

Simulated Self-Assembly of the HIV-1 Capsid: Protein Shape and Native Contacts Are Sufficient for Two-Dimensional Lattice Formation

Bo Chen and Robert Tycko*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

ABSTRACT We report Monte Carlo simulations of the initial stages of self-assembly of the HIV-1 capsid protein (CA), using a coarse-grained representation that mimics the CA backbone structure and intermolecular contacts observed experimentally. A simple representation of N-terminal domain/N-terminal domain and N-terminal domain/C-terminal domain interactions, coupled with the correct protein shape, is sufficient to drive formation of an ordered lattice with the correct hexagonal symmetry in two dimensions. We derive an approximate concentration/temperature phase diagram for lattice formation, and we investigate the pathway by which the lattice develops from initially separated CA dimers. Within this model, lattice formation occurs in two stages: 1), condensation of CA dimers into disordered clusters; and 2), nucleation of the lattice by the appearance of one hexamer unit within a cluster. Trimers of CA dimers are important early intermediates, and pentamers are metastable within clusters. Introduction of a preformed hexamer at the beginning of a Monte Carlo run does not directly seed lattice formation, but does facilitate the formation of large clusters. We discuss possible connections between these simulations and experimental observations concerning CA assembly within HIV-1 and *in vitro*.

INTRODUCTION

The capsid (CA) shell, assembled from multiple copies of one or more CA proteins, is an essential structural component of all viruses that protects the viral genome from adverse environments and is tightly associated with the viral life cycle (1). Viruses with abnormal capsids exhibit low infectivity (1–4), making CA proteins potential antiviral drug targets (5–7). The basic principles governing the construction of symmetric capsids were elucidated almost 50 years ago (8), and the structures of many viral capsids have since been determined (9). However, the mechanisms of capsid formation remain unclear due to the fact that CA self-assembly is a rapid, dynamic process that is difficult to monitor in molecular-level detail by experiments. As an alternative, simulations based on experimental observations can contribute to our understanding of the self-assembly process. As CA self-assembly involves a large number of proteins, simulations of this process with a full atomic representation of each CA protein are not currently feasible, although all-atom simulations of the dynamics within a preassembled virus have been reported (10). Therefore, CA self-assembly simulations rely principally on coarse-grained (CG) models (11–23) or kinetic models based on differential equations (24–28). Theoretical analyses of CA geometries (29–32), energy landscapes for self-assembly (33,34), and self-assembly thermodynamics (25,35) have also been reported. To our knowledge, a high-resolution visualization of the initial stages of the self-assembly process has not been reported. Such visualization requires accurate represen-

tations of both the molecular structure and the intermolecular interactions of CA proteins.

The 25.6-kDa CA protein of HIV-1 consists of N-terminal and C-terminal domains (NTD and CTD) joined by a three-residue linker segment (36–40). The 146-residue NTD includes a β -hairpin and seven α -helices (37,38). The 82-residue CTD includes one 3_{10} -helix and four α -helices and contains a dimerization surface that causes both the CTD and the full-length CA to exist predominantly as dimers in solution at low ionic strength (39–41). The CA forms a conical shell within mature HIV-1, with dimensions of ~ 120 nm \times 60 nm (42,43). Other shapes within mature HIV-1 are also observed (43). CA shells within HIV-1 contain ~ 1500 CA molecules. *In vitro*, recombinant HIV-1 CA self-assembles, at high ionic strength or upon addition of crowding agents, into various morphologies, including cones, tubes, and spheres (44–46). CA tubes and planar assemblies have been shown to be comprised of a two-dimensional (2D) triangular lattice of CA hexamers, usually called a hexagonal lattice (47–49). Within the hexagonal lattice, groups of six NTD units form the interiors of hexagons, which are joined by CTD-CTD dimerization. It is proposed that closed structures such as cones and spheres are constructed from the same lattice, but with incorporation of precisely 12 CA pentamers (42). The structures of engineered CA hexamers have been determined by x-ray crystallography (50), and detailed models for the hexagonal CA lattice have been developed by combining cryoelectron microscopy and electron diffraction data for CA lattices (48,49) with NTD and CTD structures from x-ray crystallography and NMR (36–40,49). Solid-state NMR studies of self-assembled HIV-1 CA provide further evidence that the NTD and CTD structures are retained within tubular

Submitted November 19, 2010, and accepted for publication May 12, 2011.

*Correspondence: robertty@mail.nih.gov

Editor: Benoit Roux.

© 2011 by the Biophysical Society
0006-3495/11/06/3035/10 \$2.00

doi: 10.1016/j.bpj.2011.05.025

(51) and nontubular (52) structures, and that both domains are largely immobilized in the CA lattice.

The process by which HIV-1 CA self-assembles is crucial to our understanding of both *in vivo* and *in vitro* structures. In this article, we present a CG model for HIV-1 CA that exhibits self-assembly into a hexagonal lattice in 2D Monte Carlo (MC) simulations. We focus on the 2D hexagonal lattice (rather than a closed shell that presumably includes pentamers), because such a lattice exists in HIV-1 tubes and planar assemblies that have been studied experimentally and constitutes most of the area of the conical capsids within mature HIV-1, and because hexagonal lattice formation represents the first test of our CG model. Our model includes the secondary and tertiary structure of HIV-1 CA, motivated by the hypothesis that the evolved shape of CA plays a crucial role in its self-assembly properties. Our model also includes a simple representation of attractive intermolecular interactions that derives from experimentally determined intermolecular contacts. Using this model, we characterize the stabilities of various CA oligomers, computationally derive an approximate concentration/temperature phase diagram for 2D lattice formation, and investigate the pathway by which an ordered lattice develops from an initial state in which CA dimers are well separated.

METHODS

Description of the CG model

We represent HIV-1 CA by a bundle of solid cylinders, with the position, orientation, radius, and length of each cylinder adjusted to enclose the α -carbons of one helical segment of the protein, based on coordinates in Protein Data Bank (PDB) 3H47 (50) (Fig. 1 A). Since helix 7 in the NTD has a significant bend, we represent this helix by two cylinders, called 7a and 7b. Since CA molecules in solution exist as dimers (40,44,49), we use the twofold symmetric dimer structure (Fig. 1 B) as the basic unit in MC simulations, with geometry derived from PDB 3H47. Our own dynamic light scattering data indicate no detectable CA monomer at concentrations

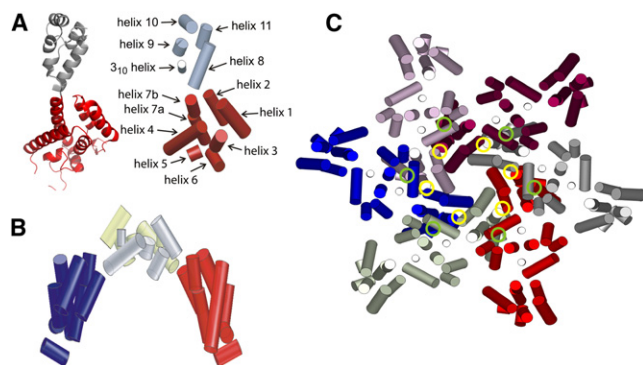


FIGURE 1 (A) Comparison of the full HIV-1 CA structure (*left*) with the CG model (*right*), with NTD and CTD shaded dark and light (*red* and *gray* on-line), respectively. (B) CG model for CA dimer, with NTDs in different shades (*blue* and *red* on-line). (C) A hexamer formed by six CA dimers, each in a different shade. Circles indicate sites of intermolecular NTD-NTD and NTD-CTD interactions that are included in the CG model.

as low as 6 μ M in 50 mM Tris buffer, pH 8, 22°C. We neglect all dynamics within a dimer, consistent with solid-state NMR evidence that both the NTD and the CTD are immobilized in the final CA lattice (51,52). As discussed below, hinge motions about the short linker between NTD and CTD, suggested as a source of pentamer/hexamer quasiequivalence (49,50), may be included in future studies. Solution NMR spectra of full-length CA dimers reported by Byeon et al. (49) show pronounced line broadening of NTD crosspeaks (compared with spectra of the isolated NTD), suggesting that NTD/CTD hinge motions are restricted in the unassembled state.

Our MC simulations include hard-wall intermolecular repulsions among all cylinders (algorithm described in the Supporting Material) and a small number of attractive intermolecular interactions (four sites per CA monomer). By examining PDB file 3H47, we find that the shortest intermolecular NTD-NTD distances within a CA hexamer are between helices 2 and 3, whereas the shortest intermolecular NTD-CTD distances are between helices 4 and 8. Therefore, we define attractive intermolecular interactions between the corresponding cylinders of CA molecules in different dimers, indicated by green and yellow circles in Fig. 1 C. The intermolecular potential energy is expressed as a sum of terms with the form

$$V(\theta, r) = f(\theta) \times g(r) \quad (1a)$$

$$f(\theta) = \begin{cases} \frac{|\theta - \theta_0|}{\theta_{\max}} - 1, & |\theta - \theta_0| \leq \theta_{\max} \\ 0, & |\theta - \theta_0| > \theta_{\max} \end{cases} \quad (1b)$$

$$g(r) = 4\epsilon_0 \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (1c)$$

$f(\theta)$ is an angular weighting function that represents the orientation dependence of CA interactions, where θ is the angle between cylinders, θ_0 is the angle in the experimental CA hexamer structure (50), and θ_{\max} is a cutoff angle (set to $2\pi/3$ in our simulations) beyond which attractive interactions are zero. $g(r)$ is a Lennard-Jones potential, where r is the distance between the centers of the interacting cylinders, and σ (which determines the minimum-energy distance $r = 1.1225\sigma$) is set to match helix-helix distances in the hexamer structure. For NTD-NTD interactions (involving cylinders 2 and 3), $\sigma = 0.90673$ nm, $\epsilon_0 = 2.0$ kcal/mol, and $\theta_0 = 39.9^\circ$. For NTD-CTD interactions (involving cylinders 4 and 8), $\sigma = 17.1305$ nm, $\epsilon_0 = 2.0$ kcal/mol, and $\theta_0 = 11.7^\circ$.

CA self-assembly is affected by ionic strength and pH (44,53), both of which modulate intermolecular interactions. In our model, the minimum interaction energy is $-4\epsilon_0/\text{CA monomer}$, since each monomer can interact via two NTD-NTD and two NTD-CTD contacts. To capture the dependence on solution conditions in our simulations, we varied the dimensionless interaction strength $\epsilon' \equiv 4\epsilon_0/RT$ by varying the temperature T .

Simulations

We performed MC simulations in 2D, with the center of mass (COM) of each CA dimer restricted to a square with periodic boundary conditions and with the twofold symmetry axis of each dimer perpendicular to the square. In each MC step, a translational and rotational move was applied to one CA dimer (chosen sequentially) and accepted or rejected according to the standard Metropolis scheme (54) using the potential energy described above. Translational moves were randomly chosen between 0.0 and 0.1 nm, in random directions. Rotational moves were randomly chosen between -10° and $+10^\circ$ about the two-fold symmetry axis. Typical acceptance probabilities were 25%. For simplicity, MC simulations were used instead of full molecular dynamics (16,17,21), discontinuous molecular dynamics (13,14,22) or Brownian dynamics (11,12,18,23). It has been shown that MC simulations with small moves are equivalent to Brownian dynamics (55,56). Moreover, a more elaborate treatment of the dynamics is unwarranted, given the approximations in our CG model (e.g., the

absence of internal motions and solvent). By equating the translational diffusion constant for a single dimer in our MC simulations with the experimental value from dynamic light scattering ($4.47 \times 10^{-7} \text{ cm}^2/\text{s}$ at 22°C), we estimate that one MC step per CA dimer corresponds to ~ 100 ps of real time.

To investigate the stability of putative intermediates in the self-assembly process, CA oligomers shown in Fig. 2 were generated by simulated annealing, starting with the appropriate number of CA dimers at random positions. Two distinct configurations were identified for groups of three, four, and six CA dimers, called trimer1, trimer2, tetramer1, tetramer2, hexamer1, and hexamer2. One thousand constant-temperature MC simulations were then performed on each oligomer until it dissociated. An oligomer was considered to have dissociated when the interaction energy of any dimer with its closest neighbor became $> -0.005\epsilon_0$ and the distance between their COMs became > 10 nm. Fig. 2 I shows an example of a dissociated hexamer2.

In MC simulations for isolated dimer, trimer2, and hexamer2 units, the root mean-squared displacement is proportional to the square root of the number of MC steps per CA dimer, as expected for diffusional motion, with diffusion constants nearly proportional to $n^{-1/2}$, where n is the oligomer size (Fig. S1 in the Supporting Material).

To investigate lattice formation, constant-temperature MC simulations were performed on sets of 36 CA dimers in (periodic) square areas that were varied to produce variations in protein concentration. Starting with CA dimers on a square lattice, the system was heated to 10,000 K for 10^5 MC steps to randomize positions and orientations before the temperature was set to the desired value. To monitor the self-assembly mechanism, multiple simulations were performed for 2.6×10^6 steps/CA dimer, sufficient to observe the nucleation and growth of the hexagonal lattice, at various CA concentrations and temperatures. COM coordinates and orientations were stored every 5 s (~ 170 steps/dimer), and movies were generated from the stored configurations at 10 frames/s (see Supporting Material). To construct the phase diagram, simulations were run for 7 days ($\sim 2 \times 10^7$ steps/dimer), and the system configuration was stored every 600 s.

Simulation programs were coded in Fortran95 and run on the National Institutes of Health Biowulf cluster, with 2.8-GHz AMD Opteron CPUs. Figures and movies were generated in MATLAB (The MathWorks, Natick, MA), with the exception of Fig. 1 A, which was created with PyMOL.

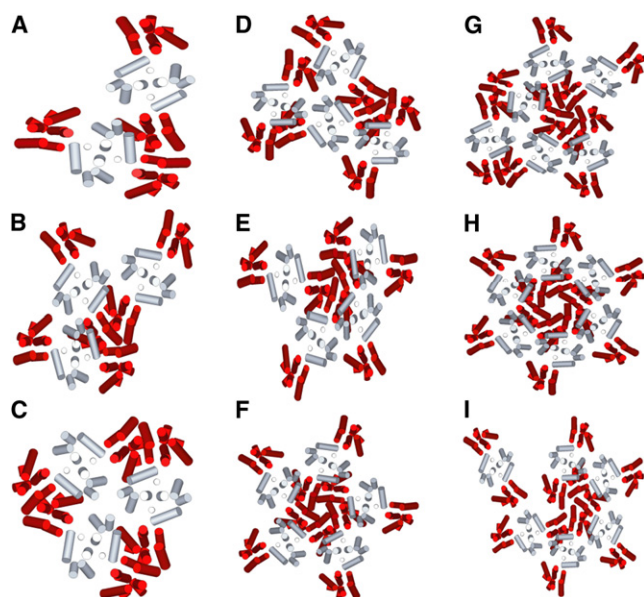


FIGURE 2 CA oligomer configurations identified by simulated annealing. (A–F) Dimer, trimer1, trimer2, tetramer1, tetramer2, pentamer, hexamer1, and hexamer2 configurations, respectively, with the oligomerization number being the number of CA dimers. (I) Dissociated hexamer2.

RESULTS

Stability of CA oligomers

For each annealed oligomer configuration in Fig. 2, 1000 dissociation simulations were performed at fixed temperatures. Fig. 3, A and B, shows plots of the percentage of simulations in which the dissociation time (see above) was less than a given value, in MC steps per CA dimer, at various temperatures. As expected, the average lifetime, τ , of each oligomer (defined as the time by which 50% of the simulations had reached dissociation) decreased with increasing temperature. Fig. 3 C shows that $\ln\tau$ is approximately proportional to $1/T$ for each oligomer, allowing activation energies, E_a , for dissociation of each oligomer to be determined from an Arrhenius law, $\tau = \tau_0 \exp(E_a/RT)$. As shown in Fig. 4 D, oligomer configurations fall into two groups. The more stable group, with $E_a > 4.5$ kcal/mol, includes hexamer2, pentamer, and trimer2 (Fig. 2, C, F, and H). The less stable group, with $E_a < 3.6$ kcal/mol, includes the remaining configurations. It is interesting that the more stable oligomers have structures that appear in the hexagonal CA lattice (hexamer2 and trimer2) or in models for closed CA shells (hexamer2, pentamer, and trimer2). Each CA dimer in the more stable oligomers interacts with two neighboring dimers, compared with 1–1.5 neighboring dimers in the less stable oligomers.

This stability analysis predicts that trimer2 may be an important intermediate configuration in the CA self-assembly process. Since trimer2 units occur in the 2D hexagonal lattice, such intermediates can be on-pathway. Pentamers are also predicted to be important intermediates, which can be either off-pathway or on-pathway depending on whether they tend to dissociate or to convert to hexamers by addition of one more CA dimer. In three-dimensional (3D) simulations, pentamers could persist as on-pathway intermediates.

Assembly kinetics

To investigate CA self-assembly, 2D MC simulations on 36 CA dimers were performed at various temperatures and concentrations. As a measure of self-assembly progress, the time dependence of the number of assembled hexamers was calculated from stored system configurations, using a straightforward geometric algorithm (see Supporting Material). In principle, 36 CA dimers can form a maximum of eight complete hexamers. In practice, fully assembled lattices contained six to eight hexamers in our simulations.

Fig. 4 A shows the time dependence of lattice formation in a representative run at 300 K ($\epsilon' = 13.4$) and at a 2D concentration ($[\text{CA}]_{2D}$) of 0.016 nm^{-2} (same average intermolecular distance as $[\text{CA}] = 2.38 \text{ mM}$ in 3D). For comparison, the experimentally determined hexagonal lattice structure has $[\text{CA}]_{2D} = 0.081 \text{ nm}^{-2}$ (48). Fig. 4 B shows

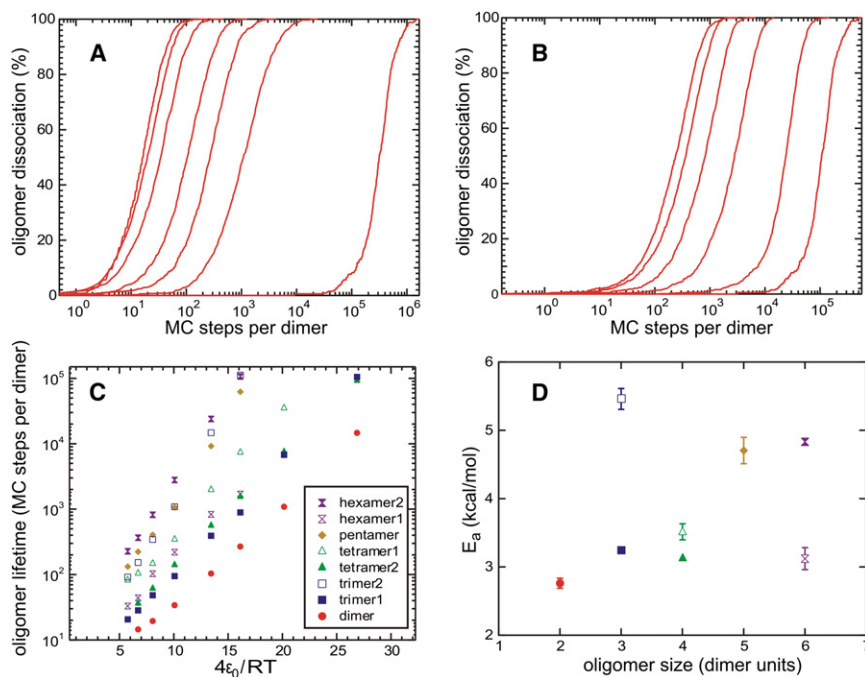


FIGURE 3 (A) Time dependence of dissociation probability for a dimer of CA dimers. Curves from left to right are from simulations at $\epsilon' = 6.72, 8.06, 10.08, 13.44, 16.12, 20.16,$ and 26.87 . (B) Time dependence of dissociation probability for a hexamer2 of CA dimers. Curves from left to right are from simulations at $\epsilon' = 5.76, 6.72, 8.06, 10.08, 13.44,$ and 16.12 . (C) Lifetime versus ϵ' for oligomer configurations in Fig. 2. (D) Activation energies for dissociation, extracted from C.

the initial randomized configuration of the system. After starting the self-assembly simulation by switching to 300 K, CA dimers form dynamically disordered clusters during a short lag period, until one hexamer appears within a cluster of 14 dimers (Fig. 4 C). Additional hexamers appear in a stepwise manner as the cluster grows, reaching a 2D lattice with 5 ± 1 complete hexamers within $\sim 5 \times 10^5$ MC steps/dimer. The structure at the boundaries of the lattice continues to fluctuate. Fig. 4 D shows a typical configuration in this

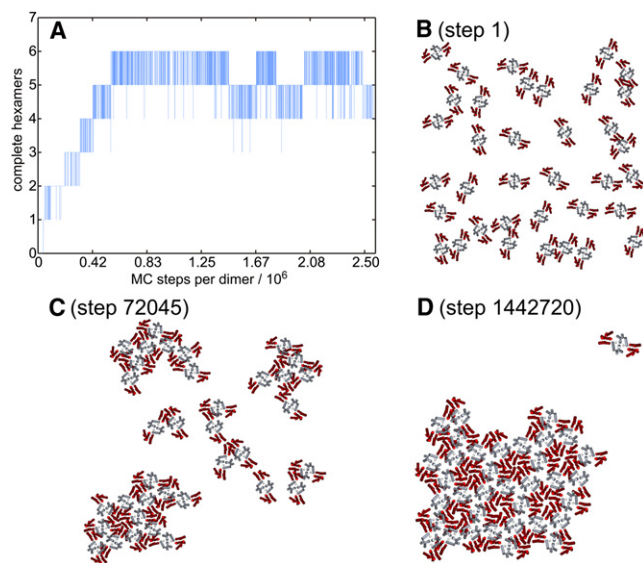


FIGURE 4 (A) Progress curve for self-assembly of 36 CA dimers at $\epsilon' = 13.44$ and $[CA]_{2D} = 0.016 \text{ nm}^{-2}$. (B–D) Configurations after indicated numbers of MC steps per dimer.

asymptotic equilibrium state. The successful assembly of CA dimers into hexagonal lattice shows that our CG model captures the essential structural features of the HIV-1 CA ground state in 2D.

The dependence of self-assembly kinetics on CA dimer concentration is shown in Fig. 5 A, which compares progress curves at 300 K for $[CA]_{2D} = 0.016 \text{ nm}^{-2}$ and $[CA]_{2D} = 0.052 \text{ nm}^{-2}$. Each curve is the average of 10 independent MC runs, each of which extends to 2.6×10^6 MC steps/dimer. On average, higher concentration leads to a shorter lag period and a more rapid initial lattice growth. The longer single runs in Fig. 5 B (to 2.2×10^7 MC steps/dimer) show that the lower concentration allows lattice formation to proceed nearly to completion, whereas the higher concentration can lead to a state in which lattice formation is blocked by congestion.

The dependence of self-assembly kinetics on ϵ' is shown in Fig. 5 C, which compares progress curves at 550 K, 300 K, and 150 K, all with $[CA]_{2D} = 0.052 \text{ nm}^{-2}$. Again, each curve is the average of 10 independent runs. At the lowest temperature (largest ϵ'), development of an ordered hexagonal lattice is relatively slow, as disordered clusters of CA dimers develop at the beginning and dissociate slowly. This behavior is analogous to the kinetic trapping discussed in previous simulations of capsid self-assembly (11,13,27). At the highest temperature (smallest ϵ'), strong fluctuations in the number of complete hexamers are evident, due to the relatively weak binding of CA dimers at the boundaries of the growing lattice. The longer single runs in Fig. 5 D show that the relatively high CA dimer concentration can prevent lattice growth, as in Fig. 5 B.

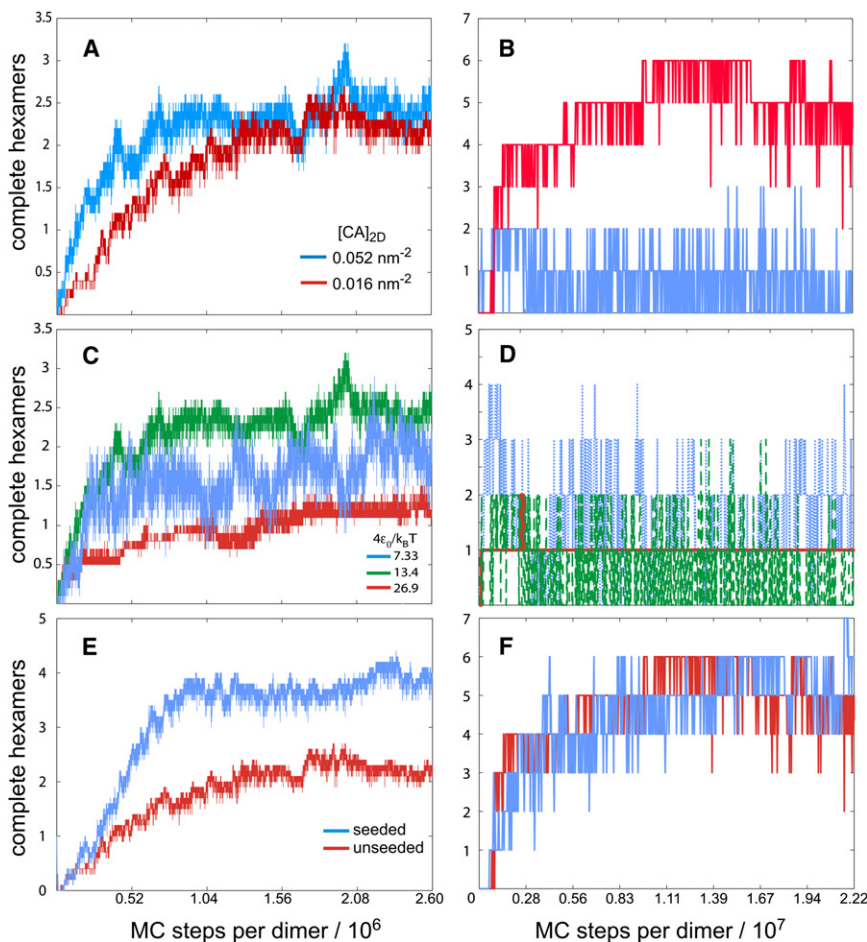


FIGURE 5 Progress curves for self-assembly under various conditions. (A and B) $\epsilon' = 13.4$ and indicated concentrations. (C and D) $[CA]_{2D} = 0.052 \text{ nm}^{-2}$ and indicated ϵ' . (E) $\epsilon' = 13.4$ and $[CA]_{2D} = 0.016 \text{ nm}^{-2}$, with and without a hexamer seed in the initial configuration. Curves in A, C, and E are averages of 10 runs. Curves in B, D, and F are single runs.

Fig. 5 E compares progress curves for seeded simulations, in which a single preformed hexamer of dimers was introduced after the initial randomization, with unseeded simulations. In both cases, $T = 300 \text{ K}$ and $[CA]_{2D} = 0.016 \text{ nm}^{-2}$. Although the hexamer seed dissociates quickly, and therefore does not directly nucleate the hexagonal lattice, the initial stage of lattice growth is more rapid on average in the seeded simulations. We attribute this effect to the earlier appearance of larger clusters of CA dimers in the seeded simulations, which allows a larger number of hexamers to form quickly. The longer single runs in Fig. 5 F show that both seeded and unseeded simulations can approach the maximum lattice size.

Phase diagram for 2D lattice formation

A series of long simulations ($1.7\text{--}2.2 \times 10^7$ MC steps/dimer) was performed over ranges of ϵ' and $[CA]_{2D}$. Fig. 6 A shows the number of hexamers formed in each simulation (averaged over the plateau period of each simulation), which we define to be the assembly efficiency. With $\epsilon' \leq 8$, assembly efficiencies are low because large clusters are unstable even at the highest concentrations. With $\epsilon' > 8$, assembly efficiencies can be high, but the value of $[CA]_{2D}$

that produces the highest efficiency shifts from $[CA]_{2D} > 0.04 \text{ nm}^{-2}$ when $8 < \epsilon' \leq 12.4$ to $[CA]_{2D} < 0.02 \text{ nm}^{-2}$ when $\epsilon' \geq 13.0$. This effect arises from congestion that prevents complete lattice formation within the accessible simulation time at high concentrations and high interaction strengths. At $[CA]_{2D} \approx 0.05 \text{ nm}^{-2}$ ($250 \text{ dimers}/10^4 \text{ nm}^2$ in Fig. 6 A), the drop in assembly efficiency above $\epsilon' = 12.4$ resembles a glass transition (18,19,22).

Fig. 6 B shows the assembly efficiency as a function of ϵ' and $[CA]_{2D}$ in a representation that we interpret as an approximate phase diagram for CA self-assembly in 2D, with blue and green points being conditions where the hexagonal lattice is thermodynamically stable, and red and yellow points being the approximate phase boundary. Black crosses are points where we verified that a preassembled hexagonal lattice dissociates completely in MC simulations. The lattice is stable over a factor of >2 in ϵ' and a factor of >10 in $[CA]_{2D}$.

Self-assembly mechanism

Fig. 7 shows snapshots from a trajectory at $[CA]_{2D} = 0.016 \text{ nm}^{-2}$ and $\epsilon' = 13.4$. As predicted by the oligomer stability analysis (see above and Fig. 3), trimer2 units are

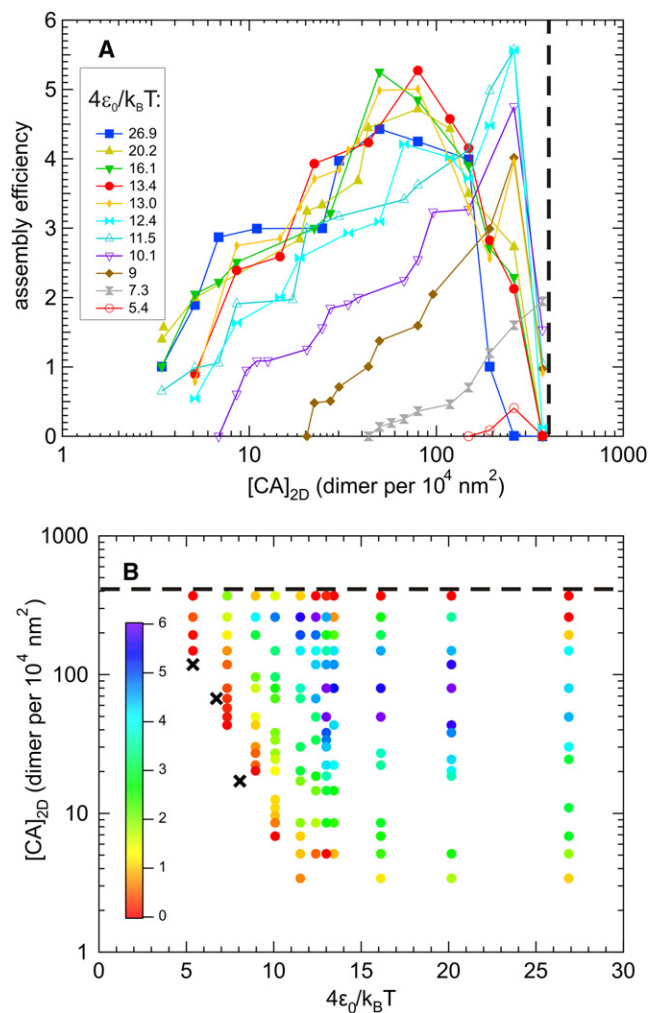


FIGURE 6 (A) Dependence of assembly efficiency on CA concentration at indicated values of ϵ' . (B) Approximate phase diagram for CA self-assembly, with colors representing assembly efficiency. Dashed line is the concentration in the experimentally determined hexagonal lattice (48). Crosses are points where preassembled hexagonal lattices dissociated completely in MC simulations, proving that they are outside the phase boundary.

important metastable species in early stages of the trajectory, whereas other small oligomers are unstable (Fig. 7, A and B). Trimer2 units frequently bind together to form larger clusters. Rearrangements within these clusters can produce pentamers (Fig. 7 C), which persist for variable periods but do not lead directly to an ordered lattice. Eventually, pentamers are disrupted (Fig. 7 D) or convert to a hexamer by insertion of an additional CA dimer unit (Fig. 7 E). Hexamers can also develop within a CA cluster that does not contain any pentamers. Once a single hexamer forms within a CA cluster, it persists indefinitely and nucleates the hexagonal lattice (Fig. 7 F). Occasionally, heptamers are observed at the edge of the lattice (Fig. 7 F), but heptamers are only metastable. The stability of a hexamer within a larger cluster arises from its interactions with additional

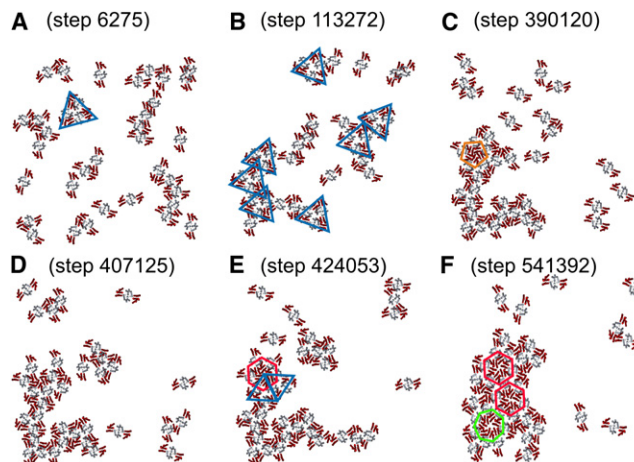


FIGURE 7 Configurations at the indicated steps/CA dimer from a typical MC run at $\epsilon' = 13.4$ and $[CA]_{2D} = 0.016 \text{ nm}^{-2}$. (A and B) Triangles indicate trimer2 configurations, which are important self-assembly intermediates. (C) Pentagon indicates metastable pentamer configuration. (D) Pentamers are disrupted. (E) Trimer2 configurations stabilize hexamers within CA clusters. (F) Heptagon indicates metastable heptamer configuration. Once a single hexamer forms within a cluster, it persists and nucleates additional hexamers.

CA dimers, which result in the formation of trimer2 units around the perimeter of the hexamer (Fig. 7 E, blue triangles). The center of each of these units ultimately becomes a point of threefold symmetry in the hexagonal lattice.

Movie S1 in the Supporting Material contains another trajectory at $[CA]_{2D} = 0.016 \text{ nm}^{-2}$ and $\epsilon' = 13.4$, in which a similar pathway is observed. Movie S2 contains a trajectory from a seeded simulation, also at $[CA]_{2D} = 0.016 \text{ nm}^{-2}$ and $\epsilon' = 13.4$. In this trajectory, the hexamer that is initially present quickly dissociates. The self-assembly mechanism then proceeds as in the unseeded simulation described above.

Movie S3 contains a trajectory at $[CA]_{2D} = 0.016 \text{ nm}^{-2}$ and $\epsilon' = 26.9$. At this lower temperature, disordered clusters of CA dimers persist for longer periods. The clusters form linear chains, which then fold on themselves to form more compact configurations. Rearrangements within clusters are slow, as is dissociation of individual dimers from clusters. Therefore, lattice formation is inefficient and is not observed before the end of the simulation. In this trajectory, one chain folds to form a heptamer, which eventually converts to a single, stable hexamer. Pentamers also develop independently, but they are not stable.

Movie S4 contains a trajectory at $[CA]_{2D} = 0.052 \text{ nm}^{-2}$ and $\epsilon' = 7.3$. At this higher concentration, the system is initially a single, porous cluster. Trimer2 units form frequently within the system, but they are transient. Pentamers also appear transiently. Once one hexamer develops, the hexagonal lattice spreads through the system rapidly. Since the lattice does not fill the available area, CA dimers at the edges continue to dissociate and rebind.

The importance of trimer2 units as intermediates in the self-assembly process is evident from an analysis of multiple trajectories at $[CA]_{2D} = 0.016 \text{ nm}^{-2}$ and $\epsilon' = 13.4$ (Fig. S2), which shows that large trimer2 populations develop consistently at early stages ($\sim 2 \times 10^5$ MC steps/dimer), whereas hexamers develop significantly later ($\sim 2 \times 10^6$ steps/dimer).

DISCUSSION

Principal results

The simulations described above demonstrate that a CG model that captures the overall shape of HIV-1 CA and includes a simple representation of intermolecular interactions to favor native contacts is sufficient to produce self-assembly into a 2D hexagonal lattice. The protein has evolved to form such a lattice. The fact that helices 1–3 of the NTD form a wedge with an angle of $\sim 60^\circ$, which allows six NTDs to form the interior of a hexagon, appears to be one important evolved property. Dimerization through CTD-CTD interactions imparts positions and orientations to six additional NTDs that are appropriate for their participation in six additional hexagons. Thus, the relatively strong dimerization interactions are apparently another evolved property that facilitates capsid formation in HIV-1. As pointed out by Ganser-Pornillos et al. (48), CTD dimerization is not a universal property of retroviral capsid proteins. Moreover, mutations that abolish dimerization interfere with HIV-1 CA self-assembly in vitro and reduce infectivity in vivo (2,4,45). The precise details of other intermolecular interactions may be less important, provided they are not strongly repulsive in a hexagonal lattice configuration. Our observation that lattice formation is relatively efficient over broad ranges of concentration and temperature supports this idea (Fig. 6 B). It has been observed in experiments that HIV-1 CA self-assembly into ordered lattices occurs over ranges of ionic strength and pH and is not prevented by changes in overall charge (57–59).

Before performing MC simulations, we hypothesized that hexamers of CA dimers might act as critical nuclei for lattice formation, as in standard models of nucleated polymerization (24,26,60). In other words, a single hexamer (as in Fig. 2 H) might be sufficiently stable to grow into a lattice by binding additional dimers, whereas smaller oligomers would usually dissociate before growing. The simulations do not support this hypothesis. Instead, CA dimers first coalesce into disordered clusters. Lattice formation is then nucleated by the appearance of a single hexamer within a larger cluster. Isolated hexamers of CA dimers are not stable and do not nucleate lattice formation.

It was surprising to find that trimer2 units (as in Fig. 2 C) are prevalent intermediate structures, both as discrete entities in the early stages of self-assembly and as components of the CA clusters in intermediate stages. Trimer2 configurations

also contribute to the enhanced stability of hexamers within clusters (as in Fig. 7 E). Dimerization through CTD-CTD interactions is essential for the stability of trimer2 configurations, perhaps contributing to the observed effects of CTD mutations on capsid assembly and viral infectivity (2,4,45).

Comparison with experimental observations

Gross et al. (53) estimated the critical concentration for HIV-1 CA self-assembly in vitro to be 2 mg/ml (78 μM) at pH 6 or 1 mg/ml (39 μM) at pH 8 and 1 M NaCl. Del Álamo et al. (57) found a critical concentration of 5.6 μM at pH 7.8 and 2.25 M NaCl. The lower concentration limit of ~ 2 dimers/ 10^4 nm^2 (Fig. 6 B) corresponds to a 3D concentration of 9.4 μM . Interaction energies in our simulations are comparable to values reported for the hepatitis B virus (61) and calculated for other systems (33).

Experimental timescales for CA self-assembly in vitro are 1–10 min (59), corresponding to $6\text{--}60 \times 10^{11}$ MC steps/dimer according to the estimate above. Thus, intermolecular separations in our simulations have values comparable to those seen in experiments, but self-assembly is more rapid by factors of $10^2\text{--}10^6$ (see Fig. 5). This acceleration may be attributed, at least in part, to the restriction of all motions to 2D. In particular, confinement to 2D reduces the importance of orientational and translational entropy and may eliminate local minima in the intermolecular interaction potentials that would otherwise impede self-assembly. It is worth noting that capsid assembly within HIV-1 takes place in a quasi-2D volume, as CA is released from the Gag polyprotein by proteolytic cleavage between the matrix (MA) and CA segments of Gag, starting from a 2D Gag layer that lines the inner surface of the viral membrane. Immature HIV-1 has been estimated to contain ~ 5000 Gag proteins (62). Based on the dimensions of the virus (63), the average 2D concentration of CA near the viral membrane is then $\sim 0.09 \text{ nm}^{-2}$, even higher than in our simulations. Little is known about the kinetics of CA self-assembly within HIV-1, but the timescale may be considerably shorter than that observed in vitro.

Experimental data of Douglas et al. (58) indicate that the in vitro self-assembly rate of wild-type HIV-1 CA at 2.25 M NaCl is roughly proportional to the third power of the protein concentration. These data may be interpreted as evidence that formation of trimers of CA dimers is a rate-limiting step for self-assembly. Direct experimental evidence for trimer2 species under conditions that favor self-assembly has not been reported, but our simulations provide stimulus for future experiments to detect such species, for example by dynamic light scattering, analytical ultracentrifugation, or rapid photochemical cross-linking (64). The importance of CTD-CTD interactions at sites of threefold symmetry in the hexagonal lattice (corresponding to local configurations similar to trimer2) has been emphasized by Byeon et al.

(49). Their cryo-EM data suggest direct interactions between helices 10 and 11. Such interactions are not included as attractive potentials in our current CG model, but could be included in future studies. Trimer2 configurations are stabilized by NTD-NTD and CTD-NTD interactions in our simulations.

Comparison with previous simulations

Previous simulations of capsid self-assembly using CG models have focused primarily on the formation of capsids with icosahedral symmetry in other viruses (11–20,22,23), whereas our simulations focus on the planar hexagonal lattice found in HIV-1 CA assemblies in vitro and as a major component of conical capsids within HIV-1. The representation of HIV-1 CA as two bundles of cylinders with dimensions and geometry that mimic the helices of the NTD and CTD (Fig. 1 A) is an important novel aspect of our CG model that captures the shape of the real protein, allowing intermolecular interactions to be represented by a small number of terms with the simple form in Eq. 1 to favor native contacts. Previous CG models have used representations that do not take into account the details of the protein shape, instead using spheres (18,23), collections of spheres (13,22), or simple shapes such as trapezoids, pentagons, and hexagons (13,19,34). Representations that do not mimic the true protein shape require the imposition of anisotropic interactions that are engineered to produce the desired intermolecular packing (18,23). It is interesting to note that in MC simulations of icosahedral self-assembly by anisotropically interacting spheres, Wilber et al. observed a self-assembly mechanism that resembles the mechanism described above for 2D self-assembly of HIV-1 CA, i.e., icosahedra form within larger disordered clusters (65). An important difference is that the icosahedra subsequently separate from the larger cluster in the simulations of Wilber et al., whereas the 2D cluster crystallizes into a hexagonal lattice in our simulations.

Krishna et al. (21) have performed simulations on HIV-1 capsids using models in which CA dimers (permanently joined, as in our simulations) were represented by four or eight spheres, with intramolecular interactions derived from all-atom simulations and with anisotropic and three-body intermolecular interactions. The starting points for these simulations were preassembled CA tubes or shells in 3D, and the goal was to determine the equilibrium lattice structures for closed capsids. Thus, our simulations address different issues than those addressed by Krishna et al. It is interesting that Krishna et al. found hexagonal lattices at equilibrium (rather than close-packed lattices) only when the NTD was represented by three spheres with the wedge-like shape of the real NTD. A model for CA with additional sites was used by Ayton and Voth in CG simulations of immature HIV-1, in which partial hexagonal order was observed for membrane-associated Gag (66).

Extensions

Future refinements of our CG model may include incorporation of hinge motions between NTD and CTD units and development of more realistic sets of intermolecular interactions. An obvious next step is to extend our simulations to 3D to investigate how lattice curvature develops and is affected by hinge motions (11,49,50), the relative strengths of various intermolecular interactions, and the balance between pentamer and hexamer formation. Questions yet to be answered include Why does HIV-1 CA form tubes rather than jelly rolls? Why do certain mutations or solvent conditions promote self-assembly into closed structures rather than tubes? Are closed lattices nucleated by stable pentamers rather than hexamers? and What limits the number of pentamers to 12?

For computational feasibility, 3D simulations may be performed in a quasi-2D manner, for example by confining CA dimers close to a surface (15), attracting CA dimers to a surface (12), or starting with a preassembled tube or shell (21). Given that CA is released from a membrane-associated Gag lattice by proteolysis during HIV-1 maturation (62,63), such approaches may be biologically relevant. MC simulations with translational and rotational motions in 3D indicate that preformed hexagonal lattice remains stable in 3D for our current model of CA dimer interactions (Fig. S3), although formation of lattice from a disordered initial state has not yet been achieved.

SUPPORTING MATERIAL

Algorithms, three figures, and four movies are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(11\)00595-9](http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)00595-9).

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health, and by the National Institutes of Health Intramural AIDS Targeted Antiviral Program. Computational facilities of the Helix Systems at the National Institutes of Health were used in this work.

REFERENCES

- Göttlinger, H. G., J. G. Sodroski, and W. A. Haseltine. 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA.* 86:5781–5785.
- Abdurahman, S., S. Höglund, ..., A. Vahlne. 2004. Selected amino acid substitutions in the C-terminal region of human immunodeficiency virus type 1 capsid protein affect virus assembly and release. *J. Gen. Virol.* 85:2903–2913.
- von Schwedler, U. K., T. L. Stemmler, ..., W. I. Sundquist. 1998. Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.* 17:1555–1568.
- von Schwedler, U. K., K. M. Stray, ..., W. I. Sundquist. 2003. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.* 77:5439–5450.
- Tang, C., E. Loeliger, ..., M. F. Summers. 2003. Antiviral inhibition of the HIV-1 capsid protein. *J. Mol. Biol.* 327:1013–1020.
- Sticht, J., M. Humbert, ..., H. G. Kräusslich. 2005. A peptide inhibitor of HIV-1 assembly in vitro. *Nat. Struct. Mol. Biol.* 12:671–677.

7. Zhou, J., X. Yuan, ..., C. H. Chen. 2004. Small-molecule inhibition of human immunodeficiency virus type 1 replication by specific targeting of the final step of virion maturation. *J. Virol.* 78:922–929.
8. Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27:1–24.
9. Reddy, V. S., P. Natarajan, ..., J. E. Johnson. 2001. Virus Particle Explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.* 75:11943–11947.
10. Freddolino, P. L., A. S. Arkipov, ..., K. Schulten. 2006. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure.* 14:437–449.
11. Elrad, O. M., and M. F. Hagan. 2008. Mechanisms of size control and polymorphism in viral capsid assembly. *Nano Lett.* 8:3850–3857.
12. Hagan, M. F. 2008. Controlling viral capsid assembly with templating. *Phys. Rev. E.* 77:051904.
13. Nguyen, H. D., and C. L. Brooks, 3rd. 2008. Generalized structural polymorphism in self-assembled viral particles. *Nano Lett.* 8:4574–4581.
14. Nguyen, H. D., V. S. Reddy, and C. L. Brooks, 3rd. 2009. Invariant polymorphism in virus capsid assembly. *J. Am. Chem. Soc.* 131:2606–2614.
15. Zandi, R., D. Reguera, ..., J. Rudnick. 2004. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA.* 101:15556–15560.
16. Arkipov, A., P. L. Freddolino, and K. Schulten. 2006. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure.* 14:1767–1777.
17. Arkipov, A., W. H. Roos, ..., K. Schulten. 2009. Elucidating the mechanism behind irreversible deformation of viral capsids. *Biophys. J.* 97:2061–2069.
18. Hagan, M. F., and D. Chandler. 2006. Dynamic pathways for viral capsid assembly. *Biophys. J.* 91:42–54.
19. Johnston, I. G., A. A. Louis, and J. P. K. Doye. 2010. Modelling the self-assembly of virus capsids. *J. Phys. Condens. Mat.* 22:104101.
20. Kivenson, A., and M. F. Hagan. 2010. Mechanisms of capsid assembly around a polymer. *Biophys. J.* 99:619–628.
21. Krishna, V., G. S. Ayton, and G. A. Voth. 2010. Role of protein interactions in defining HIV-1 viral capsid shape and stability: a coarse-grained analysis. *Biophys. J.* 98:18–26.
22. Nguyen, H. D., V. S. Reddy, and C. L. Brooks, 3rd. 2007. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett.* 7:338–344.
23. Schwartz, R., P. W. Shor, ..., B. Berger. 1998. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* 75:2626–2636.
24. Endres, D., and A. Zlotnick. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* 83:1217–1230.
25. Hagan, M. F. 2009. A theory for viral capsid assembly around electrostatic cores. *J. Chem. Phys.* 130:114902.
26. 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.
27. Moisant, P., H. Neeman, and A. Zlotnick. 2010. Exploring the paths of (virus) assembly. *Biophys. J.* 99:1350–1357.
28. Zlotnick, A. 1994. To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* 241:59–67.
29. Mannige, R. V., and C. L. Brooks, 3rd. 2009. Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling. *Proc. Natl. Acad. Sci. USA.* 106:8531–8536.
30. Mannige, R. V., and C. L. Brooks, 3rd. 2010. Periodic table of virus capsids: implications for natural selection and design. *PLoS ONE.* 5:e9423.
31. Mannige, R. V., and C. L. Brooks, 3rd. 2008. Tiling nature of virus capsids and the role of topological constraints in natural capsid design. *Phys. Rev. E.* 77:051902.
32. Twarock, R. 2006. Mathematical virology: a novel approach to the structure and assembly of viruses. *Philos. Transact. A Math. Phys. Eng. Sci.* 364:3357–3373.
33. Reddy, V. S., H. A. Giesing, ..., J. E. Johnson. 1998. Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses. *Biophys. J.* 74:546–558.
34. Wales, D. J. 2005. The energy landscape as a unifying theme in molecular science. *Philos. Transact. A Math. Phys. Eng. Sci.* 363:357–375, discussion 375–377.
35. Kegel, W. K., and P. Schöot P. 2004. Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly. *Biophys. J.* 86:3905–3913.
36. Berthet-Colominas, C., S. Monaco, ..., S. Cusack. 1999. Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *EMBO J.* 18:1124–1136.
37. Gamble, T. R., F. F. Vajdos, ..., C. P. Hill. 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell.* 87:1285–1294.
38. Gitti, R. K., B. M. Lee, ..., W. I. Sundquist. 1996. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science.* 273:231–235.
39. Momany, C., L. C. Kovari, ..., M. G. Rossmann. 1996. Crystal structure of dimeric HIV-1 capsid protein. *Nat. Struct. Biol.* 3:763–770.
40. Worthylake, D. K., H. Wang, ..., C. P. Hill. 1999. Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* 55:85–92.
41. Gamble, T. R., S. H. Yoo, ..., C. P. Hill. 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science.* 278:849–853.
42. Ganser, B. K., S. Li, ..., W. I. Sundquist. 1999. Assembly and analysis of conical models for the HIV-1 core. *Science.* 283:80–83.
43. Briggs, J. A. G., T. Wilk, ..., S. D. Fuller. 2003. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* 22:1707–1715.
44. Ehrlich, L. S., B. E. Agresta, and C. A. Carter. 1992. Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *J. Virol.* 66:4874–4883.
45. Ganser-Pornillos, B. K., U. K. von Schwedler, ..., W. I. Sundquist. 2004. Assembly properties of the human immunodeficiency virus type 1 CA protein. *J. Virol.* 78:2545–2552.
46. Gross, I., H. Hohenberg, ..., H. G. Kräusslich. 2000. A conformational switch controlling HIV-1 morphogenesis. *EMBO J.* 19:103–113.
47. Li, S., C. P. Hill, ..., J. T. Finch. 2000. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature.* 407:409–413.
48. Ganser-Pornillos, B. K., A. Cheng, and M. Yeager. 2007. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell.* 131:70–79.
49. Byeon, I. J. L., X. Meng, ..., A. M. Gronenborn. 2009. Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell.* 139:780–790.
50. Pornillos, O., B. K. Ganser-Pornillos, ..., M. Yeager. 2009. X-ray structures of the hexameric building block of the HIV capsid. *Cell.* 137:1282–1292.
51. Chen, B., and R. Tycko. 2010. Structural and dynamical characterization of tubular HIV-1 capsid protein assemblies by solid state nuclear magnetic resonance and electron microscopy. *Protein Sci.* 19:716–730.
52. Han, Y., J. Ahn, ..., T. Polenova. 2010. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.* 132:1976–1987.
53. Gross, I., H. Hohenberg, and H. G. Kräusslich. 1997. In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur. J. Biochem.* 249:592–600.

54. Metropolis, N., A. W. Rosenbluth, ..., E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.
55. Meakin, P., H. Metiu, ..., D. J. Scalapino. 1983. The simulation of spinodal decomposition in two dimensions: a comparison of Monte Carlo and Langevin dynamics. *J. Chem. Phys.* 79:1948–1954.
56. Tiana, G., L. Sutto, and R. A. Broglia. 2007. Use of the Metropolis algorithm to simulate the dynamics of protein chains. *Physica A.* 380:241–249.
57. del Alamo, M., G. Rivas, and M. G. Mateu. 2005. Effect of macromolecular crowding agents on human immunodeficiency virus type 1 capsid protein assembly in vitro. *J. Virol.* 79:14271–14281.
58. Douglas, C. C., D. Thomas, ..., P. E. Prevelige, Jr. 2004. Investigation of N-terminal domain charged residues on the assembly and stability of HIV-1 CA. *Biochemistry.* 43:10435–10441.
59. Lanman, J., J. Sexton, ..., P. E. Prevelige, Jr. 2002. Kinetic analysis of the role of intersubunit interactions in human immunodeficiency virus type 1 capsid protein assembly in vitro. *J. Virol.* 76:6900–6908.
60. Ferrone, F. 1999. Analysis of protein aggregation kinetics. *Methods Enzymol.* 309:256–274.
61. Ceres, P., and A. Zlotnick. 2002. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry.* 41:11525–11531.
62. Briggs, J. A. G., M. N. Simon, ..., M. C. Johnson. 2004. The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* 11:672–675.
63. Briggs, J. A. G., J. D. Riches, ..., H. G. Kräusslich. 2009. Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. USA.* 106:11090–11095.
64. Lu, J. X., S. Sharpe, ..., R. Tycko. 2010. Oligomerization state and supramolecular structure of the HIV-1 Vpu protein transmembrane segment in phospholipid bilayers. *Protein Sci.* 19:1877–1896.
65. Wilber, A. W., J. P. K. Doye, ..., P. Wong. 2007. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J. Chem. Phys.* 127:085106.
66. Ayton, G. S., and G. A. Voth. 2010. Multiscale computer simulation of the immature HIV-1 virion. *Biophys. J.* 99:2757–2765.