx 6.02 \cdot 10^5 \text{ mM}^{-1} \text{ s}^{-1}$ (Fig. 2 C). This is due to the additional cycles that are stabilized in the lattice when even weak trimer contacts can form (Fig. 1). Overall, the smallest assembly unit in most of our simulations is effectively the dimer, and the kinetics proceeds as theoretically expected for the rates we have specified. The dimer contact is known from experiments to form stably under all conditions (15) and these results are consistent with the expected higher-order lattice assembly from dimer building blocks (3).

The kinetics of hexamer bond formation is accelerated by higher-order lattice assembly

Unlike the kinetics of dimer bond formation, we find that even the short-time kinetics of hexamer bond formation deviates significantly from the simple theory for independent sites (A $\neq$ B $\neq$ C). This is perhaps not surprising, as the hexamer bonds allow the assembly of higher-order structures that can become stabilized against disassembly. Even for the weak binding free energy of $\Delta G \approx 6.4 \text{ k}_\text{B}T$, or $K_D \approx 1661 \text{ mM}$ (weak relative to $[C]_0 \approx 50 \text{ mM}$), we see that dissociation of the hexamer bonds slows relative to the pairwise expectations (Fig. 2 B). In contrast, if only hexamer bonds form (we turn off dimer and trimer interactions), the model exactly reproduces theory as expected (Fig. 2). We quantified how the stabilization due to the cycles enabled by dimer and trimer interactions (see Fig. 1) can effectively accelerate hexamer association and slow hexamer dissociation. We defined an effective hexamer dissociation constant, $R_\text{hex}^\text{fit} \approx \frac{k_\text{off,hex}^\text{fit}}{k_\text{on,hex}^\text{fit}}$, with rates defined by fitting our simulated data to the functional form of the simple theory (see [materials and methods](#)). As we increase the hexamer on- and off-rates while keeping the ΔG fixed, we see that the dissociation slows systematically, allowing faster assembly of effectively more stable hexameric contacts (Fig. 2 D).

These simulations also illustrate the role that the additional trimer contacts play in stabilizing the structures against disassembly. Although we add in trimer interactions at only a weak $\Delta G_\text{trim} \approx 4.1 \text{ k}_\text{B}T$ (K_D of 16.6 mM), they nonetheless quantitatively shift the kinetics of hexamer bond formation, helping to accelerate bond formation by further slowing dissociation (Fig. 2 D). The impact of the trimer contacts is diminished if the strength of the hexamer bond formation is increased, as we quantify further below. This is also not surprising, as stronger hexamer bonds will not need “help” from other contacts to stabilize bonds. Finally, we confirm that, despite the faster formation of hexamer contacts, the equilibrium bonds formed after 100 s are the same under constant ΔG_hex (Fig. S2). Adding in the trimer interaction along with the dimer ensures 15% more hexamer bonds form for these energies (Fig. S2).

Kinetic trapping emerges even for relatively weak ΔG due to the lattice size

By comparing systems with the same energetics but distinct hexamer binding kinetics, our systems will eventually reach the same equilibrium distribution of assemblies, but the assembly pathways they follow will be distinct. However, even for a relatively weak set of interactions ($\Delta G_\text{hex} \approx 6.4 \text{ k}_\text{B}T$), we already see that our system is kinetically trapped. Specifically, for all systems, by 10 s

monomers and dimers have been depleted, leaving essentially no simple building blocks to complete the growth of intermediate-sized structures (Fig. 3, A–C and [Video S1](#)). The size distribution of these assembly intermediates shifts slightly larger with faster hexamer rates over 100 s, although for the two fastest rates the distributions show a more similar steady state (Fig. 3, D–F). To grow largely complete lattices, these stalled (Fig. S3) and monomer-starved systems now must either wait for dissociation of larger complexes or wait for the rare annealing of two intermediate-sized complexes. Given enough time, our systems will eventually reach an equilibrium steady state, where the distribution of complexes is independent of the rate constants, and below we show how these same systems can be guided around these trapped intermediates to form large single complexes over much shorter timescales. A major factor driving trapping is the total number of monomers required to complete a lattice. Although our systems can assemble hundreds of Gag monomers into a single structure (Fig. 3), the complete lattice requires 3700 monomers. For high-yield assembly, the rate of nucleating new structures should be slower than the elongation rate to complete the lattice, but adding thousands of monomers onto one structure takes significantly longer than forming a capsid that contains 10 or 100 times fewer components.

As for the assembly pathways, faster binding kinetics (both on and off rates) leads to slightly fewer and thus larger intermediates on average (Figs. S3 and S4). This is consistent with the results of Fig. 2, where faster hexamer kinetics combined with dimer and trimer contacts will stabilize early growth and thus allow faster elongation of the nucleated structures before the building blocks are used up. For slower rates, in contrast, more nuclei form while the concentration is relatively high, and growth is slow. This trend was preserved for the weak and stronger free energies tested ($\Delta G_\text{hex} \approx 6.4$; 8.7 , and $11.4 \text{ k}_\text{B}T$) and with the trimer turned on and off (Fig. S4). By speeding up both the dimer and hexamer rates, we can accelerate growth, such that only 4–6 nuclei form instead of 8–10. This reduction in the number of intermediates nucleated is more sensitive to the hexamer association rate than to dimer and trimer rates, as we see a clear increase in the sizes of complexes formed with faster hexamerization (Fig. S4). We note that, while we did not establish a formal size of a critical nucleus (in terms of a free-energy maxima following critical nucleation theory), we did consistently define a nucleus as having at least 18 Gag monomers since that size structure would contain multiple stabilizing cycles that would significantly slow dissociation ([materials and methods](#) and Fig. 1).

Rapidly assembled intermediates have less uniform growth

To compare the topology of our assembled structures, we defined a RI that measures how much a lattice deviates

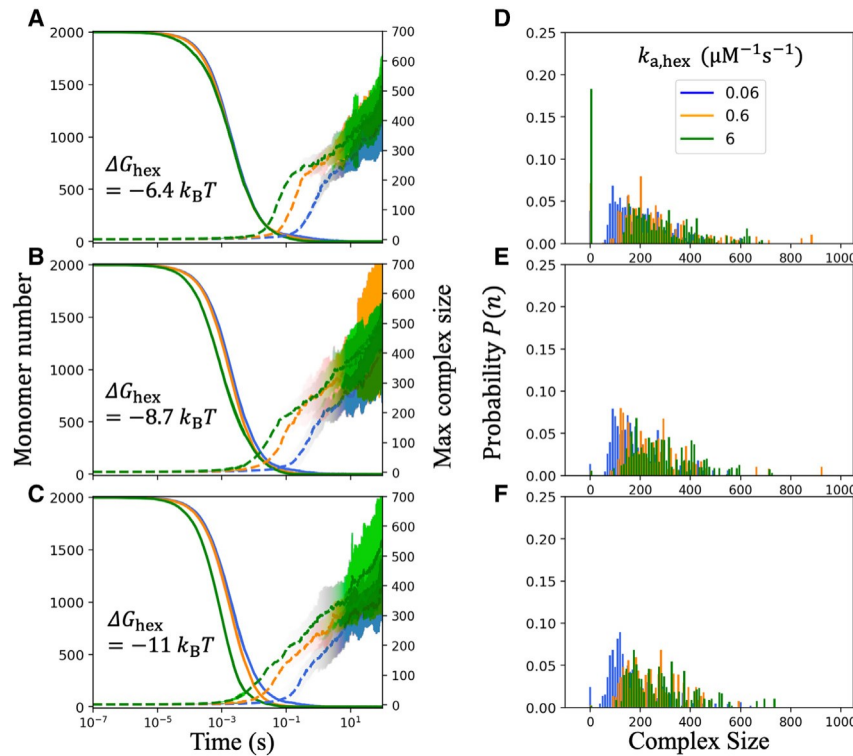


FIGURE 3 Kinetic traps emerge when all three interactions are turned on even with relatively weak DG. On the left column (A–C), we show the kinetics of monomer depletion (solid lines) and simultaneous formation of the largest assembled complex in the system (dashed lines). The hexamer rate $k_{a,hex}$ increases from $6:02 \cdot 10^2$ (blue), $0:602$ (orange), $6:02 \text{ mM}^{-1}\text{s}^{-1}$ (green), with $k_{b,hex}$ increasing to keep ΔG_{hex} constant. Then, from top (A) to bottom (C) the stability of the hexamer interaction ΔG_{hex} increases, by slowing $k_{b,hex}$. As the small building blocks are depleted by 10 s, growth stalls and the systems become trapped in incomplete intermediates (also see [Video S1](#)). The probability distributions on the right (D–F) show corresponding complex sizes of intermediates present after 100 s of simulation for the same simulations to their left. All systems exhibit multiple nucleated structures, rather than a single growing structure. With faster hexamer kinetics (green and orange), the distribution of complex sizes shifts to larger values compared with blue data (also see [Fig. S3](#)). Even for $\Delta G_{hex} \approx 6:4 \text{ k}_B T$, only a few monomers remain free in solution, leaving no room for additional growth without dissociation. All results are averaged over 16 trajectories, and standard deviation is shown around the plotted sample mean versus time. We set $k_{a,trim} \approx 6:02 \cdot 10^5 \text{ mM}^{-1}\text{s}^{-1}$, $k_{b,trim} \approx 1 \text{ s}^{-1}$, $\Delta G_{trim} \approx 4:1 \text{ k}_B T$; $k_{a,dim} \approx 6:02 \text{ mM}^{-1}\text{s}^{-1}$, $k_{b,dim} \approx 1 \text{ s}^{-1}$, $\Delta G_{dim} \approx 15:6 \text{ k}_B T$; and $D_t \approx 10 \text{ mm}^2\text{s}^{-1}$; $D_{rot} \approx 0:01 \text{ rad}^2\text{s}^{-1}$; $D_t \approx 0:5 \text{ ms}$, boxlength $\approx 405 \text{ nm}$. To see this figure in color, go online.

from a compact, ideal spherical growth (shaded with dark gray in [Fig. 4](#)). Ideally, compact structures have $RI \approx 1$, while a more extended fractal-like growth has increasingly lower RI. Compact structures have the shortest edge length, and thus they maximize the number of stabilizing cycles formed. Analyzing the same systems as in [Fig. 3](#), we find that smaller complexes of 100 Gag monomers have a higher RI on average and that the RI systematically decreases as the sizes of the complexes increases ([Fig. 4](#)). This trend occurs for each ΔG_{hex} ([Video S1](#)), but with less stable free energy there is a shift toward higher RI. This indicates that faster dissociation allows structures to grow more uniformly and compactly by correcting contacts that extend the growth and fail to satisfy as many bonds.

We further quantified growth pathways, showing that they proceed primarily through addition of monomers, dimers, and small oligomers of <18 Gags, although some rare annealing events between large structures do occur ([Fig. S5](#)). These annealing events are unlikely because they require that the two complexes have compatible edge geometries and, when such annealing does happen, highly irregular structures can form ([Fig. 4 B](#)). They are also rare because the structures must be aligned properly to avoid large-scale reorientation upon binding. Our algorithms will reject moves that cause any component of the assembly to move significantly more than expected due to translational and

rotational diffusion as specified by a threshold (see [materials and methods](#)).

A two-phase equilibrium emerges with significantly weakened ΔG_{hex}

Eliminating higher-order assembly requires dropping the interaction strength significantly lower than would be required for one type of interaction ([Fig. 5](#)). A strong dimer interaction only reduces the total components from 50 mM of monomers to 25 mM of dimers, so we focus first on the hexamer strength, comparing $\Delta G_{hex} \approx 1:7$, $4:1$, and $6:4 \text{ k}_B T$. As a baseline, we consider a system that can form up to hexamers but no higher-order lattices, meaning we turn off the homodimer and trimer contacts. At $\Delta G_{hex} \approx 6:4 \text{ k}_B T$, only 0.001% of these Gag monomers will form hexamers ([Fig. 5 A](#)). As we turn on the other interactions and increase the number of cycles possible in the lattice, the transition window from no assembly to robust assembly will shrink and shift with small changes to ΔG_{hex} . With homodimer contacts added, a larger assembly will form at $\Delta G_{hex} \approx 6:4 \text{ k}_B T$, although $>90\%$ of components present in solution are Gag monomers, whereas at $4:1 \text{ k}_B T$ we see no assemblies formed ([Fig. 5 B](#)). Finally, with all three interactions present, at $\Delta G_{hex} \approx 6:4 \text{ k}_B T$, the system consumes nearly all monomers/dimers. Large assemblies form even with $\Delta G_{hex} \approx 4:1 \text{ k}_B T$, along with a

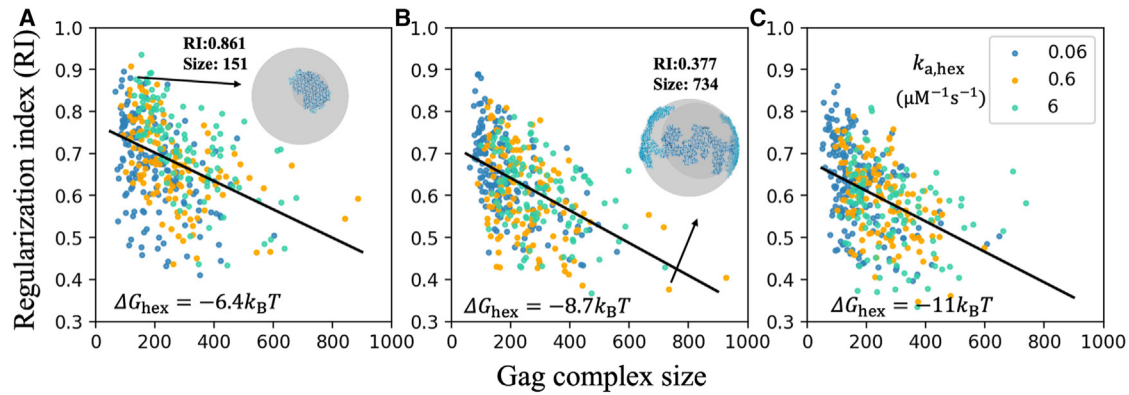


FIGURE 4 In kinetically trapped systems, growth of larger intermediates is less compact and regular. The regularization index (RI) of assembled complexes with $N_{\text{gag}} > 100$ at steady state is on average higher for smaller assemblies. The sphere in (A) illustrates a high RI showing a significant overlap of the simulated lattice in blue with the ideal spherical cap in dark gray. The one in (B) shows a low RI. Data in (A) $\Delta G_{\text{hex}} \approx 6.4 k_B T$, (B) $\Delta G_{\text{hex}} \approx 8.7$, (C) $\Delta G_{\text{hex}} \approx 11 k_B T$. Black lines are from linear regression analysis, showing a clear negative correlation between RI and Gag complex size for all panels. Pearson correlation tests further substantiate this correlation, which yields Pearson correlation coefficients of 0.42, 0.5, and 0.44 from (A) to (C). The p values are all extremely small ($< 10^{-19}$). Blue, orange, and green dots represent lattices formed in systems with association rate of 0.06, 0.6, and 6 $\text{mM}^{-1}\text{s}^{-1}$. While there is no significant trend of the RI with changing rates, there is a significant trend of higher RI values with weaker ΔG_{hex} . The best-fit lines have similar slopes from (A) to (C), but the y-intercept decreases from 0.77 to 0.72 to 0.68, respectively. Same simulation data and parameters as Fig. 3. To see this figure in color, go online.

dilute phase of monomers (Fig. 5 C). For this model, we return to a fully monomeric system when ΔG_{hex} is $1.7 k_B T$. A two-phase equilibrium with a dilute phase of monomers and dimers alongside a collection of partial assemblies is thus difficult to achieve unless only two interactions are stabilizing the lattice (Fig. S6, E–G). The stable dimer interaction combined with either hexameric interactions or trimer interactions at $\Delta G \approx 6.4 k_B T$ nucleate partial assemblies, but growth ultimately stalls as the monomer concentration dilutes (Fig. S6, A and B). These two lattices

reach the same equilibrium despite the trimer being fundamentally more stabilized against disassembly, as the higher-order cycles between dimer and trimer require more bonds than dimer and hexamer (Fig. 1).

Mimicking activation by cofactors can ensure slow nucleation and fast growth of a single lattice

Our simulations above demonstrate that, to prevent kinetic trapping in bulk simulations (where all monomers are

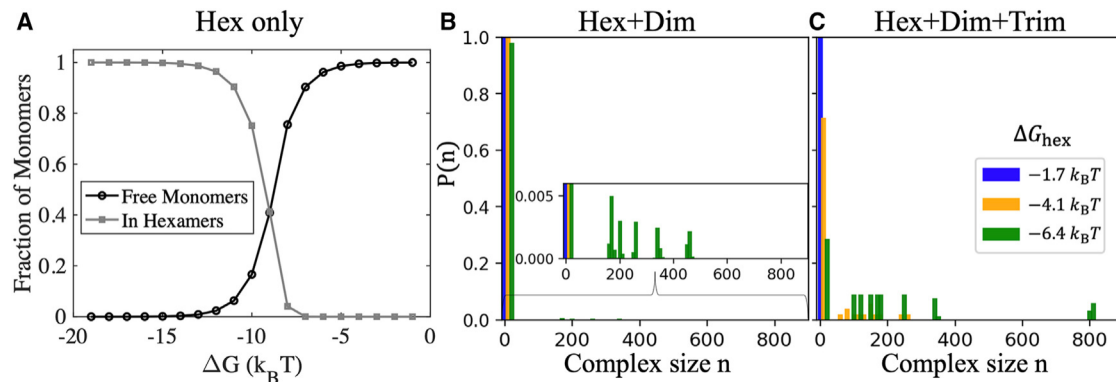


FIGURE 5 Multivalent assembly supports lattice formation with weak hexamer contacts. (A) For reference, we numerically calculated the equilibrium for Gag monomers (50 mM) that only had hexamer contacts enabled using MATLAB. The transition to $< 50\%$ fraction as free monomers occurs at $\Delta G_{\text{hex}} \approx 8.74 k_B T$. (B) For Gags that assemble the full lattice via dimer and hexamer contacts, we start to see formation of large assemblies when $\Delta G_{\text{hex}} \approx 6.4 k_B T$ (green bars), with no assembly for $1.7 k_B T$ (blue bars) and $4.1 k_B T$ (orange bars). (C) Same as (B) but now the trimer interaction is turned on at $\Delta G_{\text{trim}} \approx 4.1 k_B T$, triggering assembly at all but the lowest hexamer strength. Both (B) and (C) report distribution over a 5 s time window following steady state. We note that, due to the slow on-rates, some nucleation of lattices could still occur at longer times beyond those simulated here (100 s). The distributions are normalized over the number of distinct types of complexes, meaning that if 10 different sized complexes are present, and one is a monomer, then the probability of the monomer state is 0.1, even though only 1 in 2000 Gags are in the monomer form. Data collected from one trajectory. $k_{a,\text{hex}} \approx 6.02 \cdot 10^4$; $6.02 \cdot 10^3$; $6.02 \cdot 10^2 \text{ mM}^{-1}\text{s}^{-1}$, $k_{b,\text{hex}} \approx 100 \text{ s}^{-1}$; $D_t \approx 0.5 \text{ ms}$, and $D_t \approx 10 \text{ mm}^2\text{s}^{-1}$; $D_{\text{rot}} \approx 0.01 \text{ rad}^2\text{s}^{-1}$; boxlength $\approx 405 \text{ nm}$, $k_c \approx 60 \text{ mM}^{-1}\text{s}^{-1}$. For (B) and (C) $k_{a,\text{dim}} \approx 0.602 \text{ mM}^{-1}\text{s}^{-1}$, $\Delta G_{\text{dim}} \approx 13.3 k_B T$. For (C); $k_{a,\text{trim}} \approx 6.02 \cdot 10^5 \text{ mM}^{-1}\text{s}^{-1}$, $\Delta G_{\text{trim}} \approx 4.1 k_B T$. To see this figure in color, go online.

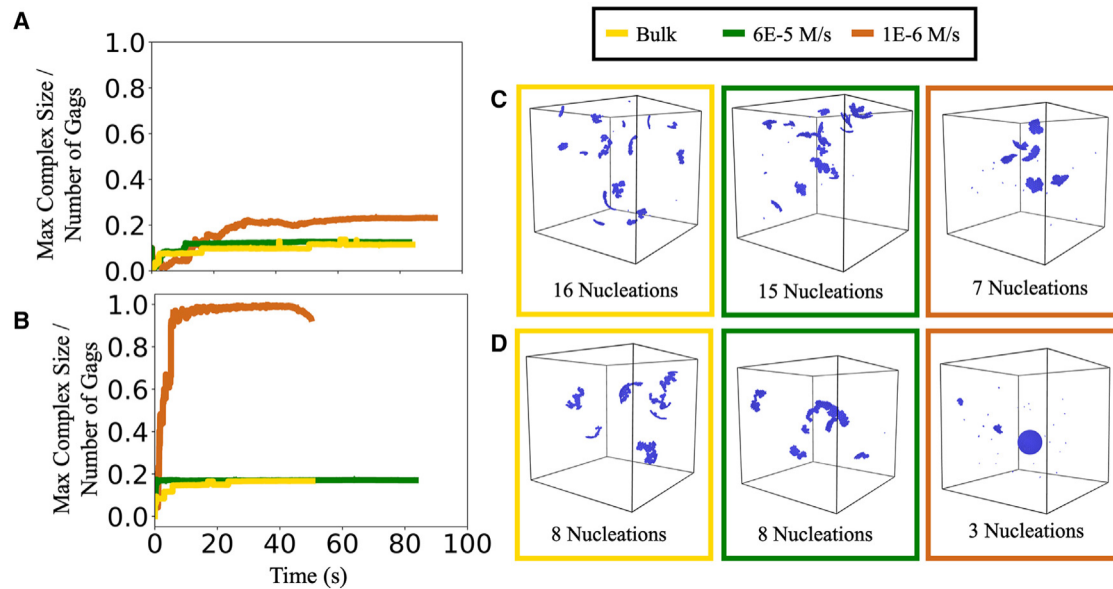


FIGURE 6 Titration in Gag monomers can reduce nucleation events and improve lattice growth. (A and B) The largest lattice formed as a function of time is normalized by the number of Gag monomers, so 1 is 100% of Gag in a single complex. The figure shows bulk simulations (yellow) and titration with 60 mM/s (green) and 1 mM/s (orange). Titration was stopped when the final concentration of 50 mM was reached, which occurred after 0.83 s (green) or after 50 s (orange). (A) Has slower binding interactions, $k_{a,hex} \approx 0.06 \text{ mM}^2\text{s}^{-1}$ and $k_{a,dim} \approx 0.006 \text{ mM}^2\text{s}^{-1}$, $k_{b,hex} \approx 1 \text{ s}^{-1}$. (B) Has faster binding, $k_{a,hex} \approx 12 \text{ mM}^2\text{s}^{-1}$ and $k_{a,dim} \approx 0.6 \text{ mM}^2\text{s}^{-1}$, $k_{b,hex} \approx 20 \text{ s}^{-1}$. For both (A) and (B) $k_{b,dim} \approx 1 \text{ s}^{-1}$, $k_{a,trim} \approx 6.02 \cdot 10^5 \text{ mM}^2\text{s}^{-1}$, $k_{b,trim} \approx 1 \text{ s}^{-1}$. For the slowest titration rate (orange) in (B), a single nuclei was grown during most of the simulation. Results from one trajectory each. In (C) and (D), we show a repre-sentative snapshot from the end of the corresponding simulations. In (B) and (D), the slowest-titration simulation uses a larger $(510 \text{ nm})^3$ box to allow a lattice to complete, which it nearly does in this single trajectory (bottom right) before a few small nuclei form. To see this figure in color, go online.

present at time zero) for the large HIV lattice, we would need to identify the very small range of ΔG_{hex} values that can complete growth of the lattice before multiple nuclei form. Without the trimer, this would require slightly more stability than $6.4 k_B T$ (Fig. 5), but weaker than $8.7 k_B T$ (Fig. S6), and with faster association rates (at 50 mM). With the trimer, values would shift to between 4.1 and $6.4 k_B T$. However, assembling from bulk monomers is not a problem the HIV lattice has evolved to solve. Both in vitro and in vivo, the Gag lattice assembly is only productive with cofactors, and therefore an alternate strategy for our model is to mimic the behavior of cofactors. Binding to cofactors effectively changes the concentration of “active” or assembly-competent Gag monomers in a time-dependent fashion. In the bulk simulations above (Figs. 3, 4, and 5), the activation was therefore instantaneous, implying a high concentration of cofactors and efficient binding.

Instead of considering the dependence of activation on both the concentration and binding rate of cofactors to the Gag monomers, we simplify our approach and use a titration rate k_c . From mass action kinetics, the rate of activated Gag production is then $\frac{d\%G^{act}}{dt} \approx k_c$. For the explicit bimolecular process, we have $\frac{d\%G^{act}}{dt} \approx k_f \frac{\%R_0\%G_0}{\%R_0 + \%G_0}$, where k_f is the binding rate between Gag and cofactor, $\%R_0$ the concentration of cofactor, and $\%G_0$ the concentration of bulk Gag before activation, with initial concentrations $\%R_0$ and

$\%G_0$, respectively, and $\%G^{act} \approx 0$. By comparing these two we can map our titration rate to the effect of cofactor binding,

$$k_c = \frac{1}{2} G_0 2k_f \frac{1}{2} R_0 \quad (6)$$

which is an accurate approximation at earlier times where the concentrations are close to their initial values. This comparison to a titration process will not directly match the time dependence of the bimolecular association process as concentrations become limited and thus deplete with time, in which case the bimolecular process slows relative to titration. Nonetheless, it establishes useful bounds on the time-scales of binding as they relate to a single titration rate: if slow titration is necessary, this would require either a lower concentration of cofactor or slower binding of cofactor to Gag. In Fig. 6 we test how introducing a titration process to “activate” the Gag monomers can reduce the number of nuclei and thus improve the formation of more complete lattices. If the titration is too fast relative to the binding between the Gag monomers, however, improvement is minimal, indicating how this rate must be calibrated relative to the Gag assembly kinetics (Fig. 6).

We note that the growth of our lattices does show some dependence on the diffusional search of the monomer. Simulations in larger boxes (510 vs. 405 nm boxlength) reach the same concentrations in time, but the larger box allows the complete 3700 monomer lattice to form. That means

that forming a single nucleus in the larger box requires the monomers to travel larger distances before “finding” the nucleus. This increased travel time slows growth of the nucleated lattice and results in higher likelihood of a new lattice forming. Thus, the titration rate should reflect the volume needed to complete a lattice. We also found that decreasing the diffusion coefficients of the monomers could shift the distribution of complexes to slightly more nucleated structures (Fig. S7). This result could not be primarily attributed to just a slower search, as a similar trend occurred when we changed the time step. Instead, we found that our criteria for rejecting association events that involve large rotational motion by larger complexes

$$k_{\text{titr}} \propto \frac{1}{tV} \quad (7)$$

where t is the timescale for a protein to bind to a nucleated structure in a volume V . We define the timescale t by mapping to a well-defined problem in the theory of diffusion-influenced reactions for binding to a reactant in a fixed volume (34). We ignore the dissociation of proteins, or assume binding to the nuclei is irreversible, and, of the two main stabilizing reactions (dimer and hexamer), we quantify binding times using the slowest rate, which in this case is the hexamer rate. Our predicted rate is then:

$$k \propto \frac{315a^2D_t}{4\pi R^6} \frac{a^3}{R^3} \frac{R^2}{a^2} \frac{R^2}{aR} \frac{R^2}{R^2} \frac{5D_t}{k_a r_0} \frac{a^3}{R^3} \frac{\#_{\text{titr}}}{k_a r_0} \quad (8)$$

was more permissive with faster diffusion ($D_t \propto 50 \text{ mm}^2/\text{s}$, $D_r \propto 0.05 \text{ rad}^2/\text{ms}$ vs. $10 \text{ mm}^2/\text{s}$ and $0.01 \text{ rad}^2/\text{ms}$). Practically, this indicates that some of the early growth of the complexes is occurring due to moderately sized oligomers combining to maintain fewer nucleated structures. However, this also means that the formula controlling the probability of accepting annealing events between large complexes shows a small dependence on the time step and diffusion constant. A more rigorous definition of this formula is needed to ensure that this acceptance criteria is fully invariant under changes to diffusion or time step. Importantly, however, we verified that the number of bonds formed is conserved (<3% error) regardless of spatial extent, diffusion constant, or time step. Because the effect of this annealing parameter on the distribution of assemblies is small, we do not see changes to our observed trends with rates or free energies, and in future work this unexpected coupling between rare annealing events and diffusion will be corrected.

The titration rate can be derived to promote nucleation of complete lattices

From Fig. 6 we clearly see that slowly titrating in or “activating” Gag monomers can improve nucleation and growth of larger lattices by keeping the concentration low. Here, we derive an optimal titration rate to ensure the growth of completed lattices. The idea is to limit the rate that a new protein appears in the simulation volume to below the rate that a protein will bind to a single nucleated structure in the volume. In this way, each new protein will contribute to growth of the single nuclei rather than formation of a distinct one. Hence, we seek to derive an expression:

where D_t is the monomer diffusion constant, $a \approx 7 \text{ nm}$ is the radius of an initially nucleated structure, $k_a \propto k_{a;\text{hex}}$ and $r_0 \approx 0.16 \text{ nm}^2$ is an approximate density of binding sites on a Gag lattice (see materials and methods for details). Finally, what is the appropriate value for V in a real system, or the length scale R that confines our assembly process? Both t and V scale with R^3 , hence the sensitivity of k_{titr} to R (Fig. S1). The volume must be physically large enough that a complete lattice can form (with $R_{\text{lattice}} \propto 50 \text{ nm}$) and contain enough monomers for the complete lattice given the concentration of Gag in solution. Thus, we have that

$$V \propto N = \frac{1}{2}G_0 \quad (9)$$

where N is the total monomers needed to complete a lattice (3700 for a 50 nm sphere), and $\frac{1}{2}G_0$ is the concentration of Gag monomers in solution. Hence, a smaller volume is needed with a higher concentration of Gag. The smaller R we choose, the faster the titration rate we can use, so ultimately, the rate of titration or “activation” must be calibrated to the concentration of Gag in solution to form completed lattices. This length scale will also ultimately control the concentration of lattices in solution, $[L]$, as assuming all monomers add to a complete lattice, $\frac{1}{2} \propto k_{\text{titr}}t_{\text{titr}}=N$ where t_{titr} is the length of time over which the monomers are titrated to reach the target Gag concentration. We used a value of $R \propto 202 \text{ nm}$ for our derivation, comparable with our simulation box size. The corresponding simulations (Fig. 7) then reached the same Gag concentrations as our previous ones, with 50 mM of Gag. We note that this volume does not contain the copies necessary for a complete lattice, which would instead be $R \propto 308 \text{ nm}$. For the larger volume, a slower titration rate is predicted, although we see below that the assembly process can partly tolerate increases in the simulation volume.

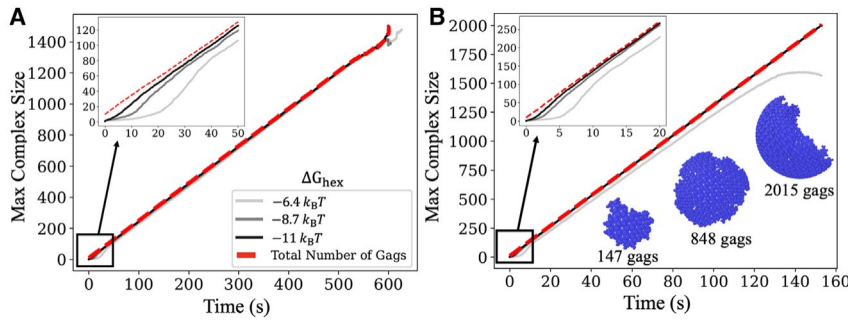


FIGURE 7 Derived titration method ensures productive nucleation and growth. (A) A single complex forms as we titrate in monomers using our derived rate $k_{\text{titr}} \approx 0.06 \text{ mM}^{-1} \text{ s}^{-1}$. All data are an average over 48 independent trajectories per model. $k_{\text{a,hex}} \approx 0.6 \text{ mM}^{-1} \text{ s}^{-1}$ for each curve, but ΔG_{hex} increases from $6.4 \text{ k}_B T$ (light gray) $8.7 \text{ k}_B T$ (gray) $11 \text{ k}_B T$ (black) (Video S2). The total number of Gags is shown as the dashed red line. The inset shows the initial lag time which extends with faster dissociation (light gray). (B) Here, the hexamer association rate was accelerated, $k_{\text{a,hex}} \approx 6 \text{ mM}^{-1} \text{ s}^{-1}$ and thus we used a faster $k_{\text{titr}} \approx 0.33 \text{ mM}^{-1} \text{ s}^{-1}$. $k_{\text{a,dim}} \approx 0.6 \text{ mM}^{-1} \text{ s}^{-1}$, $k_{\text{b,dim}} \approx 1 \text{ s}^{-1}$, $k_{\text{a,trim}} \approx 6.02 \times 10^5 \text{ mM}^{-1} \text{ s}^{-1}$, $k_{\text{b,trim}} \approx 1 \text{ s}^{-1}$. To see this figure in color, go online.

In Fig. 7 we show that this method provided excellent control to ensure the growth of a single nucleated lattice. We tested two hexamer association rates ($k_{\text{a,hex}} \approx 6$ and $0.6 \text{ mM}^{-1} \text{ s}^{-1}$) and three hexamer stabilities ($\Delta G_{\text{hex}} \approx 6.4$; 11.0 ; $13.3 \text{ k}_B T$). The systems remain free of kinetic traps as they nucleate and grow, forming a single large assembly in our volume (Video S2). The one exception is the weakest lattice ($\Delta G_{\text{hex}} \approx 6.4 \text{ k}_B T$), which after reaching a size containing 1200 monomers does start to nucleate additional structure in a fraction of the trajectories. The reason is that the association is counterbalanced by more frequent dissociation reactions for this weak interaction. Our simulations indicate that growth is therefore slowing somewhat as the lattice is getting larger, which would explain why ultimately new nuclei are able to form before the titrated monomers are added onto the existing structure, but they do not do so earlier in the trajectories. This further suggests that the growth of this lattice, with $\Delta G_{\text{hex}} \approx 6.4 \text{ k}_B T$, is not strongly stabilized against the competing disassembly, and thus assembly is less robust, requiring more tight temporal control to complete lattice formation.

Our simulations also show how the nucleation of the initial structures is dependent on both the on- and off-rates (Fig. 7). The lag time before reaching the linear growth phase is longer when off-rates are faster, consistent with more time needed to stabilize the nucleation site. For all parameter regimes these simulations produce much more compact and ideal-like growth of our structures compared with earlier simulations (Fig. 3), with average values of the regularization index now 0.85 – 0.95 that are quite close to 1 at all times (ideal growth). Our derivation of k_{titr} above is approximate, as it assumed a fixed size and reactivity for the nucleation site, whereas in reality the nucleus grows larger and the reactive sites remain always along the edge of the lattice. Nonetheless, we find that this method keeps a low enough concentration of Gag monomers that we can now reproduce smooth and productive growth, even for highly stable lattices ($\Delta G_{\text{hex}} \approx 11 \text{ k}_B T$) that would otherwise rapidly

become trapped into intermediates under bulk conditions (compare with Fig. 5).

Comparison with in vitro assembly kinetics reveals how slow binding of Gag to IP6 can support robust Gag assembly

The cofactor IP6 binds to Gag and activates it for assembly into the immature lattice (11,15,43). We therefore analyzed a recently collected set of in vitro experiments that tracked (via light scattering) the formation of Gag lattices versus time as a function of IP6 concentration (11) (Fig. 8). Light scattering can be analyzed to quantify assembly parameters (44), but cofactors will alter the observed kinetics. Based on our theoretical arguments above, if activation of the Gag monomers by IP6 binding was slow enough, then we would see a linear increase in Gag complex size, following a lag time. Furthermore, the Gag assembly kinetics should also accelerate with higher IP6 concentration if this activation is more rate limiting than the Gag-Gag assembly kinetics. We do observe both this linear early increase in Gag complex size and the acceleration of this growth rate with higher IP6 concentration (Fig. 8 A), indicating faster activation of Gag into its assembly-competent form. Therefore, we proceed to use our theoretical arguments to extract the effective titration rates. With titration rates, we can then estimate bounds on the binding rate of IP6 to Gag, $k_{\text{IP6-Gag}}$, and the binding rate of Gag to Gag, $k_{\text{a,hex}}^{\text{IP6}}$.

To extract the k_{titr} , we simply linear fit this short-time growth rate in the experimental kinetics (see materials and methods). We thus measure three titration rates, and then we can use Eq. 6 to estimate the binding rate of IP6 to Gag, $k_{\text{IP6-Gag}}$. We predict a rate $k_{\text{IP6-Gag}}$ for each titration rate, and we expect them to be identical if the growth of Gag assembly is fully rate limited by Gag-IP6 binding. The rates are indeed similar, but not identical, showing a modest slowdown in this predicted rate at higher IP6 concentrations (materials and methods). A simple interpretation is that the Gag is now being activated by IP6 a bit faster than the assembly process can keep up, so Gag activation is not fully rate limiting. We therefore use the value at the lowest

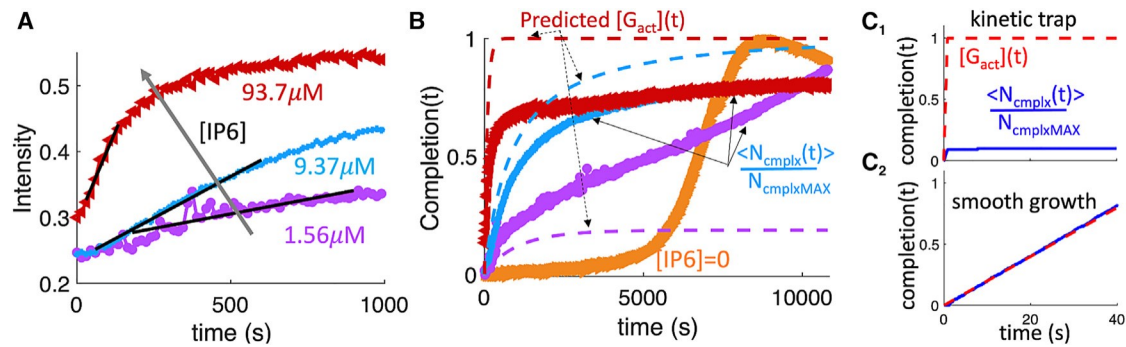


FIGURE 8 Analysis of in vitro Gag assembly kinetics indicates that IP6 can activate Gag slowly enough to support assembly with limited trapping. (A) Light scattering intensity versus time is replotted here from recent experiments measuring in vitro Gag assembly (11). Gag and RNA concentration is the same for all curves, with 50 mM Gag and 3.125 mM RNA (159 nt each) in solution. Concentration of IP6 increases from 1.56 mM (purple) to 9.37 (blue) and 93.7 (red), clearly driving faster Gag lattice assembly. A regime of relatively linear growth following a lag time for each system gives a rate estimate for IP6 “activating” Gag (black lines). (B) Same experimental data, but we convert to assembly completion fraction using 0.45 absorbance units to 100% complete (completion % 1) (see materials and methods). We subtract 0.22 or 0.24 to re-zero the y axis. Using our estimated rate of Gag binding to IP6 ($125 \text{ M}^{-1} \text{ s}^{-1}$), we predict the production of the IP6-activated Gag ($[G_{\text{act}}(t)]$) in dashed lines following a model of bimolecular association, normalized by the total Gag concentration of 50 mM (materials and methods). This activated Gag we infer as now competent to bind other Gag much more quickly, thus allowing the Gag assembly observed experimentally (solid colors). For low IP6 (purple data), it is apparent that not all Gag needs to be “activated” by IP6 to assemble, with a slower timescale of assembly occurring after the initial rapid assembly. This is known from experiment: even without IP6 (but with RNA), Gag assembly eventually occurs following a long lag time (orange data). (C) Our simulation data (solid blue lines, same as Fig. 6 B) can also report a completion fraction given the average size of Gag complexes in time normalized by the maximal complex size (e.g., $N_{\text{cmplx}}/N_{\text{cmplxMAX}} \approx 3700$ monomers). Here, the Gag is explicitly activated in time via titration, so the red dashed lines $[G_{\text{act}}(t)]$ due to titration, normalized by the total target concentration of 50 mM. In (C₁) we activate the Gag too quickly compared with its assembly, and thus the assemblies become trapped. In (C₂), we activate slowly enough that we see robust assembly growth over these shorter times. We infer that the experimental results in (B) at higher IP6 (9.37 and 93.7 mM) are a bit too fast in activating Gag, leading to assembly that results in some kinetic trapping, or a plateau that does not reach >95% yield. To see this figure in color, go online.

concentration of IP6 as the upper bound, $k_{\text{IP6-Gag}} 126 \text{ M}^{-1} \text{ s}^{-1}$. Also, because the IP6 is initially seeded with 18% of the Gag before the full mixing and measurement (11), the IP6 binding could actually be slower than this rate, so we can interpret it as an upper bound.

In Fig. 8 B we illustrate how the full-time evolution of the experimental Gag assembly can be interpreted via these binding rates. Light scattering reports on the average molecular weight of solute, so it is most directly mapped to the average mass of complexes in simulation (Fig. 8 C), where we normalize by total Gag in a complete complex to report a completion fraction. We predict the concentration of IP6-activated Gag in Fig. 8 B using our value of $k_{\text{IP6-Gag}}$, where activated Gag can then assemble complexes. At the high IP6 concentrations (9.37 and 93.7 mM), the activation is a bit faster than assembly kinetics, which leads to initially fast growth followed by some kinetic trapping, as the growth plateaus at 80% completion rather than 100% (Fig. 8 B). This is similar to our simulation results in Fig. 8 C₁ except with much higher completion fraction. However, in the lowest IP6 concentration, for short times the growth is more like Fig. 8 C₂, indicating smooth growth following activation. At the lowest IP6 concentration, we can maximally activate only 9.37 mM Gag, or 20% of the total. However, assembly still proceeds beyond these times, albeit much more slowly, meaning that not all Gag need to be bound to IP6 to promote assembly. Although we did not include this possibility in our model, this is consistent

with the known assembly dynamics of the immature Gag lattice absent of IP6 (but still with RNA), which has a very long delay but eventually a cooperative assembly into largely complete lattices (Fig. 8 B, orange curve). This comparison indicates that IP6 significantly accelerates assembly, but it can also bias away from a very slow but highly cooperative growth that occurs only due to RNA-driven activation.

Finally, we can use our fit values of k_{titr} that we extracted from the experimental data to estimate the limiting rate of Gag-Gag assembly when IP6 activates, or $k_{\text{a,hex}}^{\text{IP6}}$ via Eq. 8. To justify estimates on $k_{\text{a,hex}}^{\text{IP6}}$, if the titration rate (IP6-driven activation rate) is too fast compared with Gag assembly then we expect a more bulk-like condition that will promote kinetic trapping (Fig. 8 C₁). If the titration rate is slow enough, the assembly kinetics will be limited by the speed of activation. Then we can estimate what the slowest assembly rate could be to ensure smooth growth, although the actual rate could also be faster; thus, we can only put a lower bound on the binding rate (Fig. 8 C₂). Using the titration rate for IP6 $\approx 1.56 \text{ mM}$ over short times estimated as $k_{\text{titr}} 0.0098 \text{ mM/s}$, we have a lower bound on $k_{\text{a,hex}}^{\text{IP6}}$ of either $7.7 \cdot 10^4 \text{ M}^{-1} \text{ s}^{-1}$ using $R \approx 202 \text{ nm}$, or $1 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$ using $R \approx 308 \text{ nm}$. An upper bound from the kinetics at IP6 $\approx 9.73 \text{ mM}$ where $k_{\text{titr}} 0.028 \text{ mM/s}$ is then $2 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ using $R \approx 202 \text{ nm}$ or $3.8 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$ using $R \approx 308 \text{ nm}$. Hence, our estimates put it between $7.7 \cdot 10^4 \text{ M}^{-1} \text{ s}^{-1} < k_{\text{a,hex}}^{\text{IP6}} < 3.8 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$. This

indicates that the Gag-Gag assembly kinetics once activated by IP6 is significantly faster than IP6 binding to Gag ($k_{\text{IP6-Gag}} 125 \text{ M}^{-1} \text{ s}^{-1}$), and thus IP6 can be effective in relatively slowly activating Gag for efficient assembly even when its concentrations are higher. It is also clear that Gag assembly kinetics slows when Gag must assemble lattices without all monomers benefiting from IP6 activation (Fig. 8 B).

DISCUSSION

By developing a model of the immature Gag lattice from the cryo-ET structure, we quantified here how the strength and kinetics of the multiple Gag-Gag contacts within the lattice could support productive self-assembly from free monomers. A key component of this reaction-diffusion model is the energetic and kinetic parameters compare directly to free energies and biochemical rate constants and can thus inform pairwise interactions that are experimentally testable. A primary finding is that the Gag lattice is simply too large to assemble robustly from bulk conditions, requiring thousands of monomers to complete lattices built from monomer and dimer building blocks. Across a range of free energies and interaction rates, we showed that it is remarkably difficult to complete growth of any lattice before nucleation of competing lattices, leaving the systems starved of monomers and kinetically trapped. We further show that, at 50 mM of Gag monomers, suppressing assembly entirely requires very weak interactions, with $K_{\text{D,Hex}}$ weaker than 1.6 mM (6.4 $k_{\text{B}}T$), or weaker than 16 mM (4.1 $k_{\text{B}}T$) if trimer contacts help stabilize growth (Fig. 9). The immature Gag lattice has not evolved to assemble from bulk components, and thus we mimicked the biological roles of cofactors RNA and IP6 to define time-depen-

dent protocols that slowly “activate” Gag monomers into assembly-competent states. Our derived titration rate is quite general, applicable to a variety of self-assembling systems. This titration rate serves two general purposes: 1) designing experiments that control subunit concentration either via titrated release or via concentration of activating cofactors and 2) inverting experimental kinetics to extract protein-protein binding rates. We show that, with this additional timescale, we can keep concentrations low and dramatically improve assembly yield for a range of Gag-Gag free energies and rates, illustrating the power of cofactors in defining both assembly kinetics and yield (Fig. 9).

By providing theoretical justifications for our time-dependent titration protocol, we were able to analyze Gag assembly kinetics measured in vitro and provide here the first estimates of bounds on Gag-IP6 and Gag-Gag binding rates. Light-scattering experiments showed clearly that higher concentrations of IP6 lead to faster Gag assembly kinetics (11); this observation agrees with a model of IP6 activating Gag for fast assembly. We could thus interpret the early growth rates in terms of the $k_{\text{IP6-Gag}}$, estimating a relatively slow binding rate of $125 \text{ M}^{-1} \text{ s}^{-1}$. By also connecting to our microscopic model of titration-limited Gag assembly, we then estimated the Gag hexamer binding constant in a range of $8 \cdot 10^4$ to $4 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$. Thus, once Gag is activated, it can assemble relatively rapidly into the immature lattice. Although the effect of cofactors in our model is reflected in our choice of Gag-Gag binding parameters, cofactors are not explicitly modeled. Thus, we do not capture how one IP6 can stabilize formation of a 6-Gag hexamer. We simplified this stabilization as a bimolecular event with 6:1 stoichiometry in our theoretical analysis, but it most likely builds through stepwise and cooperative molecular interactions. Including cooperativity adds more

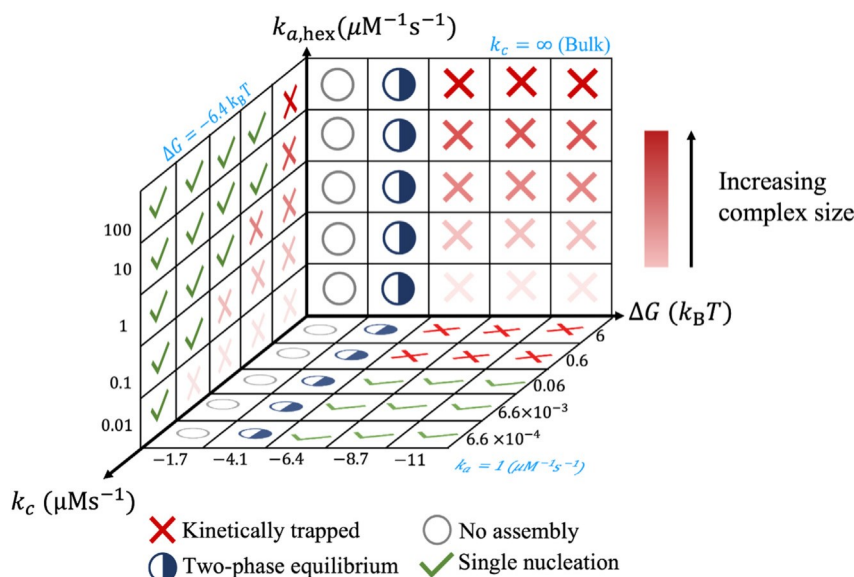


FIGURE 9 Phase diagram summarizes how self-assembled complexes will depend on free energies, association rates, and titration rates. From our simulations, we probed regimes with no assembly (open circles), a phase separated equilibrium (half circles), kinetically trapped systems (red X⁰), and completed assemblies (green checks). With no assembly possible at weak ΔG_{hex} , the stabilization of these subunit contacts moves from phase separated (a dilute monomer phase and a distribution of lattice sizes) to kinetically trapped, bracketing the narrow (and here unaccessed) region of ΔG_{hex} where optimal high-yield assembly is achieved (facing grid). Increasing the on and off-rates (z axis) produces larger intermediates on average when kinetic trapping stalls growth (facing grid, $k_c \propto N$, bulk). By titrating in components with a slow rate k_c , kinetically trapped intermediates can be circumvented to achieve complete assemblies (axis out of the page). Titration does not affect equilibrium states (no assembly, phase-separated equilibrium), as we reach the same concentrations here. The bottom grid has

a fixed association rate $k_a \propto 1 \text{ mM}^{-1} \text{ s}^{-1}$ (blue text). Faster titration rates can still circumvent traps if the association rates for subunit binding are fast (left grid, with $\Delta G_{\text{hex}} \approx 6.4 k_{\text{B}}T$). To see this figure in color, go online.

parameters that lack experimentally determined values, but our simulations are relatively efficient and can be further informed by molecular simulation. Recent coarse-grained MD simulations support that IP6 can accelerate Gag-Gag interactions, capturing coordination of one IP6 to a hexamer (35). These simulations similarly found evidence of kinetic trapping with IP6 (35), consistent with our finding above that fast activation of Gag by high concentrations of IP6 results in assembly that shows signs of kinetic trapping. While it will be important in future work to explicitly capture IP6 binding and its conditional influence on Gag interactions, our results here already suggest parameters for the interactions with and without IP6 present.

For assembly of HIV immature lattices in cells, both RNA (45) and the plasma membrane (37,46) provide additional means to temporally and conformationally control Gag activation. These are natural future extensions of our model. RNA can stimulate slower activation by introducing yet another timescale of RNA-Gag binding, which we could not quantify here because, in the *in vitro* experiments, the kinetic assays start after Gag and RNA were seeded together (11). RNA-driven activation could be important for productive assembly, as cellular IP6 concentrations are high, and we showed above that if Gag is activated too quickly then it can easily become kinetically trapped. RNA can also act to tether together components, effectively localizing and concentrating components (47), with the viral genomic RNA outcompeting cellular RNA in promoting assembly (48). Membrane binding not only helps conformationally prime Gag for immature assembly (46), it also protects Gag from degradation, with >85% of the Gag produced in the cytoplasm degraded before adsorbing to the surface (28). Restriction to the plasma membrane will fundamentally concentrate components via dimensional reduction to promote assembly (49), and the localization process introduces additional timescales that can be theoretically predicted (42) to similarly “titrate” the concentration. For the immature Gag lattice, the multiple sources of temporal and localized control most likely provide the robustness of assembly that is needed for such a large structure, protecting it against kinetic trapping and irregular growth.

Overall, our model simultaneously provides structural and kinetic details on the pathways of assembly as controlled by distinct Gag binding sites, further providing a means of quantifying *in vitro* kinetic measurements of Gag lattice assembly as stimulated by cofactors. Similar modeling studies of clathrin lattice nucleation and growth on membranes (50) and the immature Gag lattice dynamics following budding from the membrane using the same model here (51) demonstrate how this approach can be extended and provide quantitative connections to experimental kinetics. With this same model of the immature Gag lattice, we recently predicted (51) via comparison with experimental data in budded virions that activated

Gag hexamer contacts would be in a range of 8 to 10 k T, with rates above $10^4 \text{ M}^{-1} \text{ s}^{-1}$, or 10 to 12 k T with rates above $10^5 \text{ M}^{-1} \text{ s}^{-1}$, which is fully consistent with our model findings here. These advantages allow us to examine assembly kinetics here across the lifespan of immature Gag, significantly exceeding the timescales of previous computational studies of immature Gag lattices. Ultimately, simulations of self-assembly, whether using rate-based approaches (52) like here or coarse-grained energy functions (53,54), are critical to understand the dynamics of this nonequilibrium process that must proceed from unbound populations to functional assemblies, whether of viral capsids (18–20,54,55), nanoparticle assemblies (56), immature HIV lattices (35–37,51), or mature HIV capsids (57,58). Simulations can be used to construct simplified kinetic networks through diverse intermediates (59,60), allowing for design of specific pathways to reach target structures (56). Coupling to cellular factors in simulations has in several cases illustrated increased robustness to assembly (31,37,50), with modeling approaches thus offering both general and system-specific guides to interpret and integrate with further quantitative experiments.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2023.06.021>.

AUTHOR CONTRIBUTIONS

M.E.J. designed the research. Y.Q., B.M., and D.E. performed simulations. Y.Q., D.E., B.M., and Y.F. developed the model. Y.Q., D.E., Y.F., S.G., and Z.H.L. developed analysis tools. Y.Q. and D.E. generated figures. Y.Q., D.E., and M.E.J. wrote the manuscript. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

M.E.J. gratefully acknowledges funding from an NSF CAREER Award 1753174. Our work used the rockfish cluster of the ARCH supercomputing system at Johns Hopkins University, supported by NSF MRI 1920103. We are grateful to Prof. Owen Pornillos for kindly providing his data and discussing the experimental conditions.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Freed, E. O., and A. J. Mouland. 2006. The cell biology of HIV-1 and other retroviruses. *Retrovirology*. 3:77.
2. Schur, F. K. M., W. J. H. Hagen, . , J. A. G. Briggs. 2015. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature*. 517:505–508.
3. Tan, A., A. J. Pak, . , J. A. G. Briggs. 2021. Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as

- potential substrates for proteolytic maturation. *Proc. Natl. Acad. Sci. USA*. 118, e2020054118.
4. Pettit, S. C., L. E. Everitt, . , A. H. Kaplan. 2004. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J. Virol.* 78:8477–8485.
5. Sundquist, W. L., and H. G. Krusslich. 2012. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* 2:a006924.
6. Mallery, D. L., K. M. R. Faysal, . , L. C. James. 2019. Cellular IP6 levels limit HIV production while viruses that cannot efficiently package IP6 are attenuated for infection and replication. *Cell Rep.* 29:3983–3996.e4.
7. Mallery, D. L., A. B. Kleinpeter, . , L. C. James. 2021. A stable immature lattice packages IP6 for HIV capsid maturation. *Sci. Adv.* 7, eabe4716.
8. Keller, P. W., C. S. Adamson, . , A. C. Steven. 2011. HIV-1 maturation inhibitor bevirimat stabilizes the immature Gag lattice. *J. Virol.* 85:1420–1428.
9. Waheed, A. A., and E. O. Freed. 2012. HIV type 1 Gag as a target for antiviral therapy. *AIDS Res. Hum. Retroviruses*. 28:54–75.
10. Bush, D. L., and V. M. Vogt. 2014. In vitro assembly of retroviruses. *Annu. Rev. Virol.* 1:561–580.
11. Kucharska, I., P. Ding, . , O. Pornillos. 2020. Biochemical reconstitution of HIV-1 assembly and maturation. *J. Virol.* 94, e01844-19.
12. Campbell, S., and V. M. Vogt. 1997. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J. Virol.* 71:4425–4435.
13. Campbell, S., R. J. Fisher, . , A. Rein. 2001. Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc. Natl. Acad. Sci. USA*. 98:10875–10879.
14. Wagner, J. M., K. K. Zadrozny, . , O. Pornillos. 2016. Crystal structure of an HIV assembly and maturation switch. *Elife*. 5:e17063.
15. Datta, S. A. K., Z. Zhao, . , A. Rein. 2007. Interactions between HIV-1 Gag molecules in solution: an inositol phosphate-mediated switch. *J. Mol. Biol.* 365:799–811.
16. Schur, F. K. M., M. Obr, . , J. A. G. Briggs. 2016. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science*. 353:506–508.
17. Zlotnick, A. 1994. To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* 241:59–67.
18. Zlotnick, A., J. M. Johnson, . , D. Endres. 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry*. 38:14644–14652.
19. Hagan, M. F. 2014. Modeling viral capsid assembly. *Adv. Chem. Phys.* 155:1–68.
20. Mohajerani, F., B. Tyukodi, . , M. F. Hagan. 2022. Multiscale modeling of hepatitis B virus capsid assembly and its dimorphism. *ACS Nano*. 16:13845–13859.
21. Hagan, M. F., and O. M. Elrad. 2010. Understanding the concentration dependence of viral capsid assembly kinetics—the origin of the lag time and identifying the critical nucleus size. *Biophys. J.* 98:1065–1074.
22. Mohajerani, F., E. Sayer, . , M. F. Hagan. 2021. Mechanisms of scaffold-mediated microcompartment assembly and size control. *ACS Nano*. 15:4197–4212.
23. Liu, Y., and X. Zou. 2019. A new model system for exploring assembly mechanisms of the HIV-1 immature capsid in vivo. *Bull. Math. Biol.* 81:1506–1526.
24. Gartner, F. M., I. R. Graf, and E. Frey. 2022. The time complexity of self-assembly. *Proc. Natl. Acad. Sci. USA*. 119, e2116373119.
25. Hagan, M. F., O. M. Elrad, and R. L. Jack. 2011. Mechanisms of kinetic trapping in self-assembly and phase transformation. *J. Chem. Phys.* 135, 104115.
26. Kalikmanov, V. I. 2013. Nucleation theory. In *Lecture Notes in Physics* Springer, p. 316.
27. Kasai, M., S. Asakura, and F. Oosawa. 1962. The cooperative nature of G-F transformation of actin. *Biochim. Biophys. Acta*. 57:22–31.
28. Tritel, M., and M. D. Resh. 2000. Kinetic analysis of human immunodeficiency virus type 1 assembly reveals the presence of sequential intermediates. *J. Virol.* 74:5845–5855.
29. Baschek, J. E., H. C. R. Klein, and U. S. Schwarz. 2012. Stochastic dynamics of virus capsid formation: direct versus hierarchical self-assembly. *BMC Biophys.* 5:22.
30. Boettcher, M. A., H. C. R. Klein, and U. S. Schwarz. 2015. Role of dynamic capsomere supply for viral capsid self-assembly. *Phys. Biol.* 12, 016014.
31. Lazaro, G. R., and M. F. Hagan. 2016. Allosteric control of icosahedral capsid assembly. *J. Phys. Chem. B*. 120:6306–6318.
32. von Smoluchowski, M. 1917. Attempt to derive a mathematical theory of coagulation kinetics in colloidal solutions. *Z. Phys. Chem.* 92:129.
33. Rice, S. A. 1985. Diffusion limited reactions. In *Comprehensive Chemical Kinetics* Elsevier Science and Technology.
34. Szabo, A., K. Schulten, and Z. Schulten. 1980. First passage time approach to diffusion controlled reactions. *J. Chem. Phys.* 72:4350–4357.
35. Pak, A. J., M. Gupta, . , G. A. Voth. 2022. Inositol hexakisphosphate (IP6) accelerates immature HIV-1 Gag protein assembly toward kinetically trapped morphologies. *J. Am. Chem. Soc.* 144:10417–10428.
36. Ayton, G. S., and G. A. Voth. 2010. Multiscale computer simulation of the immature HIV-1 virion. *Biophys. J.* 99:2757–2765.
37. Pak, A. J., J. M. A. Grime, . , G. A. Voth. 2017. Immature HIV-1 lattice assembly dynamics are regulated by scaffolding from nucleic acid and the plasma membrane. *Proc. Natl. Acad. Sci. USA*. 114:E10056–E10065.
38. Varga, M. J., Y. Fu, . , M. E. Johnson. 2020. NERDSS: A nonequilibrium simulator for multibody self-assembly at the cellular scale. *Biophys. J.* 118:3026–3040.
39. Johnson, M. E., and G. Hummer. 2014. Free-propagator reweighting integrator for single-particle dynamics in reaction-diffusion models of heterogeneous protein-protein interaction systems. *Phys. Rev. X*. 4, 031037.
40. Johnson, M. E. 2018. Modeling the self-assembly of protein complexes through a rigid-body rotational reaction-diffusion algorithm. *J. Phys. Chem. B*. 122:11771–11783.
41. Johnson, M. E., A. Chen, . , A. M. Uhrmacher. 2021. Quantifying the roles of space and stochasticity in computer simulations for cell biology and cellular biochemistry. *Mol. Biol. Cell*. 32:186–210.
42. Mishra, B., and M. E. Johnson. 2021. Speed limits of protein assembly with reversible membrane localization. *J. Chem. Phys.* 154, 194101.
43. Dick, R. A., K. K. Zadrozny, . , V. M. Vogt. 2018. Inositol phosphates are assembly co-factors for HIV-1. *Nature*. 560:509–512.
44. Endres, D., and A. Zlotnick. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* 83:1217–1230.
45. Rein, A., S. A. K. Datta, . , K. Musier-Forsyth. 2011. Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* 36:373–380.
46. Datta, S. A. K., F. Heinrich, . , H. Nanda. 2011. HIV-1 Gag extension: conformational changes require simultaneous interaction with membrane and nucleic acid. *J. Mol. Biol.* 406:205–214.
47. Jouvenet, N., S. M. Simon, and P. D. Bieniasz. 2009. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci. USA*. 106:19114–19119.
48. Comas-Garcia, M., S. A. Datta, . , A. Rein. 2017. Dissection of specific binding of HIV-1 Gag to the ‘packaging signal’ in viral RNA. *Elife*. 6, e27055.

Qian et al.

49. Yogurtcu, O. N., and M. E. Johnson. 2018. Cytosolic proteins can exploit membrane localization to trigger functional assembly. *PLoS Comput. Biol.* 14:e1006031.
50. Guo, S.-K., A. J. Sodt, and M. E. Johnson. 2022. Large self-assembled clathrin lattices spontaneously disassemble without sufficient adaptor proteins. *PLoS Comput. Biol.* 18:e1009969.
51. Guo, S., I. Saha, . , M. E. Johnson. 2022. Defects in the HIV immature lattice support essential lattice remodeling within budded virions. Preprint at bioRxiv. <https://www.biorxiv.org/content/10.1101/2022.11.21.517392v1>.
52. Sweeney, B., T. Zhang, and R. Schwartz. 2008. Exploring the parameter space of complex self-assembly through virus capsid models. *Biophys. J.* 94:772–783.
53. Hagan, M. F., and D. Chandler. 2006. Dynamic pathways for viral capsid assembly. *Biophys. J.* 91:42–54.
54. Perlmutter, J. D., M. R. Perkett, and M. F. Hagan. 2014. Pathways for virus assembly around nucleic acids. *J. Mol. Biol.* 426:3148–3165.
55. Timmermans, S. B. P. E., A. Ramezani, . , R. Zandi. 2022. The dynamics of viruslike capsid assembly and disassembly. *J. Am. Chem. Soc.* 144:12608–12612.
56. Zeng, X., B. Li, . , X. Huang. 2016. Elucidating dominant pathways of the nano-particle self-assembly process. *Phys. Chem. Chem. Phys.* 18:23494–23499.
57. Grime, J. M. A., and G. A. Voth. 2012. Early stages of the HIV-1 capsid protein lattice formation. *Biophys. J.* 103:1774–1783.
58. Grime, J. M. A., J. F. Dama, . , G. A. Voth. 2016. Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.* 7, 11568.
59. Perkett, M. R., and M. F. Hagan. 2014. Using Markov state models to study self-assembly. *J. Chem. Phys.* 140, 214101.
60. Yang, Y. I., and Y. Q. Gao. 2015. Computer simulation studies of Ab37-42 aggregation thermodynamics and kinetics in water and salt solution. *J. Phys. Chem. B.* 119:662–670.