

**Title: “Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly”**

*Author affiliations:* John M. A. Grime<sup>1</sup>, James F. Dama<sup>1</sup>, Barbie K. Ganser-Pornillos<sup>2</sup>, Cora L. Woodward<sup>3</sup>, Grant J. Jensen<sup>3,4</sup>, Mark J. Yeager<sup>2,5</sup>, and Gregory A. Voth<sup>1\*</sup>

<sup>1</sup> Department of Chemistry, Institute for Biophysical Dynamics, James Franck Institute, and Computation Institute, The University of Chicago, Chicago, IL 60637

<sup>2</sup> Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, Charlottesville, VA 22908

<sup>3</sup> Division of Biology and <sup>4</sup>Howard Hughes Medical Institute, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

<sup>5</sup> Center for Membrane Biology, Cardiovascular Research Center, and Division of Cardiovascular Medicine, University of Virginia School of Medicine, Charlottesville, VA 22908

*Corresponding author:* Prof. Gregory A. Voth  
5735 S Ellis Ave.,  
Department of Chemistry,  
The University of Chicago,  
Chicago, IL 60637  
Phone: (773) 702-9092  
Fax: (773) 795-9106  
E-mail: gavoth@uchicago.edu

# *Abstract:*

The maturation of HIV-1 viral particles is essential for viral infectivity. During maturation, many copies of the capsid protein (CA) self-assemble into a capsid shell to enclose the viral RNA. The mechanistic details of the initiation and early stages of capsid assembly remain to be delineated. We present coarse-grained simulations of capsid assembly under various conditions, considering not only capsid lattice self-assembly but also the potential disassembly of capsid upon delivery to the cytoplasm of a target cell. The effects of CA concentration, molecular crowding, and the conformational variability of CA are described, with results indicating that capsid nucleation and growth is a multi-stage process requiring well-defined metastable intermediates. Generation of the mature capsid lattice is sensitive to local conditions, with relatively subtle changes in CA concentration and molecular crowding influencing self-assembly and the ensemble of structural morphologies.

Significant morphological changes are required to convert an “immature” virus particle (virion) of HIV-1 into the mature and infectious form <sup>1-3</sup>. During maturation, the enzymatic cleavage of Gag polypeptide <sup>4</sup> releases capsid protein (CA) to self-assemble into a conical lattice structure enclosing the viral RNA (the capsid, Fig. 1). Failure to generate a suitable capsid precludes infectivity <sup>5-9</sup> and so the details of capsid generation are of significant interest as a therapeutic target.

HIV-1 capsid protein is mainly dimeric under physiological conditions, with CA monomers connected by a well-characterised CTD/CTD interface <sup>10</sup>. Given suitable conditions in vitro, CA can spontaneously self-assemble into a wide variety of structures <sup>11-18</sup>, but the prototypical mature virion contains a single conical capsid with a complex of viral RNA and nucleocapsid protein (NC) condensed within the broader terminus. In agreement with fullerene cone models <sup>14</sup>, pentameric and hexameric CA oligomers have been identified as the basic components of mature-style CA lattice, with CA NTDs arranged into quasi-equivalent rings, <sup>15,19-21</sup> and with pentamers believed be present in regions of higher local curvature in the capsid lattice.<sup>21,22</sup> Although the NTD and CTD structures of CA are well conserved, a flexible inter-domain linker region allows CA in solution to dynamically transition between conformations that are compatible and incompatible with mature lattice <sup>23</sup>.

Controlled study of HIV-1 capsid assembly is complicated by the inherent variability of virions: non-trivial differences in the average radius ( $\approx 630 \pm 50$  Å) and Gag content ( $\approx 2400 \pm 700$  molecules) are reported in vivo <sup>24-26</sup>, and these values bound a significant potential range of CA concentrations. The large numbers of Gag, in addition to viral RNA and sundry molecules sequestered from the host cell <sup>27,28</sup>, ensure the virion interior is a

highly crowded environment with internal organic mass densities as high as  $\approx 300$  mg/mL<sup>24</sup>. Molecular crowding effects are known to influence protein behaviours<sup>29</sup> and yet, remarkably, virion maturation consistently produces similar capsid morphologies. The relatively fast maturation process<sup>30,31</sup> makes in vivo study of capsid self-assembly pathways difficult. Nonetheless, Woodward et al.<sup>32</sup> show capsid formation proceeding via a hook-shaped precursor to the broad end of the capsid, in contrast with models of capsid assembly that begin with the narrow end of the cone<sup>33</sup>.

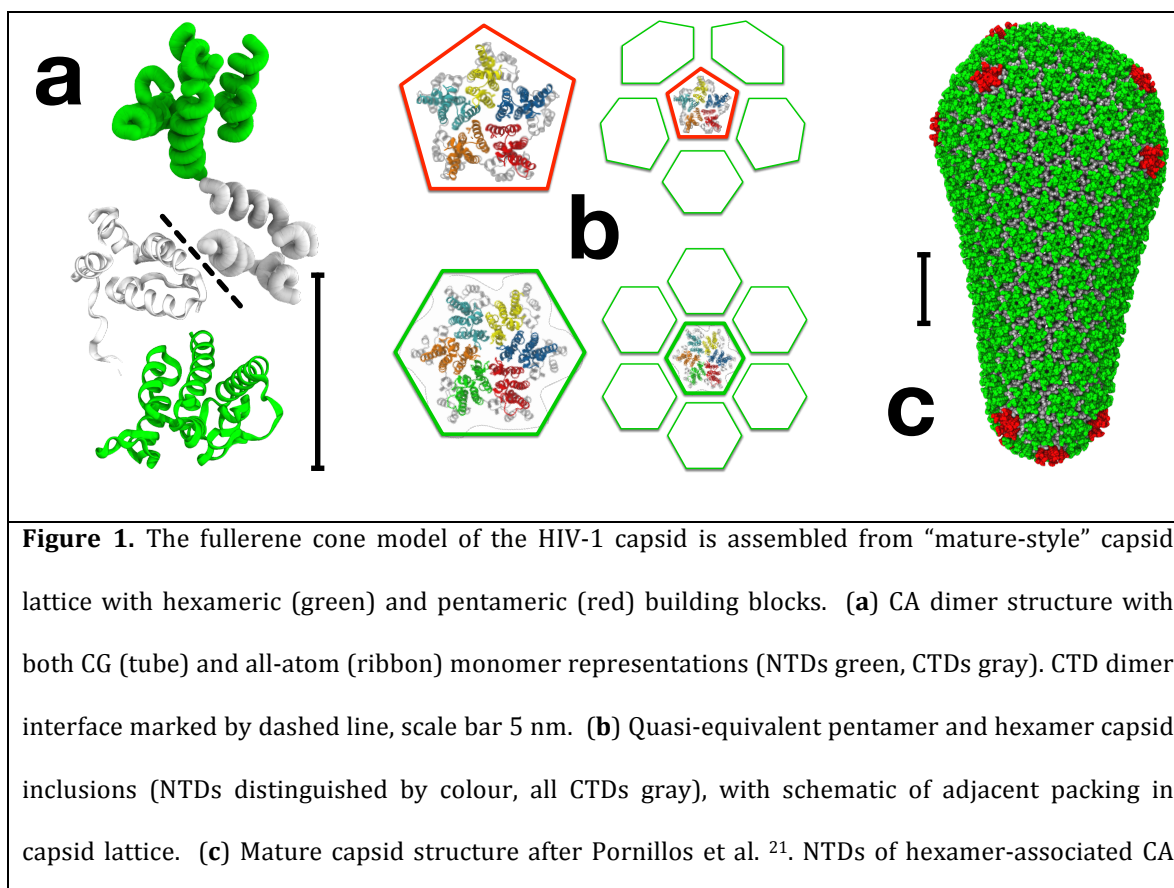
As the average virion contains approximately twice the CA present in a mature capsid, a capsid exists in equilibrium with a relatively high concentration of solution-state CA. Viral infection requires the transfer of virion contents into a target cell, producing a rapid dilution of this CA solution. Rapid dilution destabilises CA lattice in vitro under physiological salt concentrations<sup>12</sup>, but the potential significance of this effect on capsid stability is unclear.

Computer simulations of CA self-assembly offer a valuable complement to other experimental techniques. Atomic-resolution molecular dynamics (MD) simulations of pre-assembled capsids have been reported for HIV-1<sup>22</sup>, but the long time stability of such MD models has not been demonstrated and their great computational expense prevents such models being used to examine capsid self-assembly. Coarse-grained (CG) models using simpler molecular representations<sup>34,35</sup> provide an appealing alternative: for example, extensive CG simulations of icosahedral capsid self-assembly have been reported but the small, well-defined end products are in striking contrast to the structural pleomorphism observed with HIV-1 CA. Non-equilibrium CG simulations featuring irreversible subunit binding can generate capsid-relevant morphologies<sup>36,37</sup>, but may struggle to represent any



dynamic reorganization that might occur at the leading edges of lattice growth. The CG self-assembly of CA dimers was explored using Monte Carlo <sup>38</sup> and molecular dynamics simulations <sup>34</sup>, but these studies allowed CA motion only on 2-D or quasi-2D surfaces.

In this work, we consider 3-D self-assembly behaviours of CG HIV-1 capsid protein under various conditions relevant to the viral lifecycle. In particular, we examine the effects of CA concentration and molecular crowding on the generation of mature CA lattice, the influence of dynamic transitions between populations of CA that are compatible and incompatible with mature lattice formation, and the effects of rapid CA dilution on previously self-assembled CA lattice structures. In total, we present the results of 51 separate CG simulations with general trends across the data sets examined.



shown in green, with NTDs of pentamer-associated CA shown in red (all CTDs gray), scale bar 20 nm.

## Results

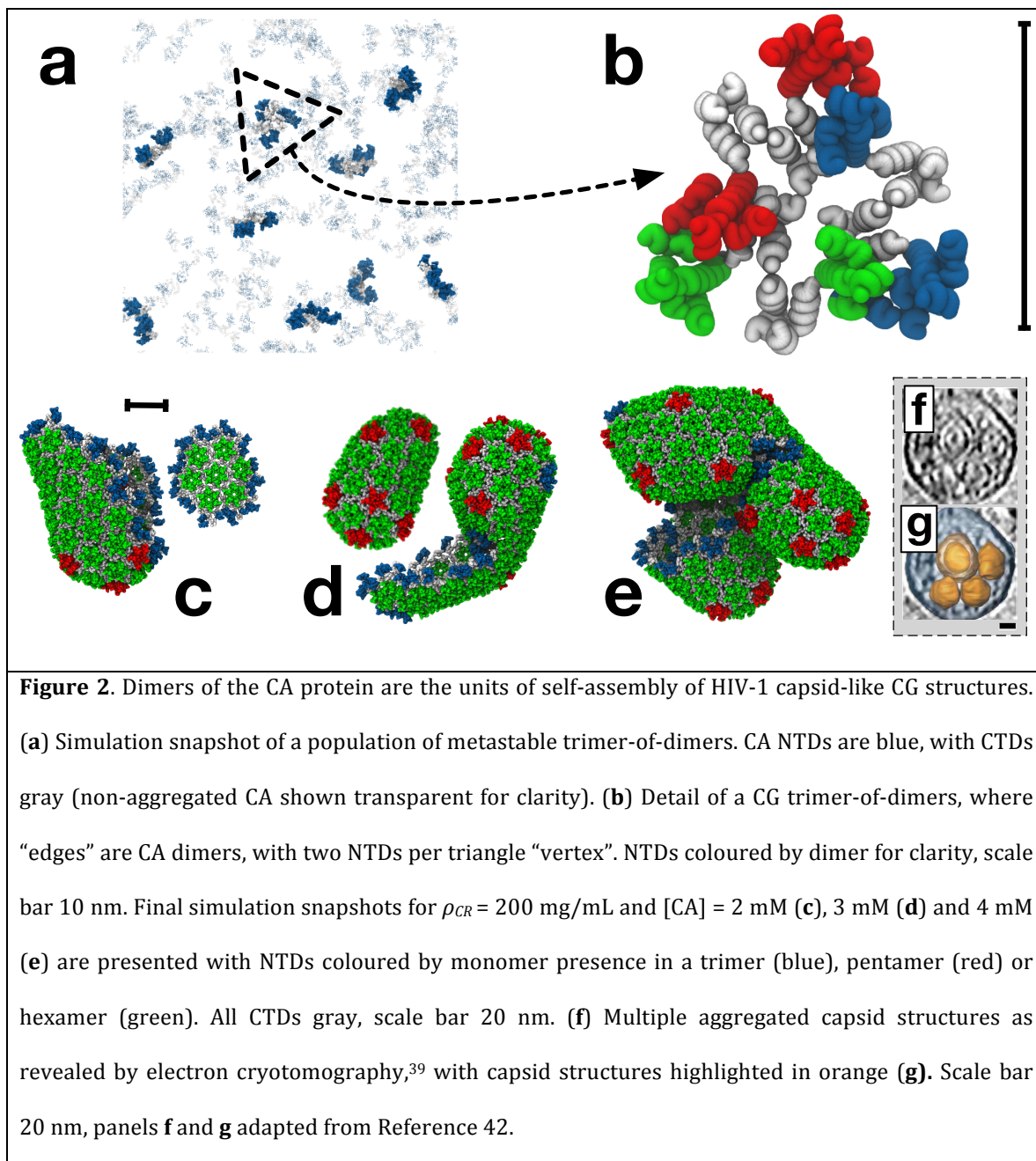
### *The effects of protein concentration and molecular crowding on the self-assembly of CG capsid protein.*

As noted, there is a dynamic population of CA conformations dictated by the flexible linker between the NTD and CTD. For our simulations, the equilibrium NTD-CTD conformation was based on that observed in the curved hexagonal lattice within CA tubes (Figure S1). CGMD simulations were performed (see Methods and Supporting Information Fig. SI 1A-C) using CA concentrations  $[CA]$  of 1 mM, 2 mM, 3 mM, and 4 mM (with  $[CA]$  in a typical virion estimated as  $\approx 3.8$  mM)<sup>24</sup>. For each  $[CA]$  studied, inert CG crowding agent densities  $\rho_{CR}$  from 0 mg/mL to 200 mg/mL were present with increments of 50 mg/mL to study the effects of molecular crowding on CG self-assembly<sup>12</sup>. A single simulation was performed for each data point for  $2 \times 10^8$  MD time steps, with results summarised in Table 1 (full data in Supporting Information, Fig. SI 2).

For  $[CA] = 1$  mM, no stable self-assembly of mature-style capsid lattice occurred for any level of crowding studied. Instead, populations of metastable “trimer-of-dimers” structures<sup>34</sup> rapidly formed (Fig. 2 a,b), with average trimer populations increasing with  $\rho_{CR}$ . A similar trend was observed for  $[CA] = 2$  mM until 200 mg/mL of crowder was present, at which point the self-assembly behaviour changed dramatically. Rather than generating a steady-state population of trimers, as was apparently the case for  $\rho_{CR} = 150$

mg/mL, a new self-assembly regime emerged: the steady production of mature-style CA lattice with quasi-equivalent pentamers and hexamers. Similar effects were observed for [CA] of 3 mM and 4 mM, with metastable trimers generated below the onset of CA lattice growth at  $100 < \rho_{CR} \leq 150$  mg/mL (for [CA] = 3 mM) and  $50 < \rho_{CR} \leq 100$  mg/mL ([CA] = 4 mM). The transition from producing only trimers to the nucleation and growth of mature-style lattice was thus a function of both [CA] and  $\rho_{CR}$ , with higher [CA] requiring lower  $\rho_{CR}$  to induce lattice growth. Differences of  $\leq 50$  mg/mL in  $\rho_{CR}$  therefore led to markedly different behaviours for [CA]  $\geq 2$  mM.

Where simulations generated only steady-state trimer populations, the steady state was rapidly attained ( $< 5 \times 10^6$  MD time steps). For simulations producing mature-style CA lattice, the appearance of stable pentameric inclusions lags the production of stable hexamers (see Fig. SI 2). In every simulation producing mature-style lattice, nucleation and growth of multiple lattice regions occurred. For example, with [CA] = 2 mM and  $\rho_{CR} = 200$  mg/mL, two independent lattice regions were generated: an incomplete cone fragment, and a small array of hexamers (Fig. 2c). For [CA] = 3 mM and  $\rho_{CR} = 200$  mg/mL, several independent lattice region appeared, resulting in a pill-shaped structure with a closed lattice surface and the fusion of an incomplete pill-shape with a cylinder fragment (Fig. 2d). For [CA] = 4 mM and  $\rho_{CR} = 200$  mg/mL, multiple lattice regions fused into a semi-amorphous assembly (Fig. 2e), reminiscent of multiple capsid structure aggregations observed in electron cryotomograms (Fig. 2 f-g).<sup>39</sup> The CG model is thus capable of producing a wide variety of curvatures, as required by the continuous curvatures observed in pleomorphic HIV-1 capsids.



Given the molecular mass of CA  $\approx 25$  kDa,  $[CA]$  of 2 mM, 3 mM, and 4 mM correspond to CA mass densities of  $\approx 50$  mg/mL,  $\approx 75$  mg/mL and  $\approx 100$  mg/mL. The onset of mature-style lattice growth therefore occurred for total mass densities of between 200 mg/mL and 250 mg/mL (for  $[CA] = 2$  mM), 175 mg/mL and 225 mg/mL ( $[CA] = 3$  mM), and

150 mg/mL and 200 mg/mL ( $[CA] = 4 \text{ mM}$ ). Total non-lipid mass density of HIV-1 virions was estimated as  $\approx 200$  to  $300 \text{ mg/mL}$  <sup>24</sup>, and so the CG model self-assembly initiated at mass density levels comparable to the virion.

These results indicate a pronounced sensitivity to both CA concentration and local molecular crowding for CG self-assembly. Well-regulated nucleation and growth pathways are therefore critical for capsid production, given the significant natural variability for HIV-1 virions <sup>24-28</sup>.

### *The effects of capsid protein concentration on CG self-assembly under constant levels of molecular crowding*

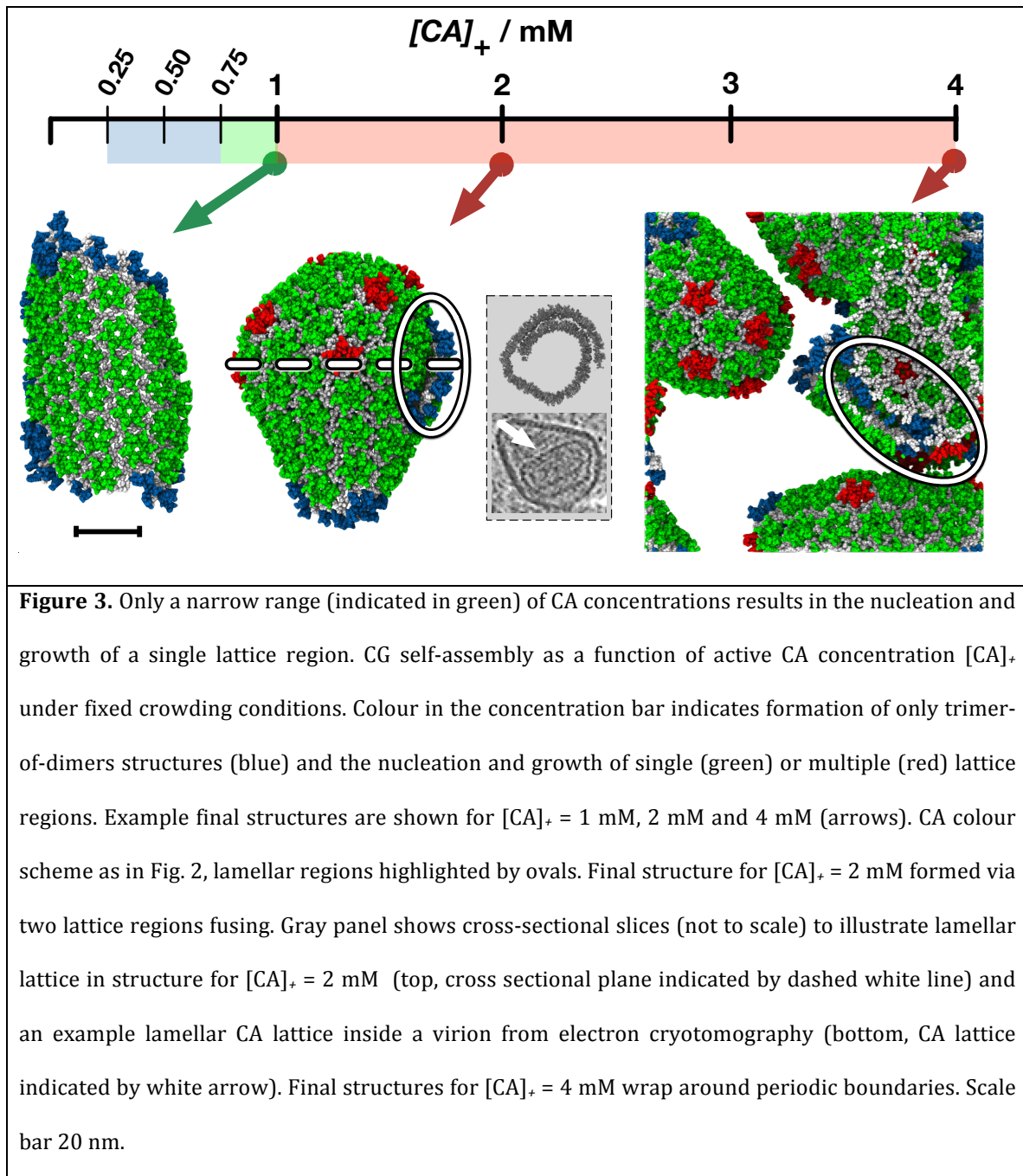
The level of inert molecular crowding can significantly alter CG self-assembly behaviours. However, CA is also a self-crowding agent as it occupies volume in the simulations. Direct comparison between self-assembly behaviours of different CA concentrations with the same quantity of inert crowder is therefore non-trivial.

To examine the effects of CA concentration on CG self-assembly under constant molecular crowding, the initial CA content of identical systems ( $4 \text{ mM CA}$ ,  $\rho_{CR} = 200 \text{ mg/mL}$ ) was partitioned into fixed “active” ( $[CA]_+$ ) and “inactive” ( $[CA]_-$ ) populations. After this initial partitioning, CA dimers therefore remain either “active” or “inactive” for the duration of the simulation. This partitioning is motivated by the experimental results of Deshmukh et al. <sup>23</sup> that suggest there is a relatively small population of “assembly-competent” CA dimers in CA solution. Inactive CA was identical to active CA in all aspects in our CG model except for the removal of attractive interactions, and thus both types of CA

occlude equivalent volume. By fixing  $[CA]_+$  between 250  $\mu\text{M}$  and 4 mM (with increments of 250  $\mu\text{M}$ ), CG self-assembly was studied as a function of CA concentration alone while maintaining fixed, virion-relevant crowding levels. The formation of trimer-of-dimers, hexamer, and pentamer structures were recorded as a function of simulation time over  $2 \times 10^8$  CGMD timesteps (one simulation per CA concentration), with selected results summarised in Fig. 3 (full data in Table 2 and Fig. SI 3).

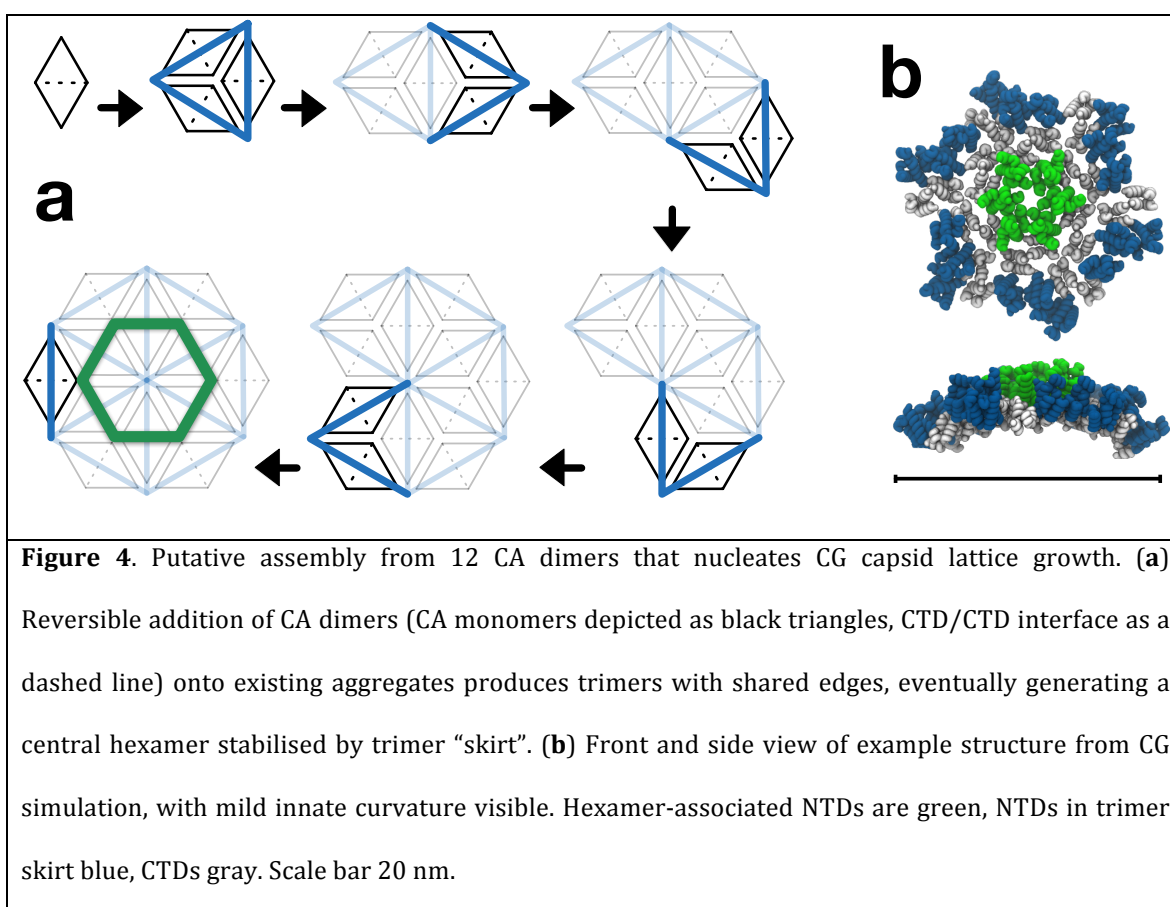
For  $250 \leq [CA]_+ \leq 500 \mu\text{M}$ , no generation of mature-style CA lattice occurred, but metastable trimer-of-dimer populations emerged. Elevating  $[CA]_+$  to 750  $\mu\text{M}$  produced an apparent initial steady-state trimer population before a single region of mature-style lattice nucleated and grew after  $\approx 4 \times 10^7$  MD timesteps. With  $[CA]_+ = 1 \text{ mM}$ , the nucleation and growth of a single lattice region began after  $\approx 1 \times 10^7$  MD timesteps. For both  $[CA]_+ = 750 \mu\text{M}$  and  $[CA]_+ = 1 \text{ mM}$ , the lattice region adopted a cylindrical conformation as growth proceeded; although pentameric inclusions regularly formed at the growing edge of the lattice, the inclusions were transient and thus the lattice was essentially hexameric. For all  $[CA]_+ \geq 1.25 \text{ mM}$  nucleation and growth of multiple lattice structures occurred, with regions of high local curvature around the pentamers. Also evident at higher  $[CA]_+$  were lamellar regions of CG lattice (Fig. 3) reminiscent of structures observed in electron cryotomograms. Lamellar and off-pathway capsid structures are indicated to be significantly more common in virion systems lacking RNA/NC complex,<sup>40</sup> as is the case in these CG simulations.





The relatively controlled nucleation and growth for lower active CA concentration  $[CA]_+$  reveals specific details of CG lattice nucleation. Visual inspection of simulation trajectories indicate lattice nucleation occurring by the addition of CA dimers onto existing trimer-of-dimers structures, producing trimers with shared edges (Fig. 4) and leading to a

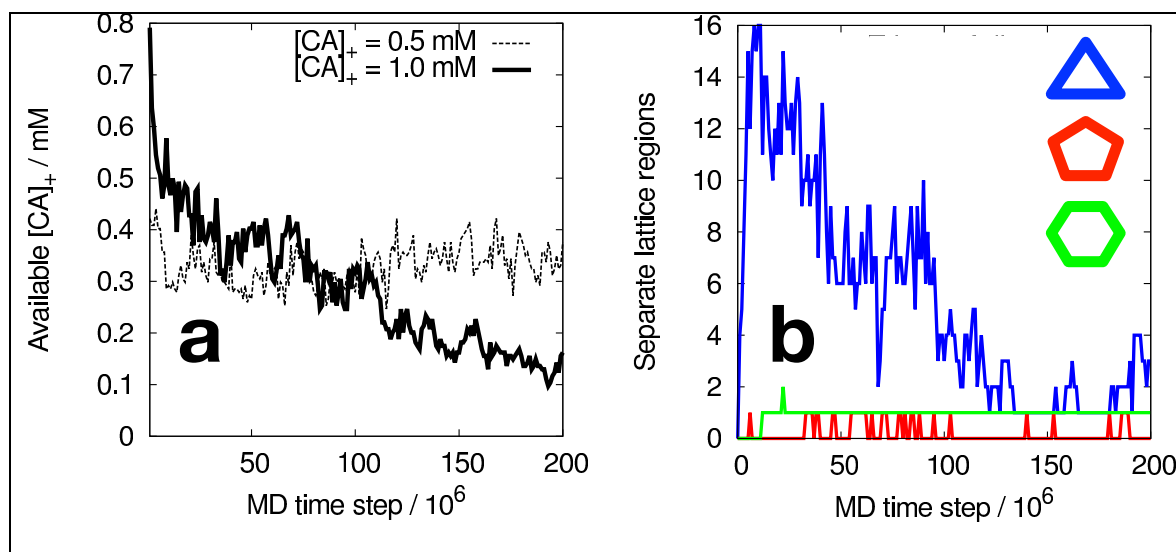
central hexamer stabilised by a peripheral trimer “skirt” from which lattice growth proceeds. Metastable trimers therefore appear to be seeds for the nucleation and growth of CG lattice <sup>34</sup>, rather than independent trimers aggregating. Transient structures along this pathway are observed in simulations that do not produce mature lattice, but these intermediates are unstable and dissociate before nucleation of lattice growth.



Examination of the solution-state  $[CA]_+$  reveals that lattice growth continues below the level of solution-state  $[CA]_+$  required for initial nucleation (e.g. Fig. 5a), in agreement with standard nucleation and growth theories <sup>41</sup>. As a region of mature lattice grows, there is also a reduction in the number of separate trimers in solution (e.g. Fig. 5b). Progressive



reduction of not only the “available” capsid protein but also the number of trimers in solution (the potential nucleating seeds) can thus provide an elegant feedback mechanism to deter multiple capsid formation.



**Figure 5.** Example CG simulation data under fixed molecular crowding. **(a)** “Available” assembly-competent CA in solution for initial  $[CA]_{+} = 0.5$  mM (resulting in no nucleation and growth of mature lattice) and initial  $[CA]_{+} = 1.0$  mM (producing nucleation and growth of a single lattice region, see main text and Table 2). Lattice growth continues below the level of available  $[CA]_{+}$  required for nucleation on the same time scale. **(b)** Number of separate lattice regions containing key structural motifs for  $[CA]_{+} = 1$  mM. The number of trimers in solution (blue curve) reduces significantly as a single region of lattice grows.

These results indicate a marked sensitivity of CG self-assembly processes to CA concentration under constant molecular crowding. Increases of as little as 250  $\mu$ M in  $[CA]_{+}$  can drive the system from steady-state populations of metastable trimer structures ( $[CA]_{+} < 750$   $\mu$ M) into the nucleation and growth of single ( $750$   $\mu$ M  $\leq [CA]_{+} \leq 1$  mM) or multiple lattice regions ( $[CA]_{+} \geq 1.25$  mM) over the time scales examined. Greater quantities of  $[CA]_{+}$

may encourage pentamer formation and/or stability over the course of the simulations: for  $[CA]_+ \leq 1$  mM, no stable pentamers were detected (Table 2).

*The effects of CA NTD/CTD conformational freedom on CG self-assembly under constant protein concentration and molecular crowding*

The NTD and CTD structures of CA are well conserved, but a flexible linker region allows significant conformational freedom in solution. Deshmukh et al. estimate that the solution-state CA with NTD/CTD arrangements compatible with mature lattice might be as low as  $\approx 5\%$  <sup>23</sup>. Furthermore, this population is inherently dynamic, with correlation times for inter-domain motions estimated as between 2 ns and 5 ns <sup>23</sup>. The CGMD simulations with fixed populations of “active” ( $[CA]_+$ ) and “inactive” ( $[CA]_-$ ) CA effectively describe a limiting case of this behaviour:  $[CA]_+$  corresponds to a population of CA whose instantaneous structure is compatible with mature lattice (and hence may directly self-assemble) and the fixed populations of  $[CA]_+$  and  $[CA]_-$  correspond to infinitely long correlation times for inter-domain motions (i.e., specific proteins remain in either assembly competent or incompetent conformations for the entire simulation).

The effects of interconversion between structural populations are now considered. Rather than initial fixing of  $[CA]_+$  and  $[CA]_-$  populations, all “free” CA in solution (i.e., CA which is not currently aggregated) is periodically identified and then randomly assigned to the  $[CA]_+$  or  $[CA]_-$  populations to maintain a specific target ratio of  $[CA]_+$  and  $[CA]_-$  in solution. The interval between random population assignments is thus a simplified proxy for time correlations in CA structural conformation: once assigned to either the  $[CA]_+$  or

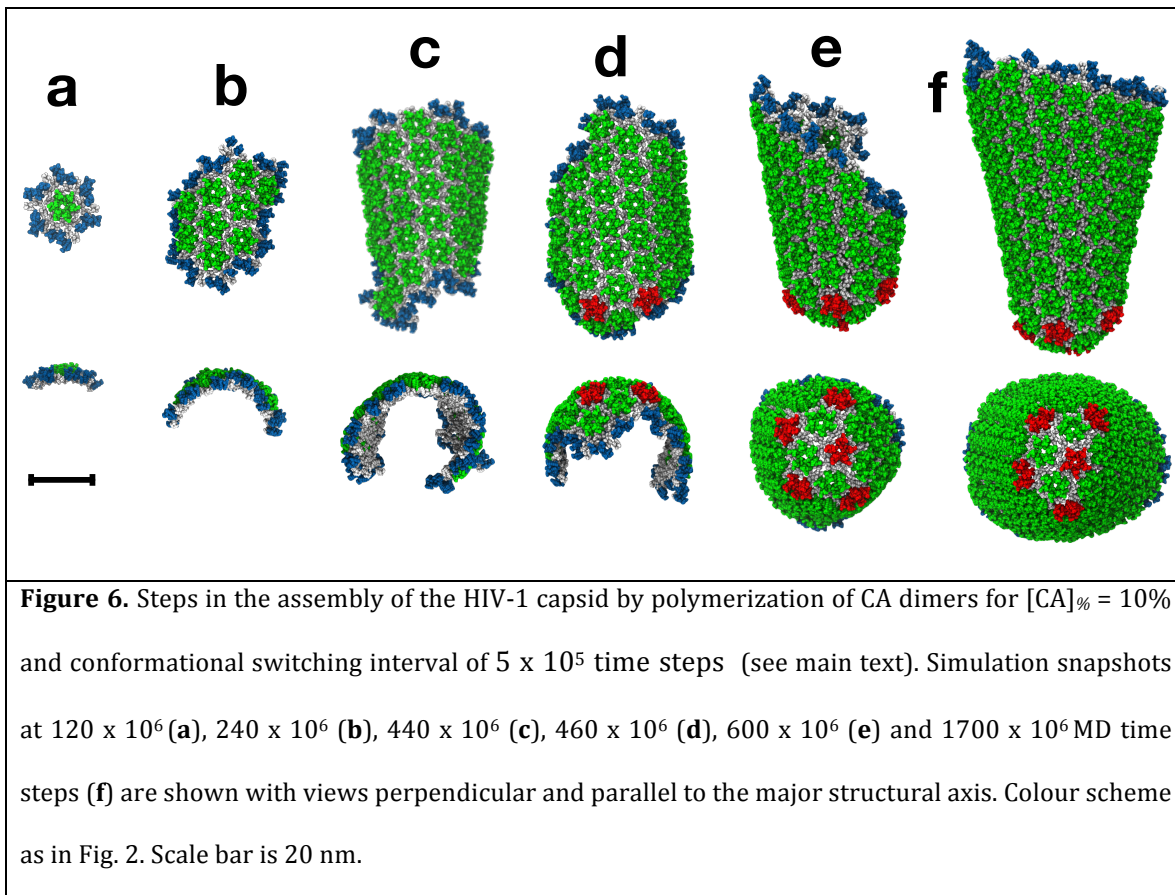
[CA]. populations, a CA dimer remains “assembly competent” or “assembly incompetent” at least until the next random population assignment. Aggregated CA is not included in this process to avoid perturbation of existing aggregates, and to reflect CA conformational stability in mature CA lattice. Specific CA can disaggregate and, in the absence of re-aggregation, would be subject to the next random population switch. This approach allows significant control over the quantity of assembly competent CA in solution in order to mimic the coexistence of assembly competent and incompetent CA populations in solution.

For an identical baseline system (4 mM total CA, 200 mg/mL crowder), target “assembly competent” proportions  $[CA]_{\%}$  of 2.5%, 5%, 10%, 25%, and 50% of solution-state CA were studied (one simulation per  $[CA]_{\%}$  value). Random population switching was performed with intervals of  $1 \times 10^4$ ,  $1 \times 10^5$ , and  $5 \times 10^5$  CGMD time steps, corresponding to 0.1 ns, 1 ns, and 5 ns of CG simulation time respectively, but caution should be exercised in any direct comparison of CG timescales to those suggested by NMR experiments (see note regarding CG time scales in SI 4). Simulations were performed for  $2 \times 10^8$  CG MD time steps unless explicitly noted, with results summarised in Table 3 and Fig. 6 (full data in Fig. SI 4).

Apparent from Table 3 is the influence of conformational switching intervals on CG self-assembly. For the most rapid switching rate (interval =  $1 \times 10^4$  time steps), no self-assembly of mature-style CA lattice emerges until some 50% of the total CA in solution is considered to be in an assembly competent form: below this level, only metastable trimer-of-dimer structures appear. With a slower switching rate (interval =  $1 \times 10^5$  time steps), the transition from steady-state trimers into mature-style lattice occurs for  $10\% < [CA]_{\%} \leq 25\%$ . For the slowest switching rate studied ( $5 \times 10^5$  time steps), the transition occurs at  $5\% < [CA]_{\%} \leq 10\%$ . For simulations producing only metastable trimers, faster switching

rates reduces the average trimer population for identical  $[CA]_{\%}$  (with an additional  $1 \times 10^8$  MD time steps performed to improve statistical estimates). These results suggest that correlation times for the NTD/CTD motions in capsid protein affect CG self-assembly, with faster switching between assembly competent and incompetent forms suppressing lattice formation for otherwise identical  $[CA]_{\%}$ .

When a target active proportion  $[CA]_{\%}$  of 10% was combined with a switching interval of  $5 \times 10^5$  time steps, a cone-shaped structure reminiscent of a mature capsid was produced (Fig. 6). A small region of hexameric lattice initially formed, which curled into a semi-cylinder under growth. Stable pentamer incorporation appeared alongside higher local curvatures at one end of the structure, generating the narrow end of a cone. At this point lattice growth essentially stopped due to insufficient assembly-competent CA in solution to drive growth, as the system contained only 50% of the CA content of a typical virion.<sup>24-26</sup> Although transient pentameric inclusions regularly formed at the exposed lattice edges, all permanent capsomer additions were hexameric after this point. It is interesting to note that pentamers were occasionally embedded slightly behind the growing edge of the lattice, but were replaced by hexamers in a process of local remodeling. The CG self-assembly can thus demonstrate a degree of “error correction”, provided the growing lattice edge does not advance too far beyond the pentamers. We note that although these conditions produced the assembly of a cone-shaped structure reminiscent of a mature capsid in this specific simulation, such results should not be considered to imply that these particular conditions are unique in producing this outcome.



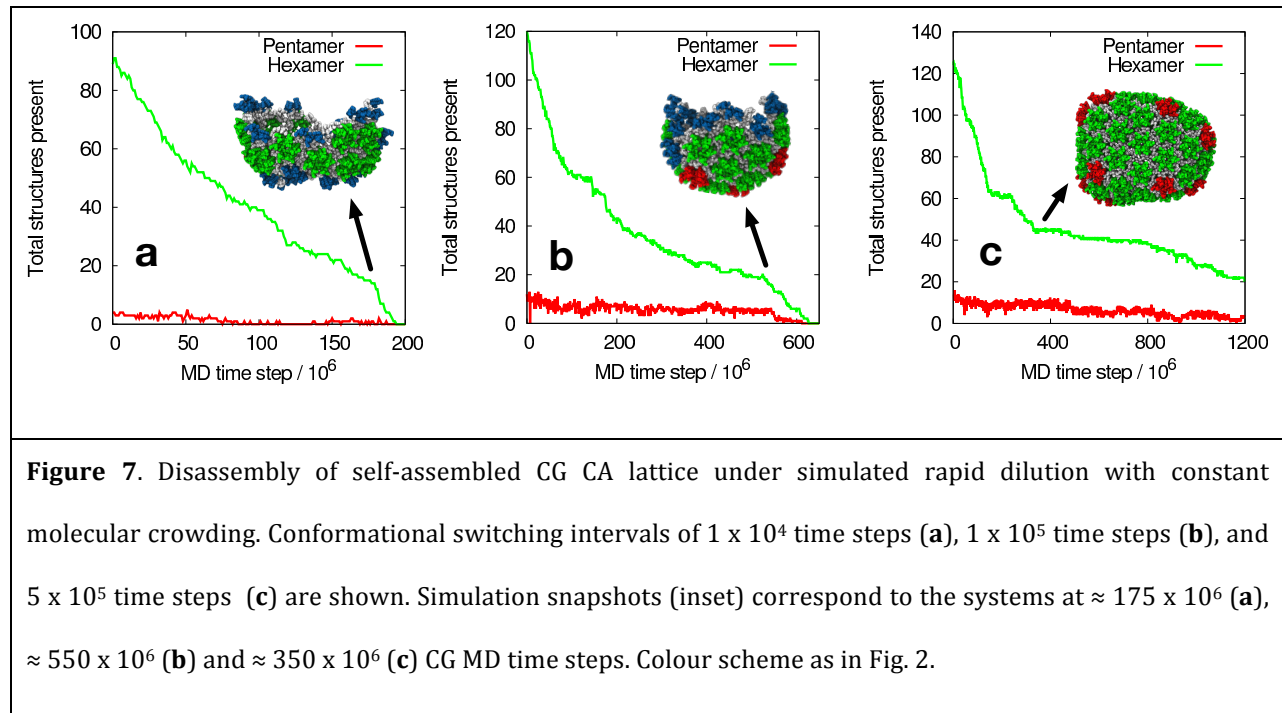
### *Stability and uncoating of self-assembled CG capsid lattice structures*

Capsid assembly is critical for HIV-1 infectivity. However, capsid destruction is also crucial to the viral lifecycle: successful infection is sensitive to the specific time at which a capsid uncoats after cell entry, with both premature and delayed uncoating detrimental to viral replication<sup>6,42</sup>. Assuming virion radius  $\approx 630 \text{ \AA}$ <sup>24</sup>, the transfer of virion contents into a cell of radius  $\approx 10 \text{ \mu m}$ <sup>43</sup> produces a rapid dilution of the CA solution around the capsid. To examine the effects of rapid dilution, simulations using previously self-assembled lattice structures were performed with  $[CA]_{\%} = 0$  to approximate the negligible effective concentration of CA in solution after virion contents are transferred into a cell. Importantly,

this approach preserves the level of molecular crowding and conformational switching rate under which the lattice structures originally formed, allowing investigation of rapid dilution on several different structures that are otherwise stable under identical conditions. This avoids the use of a single model structure, whose natural stability may differ subtly under different test conditions even in the absence of rapid dilution. These simulations therefore probe the effects of rapid dilution specifically, representing a cytoplasmic environment that is otherwise identical to the notional virion in which the structures were generated.

Maximally assembled systems (i.e., using  $[CA]_{\%} = 50\%$ ) for the three conformational switching rates described previously were used as starting configurations for rapid dilution, with  $[CA]_{\%}$  now set to 0% in each case. Fig. 7 presents the resultant time series of structural data from one rapid dilution simulation per starting conformation, with dissolution of CA lattice evident for all systems. Conformational switching intervals of  $1 \times 10^4$  time steps (Fig. 7a) and  $1 \times 10^5$  time steps (Fig. 7b) show a distinct and rapid destruction of any remaining lattice at  $\approx 175 \times 10^6$  and  $\approx 550 \times 10^6$  CGMD time steps respectively, indicating lattice dissolution is not a simple function of time (lattice structures at these points shown in Fig. 7, inset). Interestingly, the  $5 \times 10^5$  time step switching interval showed a long-lived plateau between  $\approx 350 \times 10^6$  and  $\approx 800 \times 10^6$  MD time steps corresponding to reorganization of the remaining lattice into a pill-like structure with closed surface (Fig. 7c). This enhanced stability is due to the absence of exposed lattice edges: CA in edge-free lattice requires significant reorganization of surrounding lattice components to escape, but can dissolve from exposed edges directly.

These results demonstrate the instability of CG capsid lattice under rapid dilution, even under conditions otherwise identical to those in which the lattice assembled. Furthermore, the results suggest that CA lattice structures with no exposed edges may be significantly more stable upon transfer into a cell.



## Discussion

Our study shows that CG self-assembly is sensitive to CA concentration, molecular crowding, and NTD/CTD conformational correlation times. A single CG model generates surprisingly large morphological variety, encompassing the wide range of local curvatures required in the viral capsid<sup>15</sup>. Evident in the results is a reversible multi-stage process triggering mature lattice self-assembly,<sup>44</sup> where metastable trimer-of-dimers structures

rapidly emerge to act as potential seeds for the nucleation of mature lattice growth (Fig. 4). The steady-state population of trimers increases as a function of both concentration and molecular crowding, until suitable conditions encourage the generation of mature-style capsid lattice over the simulation lengths examined, progressively reducing both the solution-state CA and the population of independent trimers to deter the nucleation of additional lattice regions (e.g., Fig. 5). The transition between self-assembly regimes appears to be quite subtle, with relatively small changes to the molecular environment capable of producing different outcomes in agreement with the sensitivity displayed e.g. in previous studies of self-assembly for icosahedral capsid models.<sup>45-50</sup>

The general effects of molecular crowding in our simulations are in agreement with those observed using other computational approaches to self-assembly. For example, high levels of crowding can reduce nucleation lag times to effectively zero (see e.g. Figs. SI 2-4), changing the basic nature of the self-assembly process from nucleation-limited to non-nucleation-limited. Schwartz and co-workers also noted such effects in discrete event simulations of icosahedral capsid self-assembly:<sup>48</sup> multiple parallel nucleation and growth events depleted the available capsid components, arresting complete capsid assembly and producing off-pathway growth. Similar kinetic traps and multiple nucleation was also observed in CG simulations of icosahedral capsid assembly.<sup>45-50</sup>

In the absence of RNA, CA is typically observed to assemble into tubes of variable diameter *in vitro*.<sup>10,12,13,15-18</sup>, albeit with occasional cone-shaped structures<sup>12</sup> and other non-tubular morphologies also reported.<sup>13,17</sup> These experiments are typically carried out with CA concentrations significantly below that expected in the virion, which is likely to assist in



the generation of regular, purely hexameric cylinders via orderly nucleation and growth to help avoid off-pathway phenomena.

Sigmoidal “assembly curve” characteristics have been observed in experiment for capsid assembly systems in e.g. HIV-1<sup>12</sup> and HPV-11.<sup>51</sup> Our results follow this general prediction, with lag periods for assembly that can be very short or effectively zero with CA concentration and/or significant crowding (see e.g. SI 2). We note that simulations lacking mature-style lattice production do not necessarily indicate that CG lattice growth cannot occur under these conditions, but rather that nucleation and growth was not detected over the simulation time scales examined. As the main focus of this study regards the nucleation and growth processes of capsid lattice, converged simulations (i.e., simulations that have reached some final steady state) do not offer significantly more information regarding these aspects. Nonetheless, examination of the number of structures relevant to mature lattice present in the simulations as a function of time (e.g., Figs. SI 2-4) indicates that many simulations are approaching the final plateau of a sigmoidal growth curve, and hence significant additional aggregation of capsid protein is unlikely.

Trimer-of-dimers structures have been reported previously in computer simulations<sup>34,38</sup>, and have also been assumed as the fundamental building block in mathematical models of capsid cylinder growth<sup>52</sup>. Our results suggest that, in isolation, these structures are metastable, and the addition of CA dimers onto these triangular templates (Fig. 4) provides the basic nucleation pathway for capsid assembly rather than the aggregation of independent trimer-of-dimer structures.

The regular appearance of transient pentamers at the unstable growing edge of CG lattice was observed, suggesting a natural aspect of lattice growth. If growth is relatively

slow, local remodeling can remove the pentamers where the expanding lattice edge does not advance too far. For more rapid assembly (e.g., higher crowding or CA concentration) the lattice may advance too quickly for reliable remodeling, producing high local curvature via stably embedded pentamers. This process can explain why relatively low concentrations of CA typically produce cylindrical structures in vitro <sup>10,12,15-20</sup>, and why increased CA concentration reduces cylinder length <sup>16</sup>: early pentamer incorporation redirects the lattice growth, suppressing the formation of longer cylinders. The inherent instability of the lattice edge suggests that CG models that allow disassembly therefore more accurately reflect capsid assembly. <sup>36,37,44</sup>

Switching between active and inactive CA populations in solution, akin to the conformational dynamism experimentally recorded in vitro <sup>23</sup>, has pronounced effects on CG lattice self-assembly. Faster random mixing of the populations appears to suppress CG lattice production (Table 3), with a low instantaneous proportion of assembly-competent CA limiting uncontrolled nucleation and growth. This effect may be due to the properties of CG CA protein immediately after dissociation from a lattice region: with longer intervals between population switching, such CA may remain in an assembly-competent conformation long enough for protein/protein interactions to encourage direct re-association, but more investigation is required to fully elucidate this potential mechanism. In any case, dynamic conformational switching ensures that growth after nucleation is not prevented by the exhaustion of an otherwise limited quantity of suitable CA. This finding suggests that CG models incorporating this phenomenon may better represent a key aspect of HIV-1 capsid assembly.

It is interesting to consider HIV-1 virion maturation in light of these results. Although the prototypical mature virion contains a single capsid, multiple capsid structures are sometimes observed<sup>9,31-33,53</sup>, both with and without RNA/NC encapsulation<sup>32</sup>, suggesting a delicate balance in the virion. Multiple capsids appear correlated with larger virions<sup>33,53</sup>, consistent with sensitivity of CG assembly to local conditions: lattice growth reduces capsid protein in solution, so larger virions (with larger numbers CA for the same overall CA concentration) could maintain CA levels above a critical nucleating value for longer periods after an initial lattice region forms. This increases the likelihood of additional nucleation events, compatible with observations that multiple cores are often comparable in size<sup>53</sup>, which might suggest they nucleated in rather quick succession with no further nucleation occurring. The sensitivity of CG self-assembly to CA concentration and local molecular crowding, in combination with natural virion variability, may thus complicate previous analyses suggesting multiple core assembly is not a function of CA concentration<sup>53</sup>.

Our CG simulations produced virion-relevant morphologies such as the extended cone-shaped structure shown in Fig. 6, including the initial generation of the narrow end of the cone. However, electron cryotomography suggests that capsid growth typically proceeds from a small region of CA lattice adsorbed to the RNA/NC complex, generating the broad end of the capsid first (although capsid growth also occurs without RNA/NC encapsulation).<sup>32</sup> The presence of such an RNA/NC complex in our simulations would prevent the high local curvature present in the narrow end of the capsid, and would thus select against the initial formation of this narrow end. Our simulations therefore do not fully recapitulate the true virion environment, but instead explore principles relevant to

capsid self-assembly. Nonetheless, our results are in general agreement with models of capsid assembly that feature an initial small sheet of CA lattice that curls under growth, producing a cup shaped structure from which assembly proceeds.<sup>32</sup>

Our results suggest the importance of relatively slow, controlled growth of the HIV-1 capsid structure. Such a process allows significant relaxation of both the local and global structure during capsid growth, to allow the generation of a metastable capsid despite the apparently weak interactions between CA dimers.<sup>11,12</sup> This phenomena suggests caution is appropriate when fitting atomic-resolution models into relatively low-resolution capsid data: for example, CTD/CTD dimer interfaces present in the atomistic MD simulations of HIV-1 viral capsids by Zhao and co-workers<sup>43</sup> can display significant shearing and distortions from the expected structures<sup>10,23</sup> (Fig. SI 5). It is possible that such deformations occur naturally in a viral capsid, but further experimental support is needed to clarify this situation.

The stability of CG lattice structures was sensitive to CA concentration in the supporting solution. Rapid dilution under otherwise identical conditions destabilised capsid lattice, indicating the metastability of CA structures in agreement with experimental observations<sup>12</sup>. Similar effects have also been observed in simulations of icosahedral capsids,<sup>54</sup> albeit these simulations did not examine this effect under constant crowding conditions. Sealed, edge-free CA lattices appear more resistant to this process (Fig. 7c). Cellular responses to HIV-1 include the weak binding of TRIM5 $\alpha$  protein to capsids, forming a hexameric super-lattice on the capsid exterior<sup>55</sup>. TRIM5 $\alpha$  binding promotes rapid capsid uncoating<sup>42,56-61</sup>, and while TRIM5 $\alpha$  is digested by the proteasome<sup>62</sup> capsid protein does not appear to be destroyed in this manner<sup>56,57,61</sup>. TRIM5 $\alpha$  restriction is less

efficient under proteasomal inhibition <sup>63-65</sup>, and we speculate this somewhat counterintuitive data can be explained by TRIM5 $\alpha$  inducing local faults in the capsid lattice: slight mismatches in CA and TRIM5 $\alpha$  lattice spacing induces additional stress in the metastable CA lattice <sup>66</sup>, with local tears providing an exposed edge from which CA can dissolve. Proteasomal degradation then ensures that a TRIM5 $\alpha$  “scaffold” cannot exert any stabilizing influence on the overall capsid superstructure, enabling fragmentation of the capsid to further accelerate uncoating. Partial support for this hypothesis is offered by TRIM5 $\alpha$  inducing only mild disruption of CA cylinders in vitro <sup>59</sup>, but as these studies used above-physiological salt concentrations with no rapid dilution (conditions in which the CA lattice may be over-stable) <sup>12</sup> complete breakdown of CA lattice was not observed. Experimental data suggests HIV-1 capsids often feature exposed lattice edges, with an estimated 25% of cytoplasmic capsids featuring seams or holes large enough to allow the escape of GFP,<sup>37</sup> and so TRIM5 $\alpha$  could help to accelerate natural lattice dissipation under rapid dilution. For capsid structures lacking exposed lattice edges, however, TRIM5 $\alpha$  would provide a powerful accelerator of uncoating. This process would offer a relatively simple description of the general effects of TRIM restriction on HIV-a capsids, but we note that the natural TRIM restriction processes could be significantly more complicated.

## Methods

*Model Details.* Full CG model details are provided in SI 1. Briefly, CG capsid protein is modeled directly from experimental hexamer (PDB 3H4E) and pentamer (3P0A) data <sup>20,21</sup>, and also heavily influenced by work from the same authors (B. K. G-P., M.Y.) with

cylindrical CA assemblies in vitro<sup>14-16,19,20</sup>. CA is represented as a dimer, the prevalent species in solution<sup>10,23</sup>, which is required for CA lattice assembly<sup>11</sup>. The NTD and CTD of a CA monomer are represented as independent stiff elastic network models (ENMs) comprising the C $\alpha$  atoms from CA alpha helices, which are well conserved across experimental structures<sup>34</sup>. An additional weaker ENM connects the NTD and CTD in each monomer to provide limited inter-domain structural flexibility. The CTD/CTD dimer interface is represented by a stiff ENM based on PDB structure 2KOD<sup>10</sup> relevant for mature lattice<sup>23</sup>. CG beads have excluded volume radii to deter unphysical overlaps as indicated by experimental structural data. Important CA protein/protein interfaces required for self-assembly<sup>9,10</sup> are represented by specific attractive interactions, with the locations of energy minima parameterized using experimental structures<sup>20,21</sup> and attractive strengths chosen to allow CA self-assembly under the experimentally relevant mass densities and conditions we study here (SI 1). These attractive interactions can be dynamically enabled and disabled, allowing the generation of a fixed solution-state ratio of assembly competent and assembly incompetent populations as indicated by NMR data.<sup>23</sup> Inert crowding agent is based on the excluded volume and relative mass of Ficoll 70<sup>12</sup>. The combination of internal NTD/CTD flexibility in CA monomers, “switching” capability, and freedom of motion in a 3-D simulation domain with explicit crowding molecules provides an advanced and relatively detailed molecular model with which to study CA self-assembly.

*Simulation details.* All simulations performed with our UCG-MD code<sup>67</sup>. Simulation domain was an 800 Å cube ( $\approx$  50% of typical virion volume for computational efficiency, with max. 616 CA dimers present at [CA] = 4mM) with periodic boundaries. Temperature of 300 K

maintained by Langevin dynamics with relaxation period 100 ps<sup>-1</sup> to minimise influence of thermostat, integration timestep 10 fs. Visualizations of simulation data were created using VMD,<sup>68</sup> with CG CA depicted as tubes connecting CG particles in each notional alpha helix to emphasize the local secondary structure of the CG model.

## References

- 1 Briggs, J. A. & Krausslich, H. G. The molecular architecture of HIV. *J. Mol. Biol.* **410**, 491-500 (2011).
- 2 Ganser-Pornillos, B. K., Yeager, M. & Sundquist, W. I. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* **18**, 203-217 (2008).
- 3 Sundquist, W. I. & Krausslich, H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).
- 4 Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* **68**, 8017-8027 (1994).
- 5 Dorfman, T., Bukovsky, A., Ohagen, A., Hoglund, S. & Gottlinger, H. G. Functional domains of the capsid protein of human immunodeficiency virus type 1. *J. Virol.* **68**, 8180-8187 (1994).
- 6 Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667-5677 (2002).
- 7 Mammano, F., Ohagen, A., Hoglund, S. & Gottlinger, H. G. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J. Virol.* **68**, 4927-4936 (1994).
- 8 Tang, S. *et al.* Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. *J. Virol.* **75**, 9357-9366 (2001).
- 9 von Schwedler, U. K., Stray, K. M., Garrus, J. E. & Sundquist, W. I. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.* **77**, 5439-5450 (2003).
- 10 Byeon, I. J. *et al.* Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780-790 (2009).
- 11 Bocanegra, R. *et al.* Association equilibrium of the HIV-1 capsid protein in a crowded medium reveals that hexamerization during capsid assembly requires a functional C-domain dimerization interface. *Biophys. J.* **104**, 884-893 (2013).
- 12 del Alamo, M., Rivas, G. & Mateu, M. G. Effect of macromolecular crowding agents on human immunodeficiency virus type 1 capsid protein assembly in vitro. *J. Virol.* **79**, 14271-14281 (2005).



- 13 Ehrlich, L. S., Agresta, B. E. & Carter, C. A. Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *J. Virol.* **66**, 4874-4883 (1992).
- 14 Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T. & Sundquist, W. I. Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80-83 (1999).
- 15 Ganser-Pornillos, B. K., Cheng, A. & Yeager, M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell* **131**, 70-79 (2007).
- 16 Ganser-Pornillos, B. K., von Schwedler, U. K., Stray, K. M., Aiken, C. & Sundquist, W. I. Assembly properties of the human immunodeficiency virus type 1 CA protein. *J. Virol.* **78**, 2545-2552 (2004).
- 17 Grattinger, M. *et al.* In vitro assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin A. *Virology* **257**, 247-260 (1999).
- 18 Li, S., Hill, C. P., Sundquist, W. I. & Finch, J. T. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409-413 (2000).
- 19 Pornillos, O., Ganser-Pornillos, B. K., Banumathi, S., Hua, Y. & Yeager, M. Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus. *J. Mol. Biol.* **401**, 985-995 (2010).
- 20 Pornillos, O. *et al.* X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282-1292 (2009).
- 21 Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469**, 424-427 (2011).
- 22 Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646 (2013).
- 23 Deshmukh, L. *et al.* Structure and dynamics of full-length HIV-1 capsid protein in solution. *J. Am. Chem. Soc.* **135**, 16133-16147 (2013).
- 24 Carlson, L. A. *et al.* Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **4**, 592-599 (2008).
- 25 Carlson, L. A. *et al.* Cryo electron tomography of native HIV-1 budding sites. *PLoS Pathog.* **6**, e1001173 (2010).
- 26 Wright, E. R. *et al.* Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* **26**, 2218-2226 (2007).
- 27 Ott, D. E. Cellular proteins detected in HIV-1. *Rev. Med. Virol.* **18**, 159-175 (2008).
- 28 Ott, D. E. *et al.* Cytoskeletal proteins inside human immunodeficiency virus type 1 virions. *J. Virol.* **70**, 7734-7743 (1996).
- 29 Minton, A. P. Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. *J. Pharm. Sci.* **94**, 1668-1675 (2005).
- 30 Konnyu, B. *et al.* Gag-Pol processing during HIV-1 virion maturation: a systems biology approach. *PLoS Comput. Biol.* **9**, e1003103 (2013).
- 31 Mattei, S. *et al.* Induced maturation of human immunodeficiency virus. *J. Virol.* **88**, 13722-13731 (2014).
- 32 Woodward, C. L., Cheng, S. N. & Jensen, G. J. Electron cryotomography studies of maturing HIV-1 particles reveal the assembly pathway of the viral core. *J. Virol.* **89**, 1267-1277 (2015).



- 33 Briggs, J. A. *et al.* The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure* **14**, 15-20 (2006).
- 34 Grime, J. M. & Voth, G. A. Early stages of the HIV-1 capsid protein lattice formation. *Biophys. J.* **103**, 1774-1783 (2012).
- 35 Saunders, M. G. & Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **42**, 73-93 (2013).
- 36 Levandovsky, A. & Zandi, R. Nonequilibrium assembly, retroviruses, and conical structures. *Phys. Rev. Lett.* **102**, 198102 (2009).
- 37 Yu, Z. *et al.* Unclosed HIV-1 capsids suggest a curled sheet model of assembly. *J. Mol. Biol.* **425**, 112-123 (2013).
- 38 Chen, B. & Tycko, R. Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation. *Biophys. J.* **100**, 3035-3044 (2011).
- 39 Benjamin, J., Ganser-Pornillos, B. K., Tivol, W. F., Sundquist, W. I. & Jensen, G. J. Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. *J. Mol. Biol.* **346**, 577-588 (2005).
- 40 Mattei, S. *et al.* RNA and Nucleocapsid Are Dispensable for Mature HIV-1 Capsid Assembly. *J Virol* **89**, 9739-9747 (2015).
- 41 Stokes, R. J. & Evans, D. F. *Fundamentals of interfacial engineering.* (Wiley-VCH, 1997).
- 42 Yufenyuy, E. L. & Aiken, C. The NTD-CTD intersubunit interface plays a critical role in assembly and stabilization of the HIV-1 capsid. *Retrovirology* **10**, 29 (2013).
- 43 Zhao, L. *et al.* Intracellular water-specific MR of microbead-adherent cells: the HeLa cell intracellular water exchange lifetime. *NMR Biomed.* **21**, 159-164 (2008).
- 44 Rapaport, D. C. Role of reversibility in viral capsid growth: a paradigm for self-assembly. *Phys Rev Lett* **101**, 186101 (2008).
- 45 Johnston, I. G., Louis, A. A. & Doye, J. P. K. Modelling the self-assembly of virus capsids. *Journal of Physics-Condensed Matter* **22** (2010).
- 46 Nguyen, H. D., Reddy, V. S. & Brooks, C. L. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Letters* **7**, 338-344 (2007).
- 47 Nguyen, H. D., Reddy, V. S. & Brooks, C. L. Invariant Polymorphism in Virus Capsid Assembly. *J. Am. Chem. Soc.* **131**, 2606-2614 (2009).
- 48 Smith, G. R., Xie, L., Lee, B. & Schwartz, R. Applying molecular crowding models to simulations of virus capsid assembly in vitro. *Biophys J* **106**, 310-320 (2014).
- 49 Wilber, A. W., Doye, J. P. K., Louis, A. A. & Lewis, A. C. F. Monodisperse self-assembly in a model with protein-like interactions. *Journal of Chemical Physics* **131** (2009).
- 50 Wilber, A. W. *et al.* Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *Journal of Chemical Physics* **127** (2007).
- 51 Casini, G. L., Graham, D., Heine, D., Garcea, R. L. & Wu, D. T. In vitro papillomavirus capsid assembly analyzed by light scattering. *Virology* **325**, 320-327 (2004).
- 52 Tsiang, M. *et al.* A trimer of dimers is the basic building block for human immunodeficiency virus-1 capsid assembly. *Biochemistry* **51**, 4416-4428 (2012).
- 53 Briggs, J. A., Wilk, T., Welker, R., Krausslich, H. G. & Fuller, S. D. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* **22**, 1707-1715 (2003).

- 54 Hagan, M. F. & Chandler, D. Dynamic pathways for viral capsid assembly. *Biophys J* **91**, 42-54 (2006).
- 55 Ganser-Pornillos, B. K. *et al.* Hexagonal assembly of a restricting TRIM5alpha protein. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 534-539 (2011).
- 56 Diaz-Griffero, F. *et al.* Comparative requirements for the restriction of retrovirus infection by TRIM5alpha and TRIMCyp. *Virology* **369**, 400-410 (2007).
- 57 Diaz-Griffero, F. *et al.* Modulation of retroviral restriction and proteasome inhibitor-resistant turnover by changes in the TRIM5alpha B-box 2 domain. *J. Virol.* **81**, 10362-10378 (2007).
- 58 Kutluay, S. B., Perez-Caballero, D. & Bieniasz, P. D. Fates of retroviral core components during unrestricted and TRIM5-restricted infection. *PLoS Pathog.* **9**, e1003214 (2013).
- 59 Langelier, C. R. *et al.* Biochemical characterization of a recombinant TRIM5alpha protein that restricts human immunodeficiency virus type 1 replication. *J. Virol.* **82**, 11682-11694 (2008).
- 60 Nakayama, E. E. & Shioda, T. Anti-retroviral activity of TRIM5 alpha. *Rev. Med. Virol.* **20**, 77-92 (2010).
- 61 Stremlau, M. *et al.* Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5514-5519 (2006).
- 62 Diaz-Griffero, F. *et al.* Rapid turnover and polyubiquitylation of the retroviral restriction factor TRIM5. *Virology* **349**, 300-315 (2006).
- 63 Javanbakht, H., Diaz-Griffero, F., Stremlau, M., Si, Z. & Sodroski, J. The contribution of RING and B-box 2 domains to retroviral restriction mediated by monkey TRIM5alpha. *J. Biol. Chem.* **280**, 26933-26940 (2005).
- 64 Lienlaf, M. *et al.* Contribution of E3-ubiquitin ligase activity to HIV-1 restriction by TRIM5alpha(rh): structure of the RING domain of TRIM5alpha. *J. Virol.* **85**, 8725-8737 (2011).
- 65 Stremlau, M. *et al.* The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).
- 66 Zhao, G. *et al.* Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces. *PLoS Pathog* **7**, e1002009 (2011).
- 67 Grime, J. M. A. & Voth, G. A. Highly Scalable and Memory Efficient Ultra-Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **10**, 423-431 (2014).
- 68 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, 27-38 (1996).

## Acknowledgements

This research was supported by National Institutes of Health grant P50-GM082545. The computations in this work are part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. This work is also part of the “Ultra-Coarse-Grained Simulations of Biomolecular Processes at the Petascale” Petascale Computing Resource Allocation (PRAC) support by the National Science Foundation (award number OCI-1440027).

## Figure Legends

**Figure 1.** The fullerene cone model of the HIV-1 capsid is assembled from “mature-style” capsid lattice with hexameric (green) and pentameric (red) building blocks. **(a)** CA dimer structure with both CG (tube) and all-atom (ribbon) monomer representations (NTDs green, CTDs gray). CTD dimer interface marked by dashed line, scale bar 5 nm. **(b)** Quasi-equivalent pentamer and hexamer capsid inclusions (NTDs distinguished by colour, all CTDs gray), with schematic of adjacent packing in capsid lattice. **(c)** Mature capsid structure after Pornillos et al.<sup>21</sup>. NTDs of hexamer-associated CA shown in green, with NTDs of pentamer-associated CA shown in red (all CTDs gray), scale bar 20 nm.

**Figure 2.** Dimers of the CA protein are the units of self-assembly of HIV-1 capsid-like CG structures. **(a)** Simulation snapshot of a population of metastable trimer-of-dimers. CA NTDs are blue, with CTDs gray (non-aggregated CA shown transparent for clarity). **(b)** Detail of a CG trimer-of-dimers, where “edges” are CA dimers, with two NTDs per triangle “vertex”. NTDs coloured by dimer for clarity, scale bar 10 nm. Final simulation snapshots for  $\rho_{CR} = 200$  mg/mL and  $[CA] = 2$  mM **(c)**, 3 mM **(d)** and 4 mM **(e)** are presented with NTDs coloured by monomer presence in a trimer (blue), pentamer (red) or hexamer (green). All CTDs gray, scale bar 20 nm. **(f)** Multiple aggregated capsid structures as revealed by electron cryotomography,<sup>39</sup> with capsid structures highlighted in orange **(g)**. Scale bar 20 nm, panels **f** and **g** adapted from reference 42.

**Figure 3.** Only a narrow range (indicated in green) of CA concentrations results in the nucleation and growth of a single lattice region. CG self-assembly as a function of active CA concentration  $[CA]_+$  under fixed crowding conditions. Colour in the concentration bar indicates formation of only trimer-of-dimers structures (blue) and the nucleation and growth of single (green) or multiple (red) lattice regions. Example final structures are shown for  $[CA]_+ = 1$  mM, 2 mM and 4 mM (arrows). CA colour scheme as in Fig. 2, lamellar regions highlighted by ovals. Final structure for  $[CA]_+ = 2$  mM formed via two lattice regions fusing. Gray panel shows cross-sectional slices (not to scale) to illustrate lamellar lattice in structure for  $[CA]_+ = 2$  mM (top, cross sectional plane indicated by dashed white line) and an example lamellar CA lattice inside a virion from electron

cryotomography (bottom, CA lattice indicated by white arrow). Final structures for  $[CA]_+ = 4$  mM wrap around periodic boundaries. Scale bar 20 nm.

**Figure 4.** Putative assembly from 12 CA dimers that nucleates CG capsid lattice growth. **(a)** Reversible addition of CA dimers (CA monomers depicted as black triangles, CTD/CTD interface as a dashed line) onto existing aggregates produces trimers with shared edges, eventually generating a central hexamer stabilised by trimer “skirt”. **(b)** Front and side view of example structure from CG simulation, with mild innate curvature visible. Hexamer-associated NTDs are green, NTDs in trimer skirt blue, CTDs gray. Scale bar 20 nm.

**Figure 5.** Example CG simulation data under fixed molecular crowding. **(a)** “Available” assembly-competent CA in solution for initial  $[CA]_+ = 0.5$  mM (resulting in no nucleation and growth of mature lattice) and initial  $[CA]_+ = 1.0$  mM (producing nucleation and growth of a single lattice region, see main text and Table 2). Lattice growth continues below the level of available  $[CA]_+$  required for nucleation on the same time scale. **(b)** Number of separate lattice regions containing key structural motifs for  $[CA]_+ = 1$  mM. The number of trimers in solution (blue curve) reduces significantly as a single region of lattice grows.

**Figure 6.** Steps in the assembly of the HIV-1 capsid by polymerization of CA dimers for  $[CA]_{\%} = 10\%$  and conformational switching interval of  $5 \times 10^5$  time steps (see main text). Simulation snapshots at  $120 \times 10^6$  **(a)**,  $240 \times 10^6$  **(b)**,  $440 \times 10^6$  **(c)**,  $460 \times 10^6$  **(d)**,  $600 \times 10^6$  **(e)** and  $1700 \times 10^6$  MD time steps **(f)** are shown with views perpendicular and parallel to the major structural axis. Colour scheme as in Fig. 2. Scale bar is 20 nm.

**Figure 7.** Disassembly of self-assembled CG CA lattice under simulated rapid dilution with constant molecular crowding. Conformational switching intervals of  $1 \times 10^4$  time steps **(a)**,  $1 \times 10^5$  time steps **(b)**, and  $5 \times 10^5$  time steps **(c)** are shown. Simulation snapshots (inset) correspond to the systems at  $\approx 175 \times 10^6$  **(a)**,  $\approx 550 \times 10^6$  **(b)** and  $\approx 350 \times 10^6$  **(c)** CG MD time steps. Colour scheme as in Fig. 2.

**Table 1.** CG self-assembly for varying [CA] and inert crowding levels. Triangles ( $\Delta$ ) indicate simulations producing only steady-state CA trimer-of-dimers populations (average trimer count listed, standard deviation in parentheses). Entries marked  $\diamond\diamond$  denote production of multiple lattice regions. Hexamer (bold typeface) and pentamer (italics) counts shown after  $2 \times 10^8$  CGMD time steps.

Inert crowding level $\rho_{CR}$ / mg/mL	CA concentration / mM			
	1	2	3	4
0	$\Delta$ 0.36 (0.63)	$\Delta$ 2.98 (1.67)	$\Delta$ 10.47 (3.22)	$\Delta$ 25.49 (5.29)
50	$\Delta$ 0.57 (0.75)	$\Delta$ 5.41 (2.20)	$\Delta$ 18.74 (4.03)	$\Delta$ 42.11 (6.48)
100	$\Delta$ 1.32 (1.09)	$\Delta$ 10.73 (3.14)	$\Delta$ 33.19 (6.01)	$\diamond\diamond$ <b>68</b> , 7
150	$\Delta$ 3.38 (1.72)	$\Delta$ 21.02 (4.23)	$\diamond\diamond$ <b>78</b> , 3	$\diamond\diamond$ <b>140</b> , 6
200	$\Delta$ 6.90 (2.57)	$\diamond\diamond$ <b>52</b> , 3	$\diamond\diamond$ <b>96</b> , 16	$\diamond\diamond$ <b>146</b> , 20

**Table 2.** CG self-assembly for fixed “active” CA concentrations  $[CA]_+$  under constant crowding. Triangles ( $\Delta$ ) indicate simulations producing only steady-state populations of trimer-of-dimers structures (average trimer count listed, standard deviation in parentheses). Entries marked  $\diamond\diamond$  denote the production of multiple mature-style lattice regions. Final hexamer (bold typeface) and pentamer counts (italics) are shown after  $2 \times 10^8$  MD timesteps.

“Active” CA concentration, $[CA]_+$	Simulation results
250 $\mu$ M	$\Delta$ 1.36 (1.13)
500 $\mu$ M	$\Delta$ 6.26 (2.19)
750 $\mu$ M	<b>10</b> , 0
1.00 mM	<b>29</b> , 0
1.25 mM	$\diamond\diamond$ <b>44</b> , 2
2.00 mM	$\diamond\diamond$ <b>70</b> , 7
3.00 mM	$\diamond\diamond$ <b>118</b> , 5
4.00 mM	$\diamond\diamond$ <b>146</b> , 20

**Table 3.** CG self-assembly for varying [CA]<sub>%</sub>, (see main text) with 4 mM total CA and 200 mg/mL of inert crowder. Triangles ( $\Delta$ ) denote steady-state trimer-of-dimer populations with no stable assembly of mature lattice (mean trimer count shown, standard deviation in parentheses). Entries marked  $\diamond\diamond$  indicate nucleation and growth of multiple lattice regions. Final hexamer (bold typeface) and pentamer count (italics) shown after  $3 \times 10^8$  MD timesteps unless otherwise noted in main text.

Proportion of “active” CA in solution, [CA] <sub>%</sub>	Switching interval		
	1 x 10 <sup>4</sup> time steps	1 x 10 <sup>5</sup> time steps	5 x 10 <sup>5</sup> time steps
2.5%	$\Delta$ 0.13 (0.38)	$\Delta$ 0.33 (0.67)	$\Delta$ 0.66 (0.72)
5%	$\Delta$ 0.20 (0.43)	$\Delta$ 0.83 (0.87)	$\Delta$ 1.63 (1.45)
10%	$\Delta$ 0.39 (0.66)	$\Delta$ 2.43 (1.60)	<b>100</b> , 5
25%	$\Delta$ 1.99 (1.41)	$\diamond\diamond$ <b>82</b> , 7	$\diamond\diamond$ <b>118</b> , 7
50%	$\diamond\diamond$ <b>90</b> , 4	$\diamond\diamond$ <b>121</b> , 10	$\diamond\diamond$ <b>128</b> , 13