Role of Protein Interactions in Defining HIV-1 Viral Capsid Shape and Stability: A Coarse-Grained Analysis

Vinod Krishna, Gary S. Ayton, and Gregory A. Voth*
Center for Biophysical Modeling and Simulation, and Department of Chemistry, University of Utah, Salt Lake City, Utah

ABSTRACT Coarse-grained models of the HIV-1 CA dimer are constructed based on all-atom molecular dynamics simulations. Coarse-grained representations of the capsid shell, which is composed of ~1500 copies of CA proteins, are constructed and their stability is examined. A key interaction between carboxyl and hexameric amino terminal domains is shown to generate the curvature of the capsid shell. It is demonstrated that variation of the strength of this interaction for different subunits in the lattice can cause formation of asymmetric, conical-shaped closed capsid shells, and it is proposed that variations, in the structure of the additional carboxyl-amino terminal binding interface during self-assembly, are important aspects of capsid cone formation. These results are in agreement with recent structural studies of the capsid hexamer subunit, which suggest that variability in the binding interface is a cause of the differences in subunit environments that exist in a conical structure.

INTRODUCTION

Viral capsids are protein shells that enclose the genetic material of the virus. During infection of a host cell, the capsid shell is disassembled to release the genetic material, and in the later stages of infection, the shell is reassembled, attaining its final shape during viral maturation (1–7). Interfering with either the assembly or disassembly of the viral capsid shell has been shown to result in reduced viral infectivity (8–14). Consequently, the viral capsid has gained increasing importance as a target for antiviral agents.

Viral capsids have a highly regular and symmetric lattice structure (15). The structure of many viral capsids can be explained through the principles of quasiequivalence formulated by Crick and Watson (16) and Caspar and Klug (17). Under these theories, viral capsid subunits adopt different local conformations to combine into hexameric and pentameric subunits that form an icosahedral capsid cage. By Euler's theorem, the presence of 12 pentameric subunits is necessary to form a closed polyhedral structure. The shape of the polyhedral structure is then determined by the placement of the 12 pentameric subunits. In icosahedral capsids, the pentameric subunits are placed symmetrically such that each subunit has an identical environment. However, notable exceptions to icosahedral viral capsid structures exist. These include the capsids found in retroviruses, for example the conical capsid found in the HIV-1 viral particle (18,19). Based on the available evidence, the HIV-1 capsid cone is also considered to consist of hexameric and 12 pentameric subunits, with seven of the several hundred pentamers located at the broad end of the cone and the remaining five pentamers located at the narrow end (19). HIV-1 capsids are thought to be formed according to the same rules as those for a fullerene cone structure. The presence of hexameric lattices in HIV-1 and related lentivirus capsids have been recently verified (20–22).

The HIV-1 viral capsid appears to have a simple and regular structure; however, the physical principles responsible for the formation of a conical shape are still unknown, and have been the subject of numerous theoretical and computational studies (23–26). Experiments have demonstrated the self-assembly of HIV-1 viral capsid proteins into cones and cylinders both in vitro and in vivo, although the environments for self-assembly in vitro and in vivo are quite different (14,18,19). Furthermore, capsid proteins self-assemble in vitro into cones and cylinders in both the presence and the absence of viral RNA (27,28), indicating that capsid formation is a self-assembly process driven mainly by interactions between the constituent capsid proteins.

The formation of the conical capsid occurs over multiple scales. As capsid formation is a self-assembly process, local interactions between CA monomers (to be defined shortly) play an important role in determining the outcome of the assembly. The CA monomer is a 23-kDa protein with an amino (NTD) and carboxyl terminal domain (CTD) that are connected by a short-linker region. Experimentally, suitable mutations of either domain either disrupted cone formation or resulted in the formation of other capsid shapes (27,29). These experiments therefore provide evidence that capsid formation is sensitive to the detailed protein-protein interactions between CA protein subunits and further reinforce the concept that capsid formation is a self-assembly process (13).

The experimental evidence implies that modeling approaches to understand the self-assembly of the HIV-1 capsid must accurately include the local capsid protein interactions that are crucial to the self-assembly process. On the other hand, capsid self-assembly involves hundreds of association events between individual CA proteins. As a result the timescales over which capsid self-assembly takes place

Submitted August 26, 2009, and accepted for publication September 25, 2009.

*Correspondence: voth@chem.utah.edu

Editor: Nathan Andrew Baker.

0006-3495/10/01/0018/9 $2.00

doi: 10.1016/j.bpj.2009.09.049

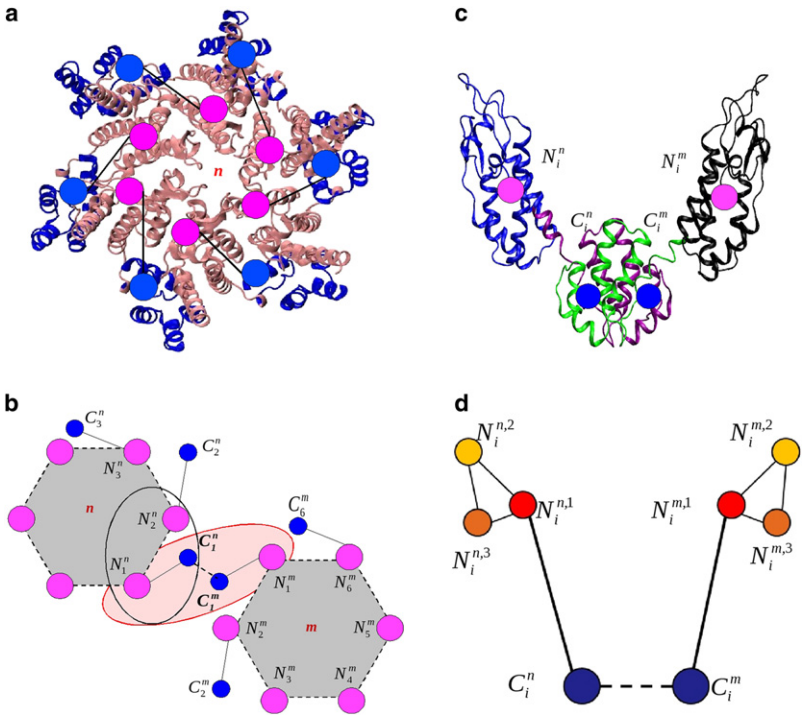

FIGURE 1 A schematic of the capsid lattice structure. A capsid hexamer (*a*) is represented by six monomers labeled with subscripts, each having a NTD (*pink*) and a CTD (*blue*). Neighboring hexamers, *n* and *m* (*superscripts*), bind each other by a dimeric interface, ($N_i^n : C_i^n : C_i^m : N_i^m$), between two CTDs as shown schematically in the shaded (*pink*) region on panel *b*. The other circled region in panel *b* is a schematic of the hexameric NTD-CTD interface, $N_{i+1}^n - C_i^n$, that is represented as a three-body potential involving the circled atoms, in CG simulations. (*c*) Atomistic structure of the capsid dimer with a four-site CG dimer representation superimposed. Each NTD is represented by a single site (*pink*) and the same is true for each CTD (*blue*). (*d*) Schematic of an eight-site CG model of the capsid dimer. Each NTD is modeled by three sites whereas a single-site model (*blue spheres*) is retained for each of the CTDs.

are long in comparison to the timescales that govern individual CA protein dynamics. Consequently, it is extremely difficult to directly study this self-assembly process at the molecular level and to draw inferences regarding the factors that determine capsid shape and symmetry. Furthermore, since there is such a great separation in time- and length scale between individual CA association events driven by protein-protein interactions, and the assembly of the capsid lattice as a whole, it is a challenging problem to specifically relate the influence of individual CA protein-protein interactions on capsid stability, shape, and symmetry.

This work addresses the problem of relating CA protein interactions to structural and shape properties of self-assembled HIV-1 and similar retroviral capsids. We investigate the role of different CA protein-protein interactions in controlling the shape of the HIV-1 capsid, and the physical factors that govern capsid lattice symmetry and structure. The relationships between specific CA protein-protein interactions and shape are predicted.

To make the connection between CA protein-protein interactions and structure, a coarse- graining approach is used here wherein CA protein-protein interactions are mimicked through interactions between coarse-grained (CG) sites chosen to represent capsid protein structure. Then, by altering different representative protein-protein interactions in CG models of the capsid, we analyze the roles of different CA protein-protein interactions in determining capsid shape and stability. Using a similar strategy, inferences are also made regarding the physical factors that determine capsid lattice symmetry and structure. Where possible, experimentally available structures of capsid protein complexes are simulated with molecular dynamics (MD) methods to determine the optimal location of the CG sites, and the MD simulations are used to determine strengths of effective interactions between various CG sites. Remaining interaction parameters that cannot be determined from MD are treated as adjustable parameters, and their strengths are altered to infer their importance in determining capsid structure and shape. All the effective interactions studied through coarse-graining are chosen to represent experimentally known or inferred protein-protein interactions as closely as possible.

The following sections of this article are organized as follows. In Methods, we explain the techniques developed and used to construct CG models of the HIV-1 CA dimer, based on MD simulations of a model built from structures of the dimerizing CTD and the NTD. The Results section presents CG MD simulations performed on CG models of the HIV-1 CA structure, showing the effects of varying strengths of different CG effective protein-protein interactions on the structure of the capsid. These simulations and their results are further addressed in the Discussions section, where predictions are made regarding the relation among CA protein-protein interactions and capsid shape, stability, and lattice structure.

METHODS

The capsid lattice consists of CA hexamers, each formed by six monomeric NTDs (21). Each CA hexamer is surrounded by another six hexamer neighbors. Two such neighboring hexamers are bound to each other, and into a lattice, through the dimerization of corresponding monomeric CTDs (30,31). The structure and interactions are illustrated in Fig. 1. The individual hexameric units are unstable in solution, and so HIV-1 CA proteins

exist predominantly in their dimeric form in such environments (30). The capsid lattice contains two main packing interfaces: The NTD-NTD interface that exists in the CA hexamer, and the CTD-CTD dimer interface that binds the CA proteins from adjacent hexamers (see Fig. 1, *a* and *b*). Additionally, experimental evidence shows (21,32) there is a third CTD-NTD interface type in the capsid lattice that is contiguous with the NTD-NTD interface (see Fig. 1, *a* and *b*).

The following notation will be employed: N_i^n will refer to the i^{th} NTD CG site on hexamer n. Likewise, C_i^n will designate the i^{th} CTD CG site on hexamer n and is shown in Fig. 1. It is possible to express the total interaction energy of N_T CG hexamers as written:

$$
\begin{aligned}
V_{\text{total}}^{CG} = & \sum_{n=1}^{N_T}\left[\sum_{i=1}^{6} V_{N_i^n C_i^n}^{\text{bond}} + \sum_{m=1}^{n}\sum_{i=1}^{6}\sum_{\substack{j=1\\ i>j,\\ m=n}}^{6} V_{N_i^n N_j^m}^{\text{pair}} + V_{C_i^n C_j^m}^{\text{pair}} + V_{N_i^n C_j^m}^{\text{pair}}\right] + \sum_{n=1}^{N_T}\sum_{m=1}^{n}\sum_{\substack{i,j=1\\ j<i\\ m=n}}^{6}\left[V_{N_i^n C_i^n C_j^m}^{\text{3body}} + V_{N_j^m N_i^n C_i^n}^{\text{3body}}\right. \\
& \left. + \sum_{l,k<j,l=m}^{6} V_{N_j^m N_i^n N_k^l}^{\text{3body}}\right] + \varepsilon\sum_{n=1}^{N_T}\left[\sum_{\substack{\alpha,\beta=1\\ \beta<\alpha}}^{3} V_{N_i^{n,\alpha} N_i^{n,\beta}}^{\text{bond}} + \sum_{\substack{m\le n,\\ l\le m}}^{N_T}\sum_{i=1}^{6}\sum_{\substack{j=1,\\ j<i,\\ m=n}}^{6}\left[V_{N_i^{n,1} N_i^{n,2} C_j^m N_i^{n,3}}^{\text{torsion}} + \sum_{\substack{k=1\\ k<j,\\ l=m}}^{6} V_{N_i^{n,2} C_k^l C_j^m N_j^{n,1}}^{\text{torsion}}\right.\right. \\
& \left.\left. + \sum_{\alpha=1}^{3} V_{N_i^{n,\alpha} N_j^{m,\alpha} N_k^{l,\alpha}}^{\text{3body}}\right]\right]. \qquad (1)
\end{aligned}
$$

In Eq. 1, $6N_T$ is the total number of monomers. The value $\varepsilon = 0$ applies for simulations of the four-CG site dimer (two sites per CG monomer) model, and is 1 otherwise (e.g., for the eight-site dimer CG model). The four sites labeled $(N_i^{n,1}, N_i^{n,2}, N_i^{n,3}, C_i^n)$ involve summations over additional indices α,β in Eq. 1 that denote the additional sites for each CG NTD. The cross terms $V_{N_i^n C_j^m}^{\text{pair}}$ are turned off for simulations involving the three-body contributions in Eq. 1 to save on computational time. Test simulations were run to confirm that they had no effect on the results of the CG simulations. The pair and three-body potentials had a distance cutoff of 8 nm, whereas the torsional potential has a distance cutoff of 4 nm. An important difference between the CG model potential in Eq. 1 and actual protein interactions is that, in contrast to protein-protein interactions in which individual proteins bind each other through specific and complementary interfaces (and which do not generally participate in forming other protein-protein interfaces), the CG potentials are nonspecific interactions. These nonspecific interactions allow, in principle, for interactions between arbitrary numbers of sites (CG interfaces), as long as they are within the appropriate distance cutoffs that are described later in this section. The distance cutoffs are the same as traditional cutoffs in the definition of nonbonded potentials used in MD simulations.

For the higher-resolution CG model (Fig. 1 *d*) described later in this work, additional three-body angular interactions are introduced to effectively represent the ''honeycomb'' hexagonal lattice symmetry of the underlying molecular lattice, and these additional interactions are given by the last term in Eq. 1. In addition to this, the other summations involving CG sites representing NTDs acquire additional indices corresponding to interactions of the different types of sites between each pair of NTDs. For example, the pair potential term, $V_{N_i^n N_j^m}^{\text{pair}}$, becomes $V_{N_i^{n,\alpha} N_j^{m,\beta}}^{\text{pair}}$ and the summation is extended and evaluated for all pairs of NTD CG site types (α,β). The three-body potentials are always required to involve sites of the type $N_i^{n,1}$, which bond to the CTD CG site C_i^n through the bonded potential, $V_{N_i^{n,1} C_i^n}^{\text{bond}}$ (Fig. 1 *d*). Hence the effective CTD-NTD potential, $V_{N_j^m N_i^n C_i^n}^{\text{3body}}$ is replaced by the summation, $\sum_{\beta=2,3} V_{N_j^{m,\beta} N_i^{n,1} C_i^n}^{\text{3body}}$, and the three-body potential $V_{N_i^n C_i^n C_j^m}^{\text{3body}}$ is replaced by $\sum_{\alpha=1}^{3} V_{N_i^{n,\alpha} C_i^n C_j^m}^{\text{3body}}$.

The types of CG potential functions used in each term in Eq. 1 are provided in Table 1. The torsional potential used can be understood from Fig. 1. CTD domains belonging to a capsid monomer are involved in the CTD-CTD dimer and the hexameric NTD-CTD interfaces. Hence torsion about the CTD-CTD dimer, or the NTD-CTD interface directly affects the hexameric NTD-CTD interface.

The general strategy to identify and construct CG representations requires the identification of CG site positions within each protein domain and the determination of reasonable values for the CG effective potential parameters. These two tasks are discussed below. Since experimental structures of the CA dimer are available, and since CA proteins exist in dimeric form when in solution, atomistic MD simulations of the dimer were performed and CG sites identified based on the simulations.

TABLE 1 A listing of the potentials used for each interaction

CG model type	Interface	Interaction	Potential type
Two-site	CTD-CTD	$(C_i^n : C_j^m)$	LJ (pair, Eq. 2)
	NTD-NTD	$(N_i^n : N_j^m)$	LJ (pair, Eq. 2)
	NTD-CTD	$(N_i^n : C_i^n$, dimer)	Harmonic (pair, Eq. 3)
	NTD-CTD-CTD	$(N_i^n : C_i^n : C_j^m$, dimer)	Three-body angular (Eq. 5)
	Hexameric NTD-CTD	$(N_j^m : N_i^n : C_i^n)$	Three-body (Eq. 4)
Four-site (additional potentials only)	NTD-CTD-CTD-NTD	$(N_i^{n,1} : C_j^m : C_k^l : N_k^{l,2})$ or $(N_i^{n,1} : N_i^{n,2} : C_j^m : N_i^{n,3})$	Torsion (four-body, Eq. 7)
	NTD-NTD-NTD	$(N_i^{n,\alpha} : N_j^{n,\alpha} : N_k^{n,\alpha})$, $\alpha = 1,2,3$	Three-body (Eq. 6)
	Intra NTD	$(N_i^{n,1}, N_i^{n,2}, N_i^{n,3}$, Fig. 1)	Harmonic (pair, Fig. 1, Eq. 3)

Shown as lines between sites in Fig. 1. For the four-site monomer model, the interactions listed are in addition to those included from the two-site monomer model.

TABLE 2 Intradimer position and potential parameters from atomistic and CG simulations for the two-site model

Interface/interaction	K_b (kcal/mol nm^2)	Equilibrium length (nm)	CG length (nm)
CTD-NTD ($C_i^n : N_i^n$)	0.19 (CG:0.185)	4.031	3.90
CTD-CTD ($C_i^n : C_i^m$)	0.99	2.148	2.85

The four-site model parameters are similar to these, and can be found in Table S1.

Identifying coarse-grained sites

To identify the CG sites, we explored two complementary methods. In the first, atomistic trajectories obtained by MD simulation of the CA dimer in solution were analyzed using quasiharmonic analysis (QHA) (33–35) (results are summarized in the Supporting Material). To obtain an idea of the frequency spectrum of the dimer dynamics, we performed a QHA of the center-of-mass motion of the helices and β-loops in the capsid dimer. Based on the QHA, a lower-resolution four-site model of the CA dimer (two CG sites per monomer) was identified (see Fig. 1 *c*). For comparison, the essential-dynamics coarse-graining method (36) was also used to construct a four-site model of the capsid dimer. The domains identified for a four-site CG model by this method were found to be nearly identical to the domains identified through the QHA analysis, thus validating the CG mapping for the present four-site model.

Calculation of CG model effective potentials

CG variables represent collective degrees of freedom of several atoms in a protein, hence their interactions must be chosen to describe large-scale protein motions, whereas the effects of local fluctuations of the individual atoms are represented in an averaged, effective fashion (37). Thus, fluctuations in each of the CG variables around their mean values were calculated from the structures obtained from atomistic MD. Intradimer CG potential parameters were then obtained from the functional fit to probability distributions evaluated from the atomistic ensemble. This ensures that the CG potentials obtained are thermodynamic averages over local modes of motion. The Lennard-Jones (LJ) potential parameters used to represent CG CTD-CTD ($C_i^n : C_j^m$) interactions were evaluated from CG bond lengths by equating the effective harmonic potential spring constants obtained from the atomistic MD to the second derivative of the LJ potentials at their minimum positions. An LJ potential modeling the hexameric NTD binding interactions ($N_i^n : N_j^m$) was added with a well-depth of 4 kcal/mol and a bond length to approximate the distance between the centers of the hexameric NTDs in the HIV-1 capsid lattice. This well-depth was chosen to keep the binding strength of the hexameric NTDs similar to the CG CTD-CTD interaction. The LJ potentials used for the CTD dimer interface in CG simulations, with a well-depth ε and a zero at σ, have the functional form

$$V^{\text{pair}}_{C_i^n C_j^m} = 4\varepsilon\left[\left(\frac{\sigma}{r_{C_i^n C_j^m}}\right)^{12} - \left(\frac{\sigma}{r_{C_i^n C_j^m}}\right)^{6}\right]. \quad (2)$$

The functional form used for the NTD binding interaction (between sites $N_i^n : N_j^m$) is the same as Eq. 2, with the CTD-CTD CG distances $r_{C_i^n C_j^m}$ replaced by the distances between NTD sites, and the potential parameters correspondingly replaced by those for the NTD pair potential. The harmonic pair potential used to bind the monomeric NTD and CTD CG sites together is

TABLE 3 LJ pair interaction parameters for the two-site CG model

CG interface	LJ depth, ε (kcal/mol)	σ (nm)
C_i^n:C_j^m	4.0	2.8
N_i^n:N_j^m	4.0	3.3

$$V^{\text{pair}}_{N_i^n C_i^n} = \frac{K_b}{2}\left(r_{N_i^n C_i^n} - r_b\right)^2. \quad (3)$$

The distance, r_b, is the equilibrium bond length between CG sites N_i^n and C_i^n. The K_b value is the spring constant. We also used the functional form of Eq. 3 to describe interactions between CG sites for a given NTD in higher-resolution CG models of the capsid (discussed later in this article). The type of potential used for different interfaces (Fig. 1) is given in Table 1, and the parameters are provided in Tables 2–4.

Four-site dimers

In addition to the pair potentials from Eqs. 2 and 3 (see Table 1), we used a three-body interaction to represent the hexameric NTD-CTD interface simulations involving two-site CG models of the CA monomer (four sites for a dimer, see Fig. 1 *d*). This potential involves CG sites in the sequence$N_j^m - N_i^n - C_i^n$, as shown in Fig. 1. The angular part of this potential involves the angle,$\varphi_{N_j^m N_i^n C_i^n}$, subtended around the central site, N_i^n, by the other two sites, and the spatial part is a function of the spatial distances, $r_{N_j^m N_i^n}$,$r_{N_i^n C_i^n}$, from the central CG site,N_i^n. The functional form is

$$V^{\text{3body}}_{N_j^m N_i^n C_i^n} = \frac{K_\varphi}{2}\left(\varphi_{N_j^m N_i^n C_i^n} - \varphi_0\right)^2 \exp\left(-\frac{r^8_{N_j^m N_i^n} + r^8_{N_i^n C_i^n}}{a^8}\right). \quad (4)$$

Similarly, the $N_i^n - C_i^n - C_j^m$ dimeric angular interaction has the functional form

$$V^{\text{angle}}_{N_i^n C_i^n C_j^m} = \frac{K_\theta}{2}\left(\theta_{N_i^n C_i^n C_j^m} - \theta_0\right)^2 \exp\left(-\frac{r_{N_i^n C_i^n} + r_{C_i^n C_j^m}}{r_0}\right). \quad (5)$$

Except for the damping constants (a, r_0), the three-body parameters in Eqs. 4 and 5 were treated as free variables. Fig. 2 displays the final CG structures resulting from different values of the three-body spring constants. We performed simulations for different values of the angles, θ_0,ϕ_0 between 40 and 90°, and the results were found to be insensitive to their values. The damping constants (a, r_0) in Eqs. 4 and 5 were chosen to be ~11 nm.

Eight-site capsid dimers

Results from studies of the capsid lattice constructed from the two-site CG monomers described later indicate that a higher resolution CG model is required to describe the honeycomb p6 lattice symmetry properties of the CA lattice. Thus, CG potentials were also constructed for a four-site (three-site NTD and one-site CTD) monomer model (eight-site dimer), shown schematically in Fig. 1 *d*, of the CA protein. These potentials are described here.

TABLE 4 Three- and four-body potential parameters for the eight-site CG dimer

Interface	Interaction	Parameters	Potential type
NTD-CTD-CTD-NTD	($N_{i+1}^{n,1} : C_i^n : C_i^m : N_i^{n,2}$) in Fig. 1	3.2 kcal/mol (K_ω)	Torsion(four-body, Eq. 7)
NTD-NTD-NTD	($N_i^{n,\alpha} : N_j^{m,\alpha} : N_k^{l,\alpha}$), α = 1,2,3 in Fig. 1	1.2 kcal/mol deg^2(K_ϕ in Eq. 6)	Three-body(Eq. 6)
Inter NTD	(pair,$N_i^{n,\alpha} : N_{i+1}^{n,\beta}$, with α, β = 1,2,3)	4.0 kcal/mol(ε in Eq. 2)	LJ (Eq. 2)
Intra NTD	($N_i^{n,\alpha}$,$N_i^{n,\beta}$), $\alpha \neq \beta$ as in Fig. 1	0.185 kcal/mol nm^2(K_b in Eq. 3)	Harmonic(pair, Fig. 1, Eq. 3)

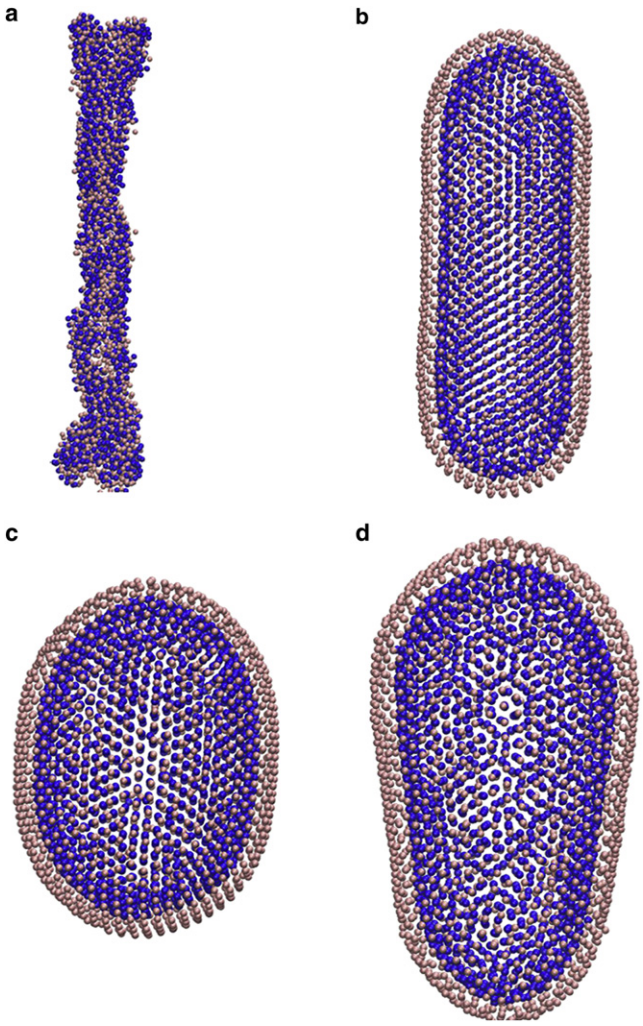

FIGURE 2 Final structures resulting from CG simulations of a two-site CG representation of the HIV-1 CA protein. (*a*) Structure resulting from a model with only intradimer and NTD-NTD pairwise interactions. (*b*) Structure resulting from a model with intradimer, NTD-NTD and an additional NTD-CTD three-body interaction (Eq. 4). The harmonic spring constant is $K_\phi = 2.24$ kcal/mol deg^2. (*c*) Structure resulting from a model with the same set of interactions as panel *b* except that the NTD-CTD interaction is stronger ($K_\phi = 5.24$ kcal/mol deg^2, Eq. 4). (*d*) Resulting structure from simulation with a linearly varying NTD-CTD interaction along the axis of symmetry ($K_\phi = 1.3$–3.9 kcal/mol deg^2).

For the three-site NTD representations to favor a honeycomblike p6 lattice, a three-body NTD-NTD-NTD potential was introduced. This potential decays exponentially in distance, while its angular term has minima at 120°. The potential is designed to represent NTD-NTD hexameric packing interactions in the CA lattice and is based on similar potentials used to simulate liquid crystal symmetries (38). The mathematical form of this potential is given below, with $\phi = \theta_{N_i^{n,\alpha}N_j^{m,\alpha}N_k^{l,\alpha}}$,

$$V^{3body}_{N_i^{n,\alpha}N_j^{m,\alpha}N_k^{l,\alpha}} = \frac{K_\varphi}{2}\left[1-\cos(6\varphi)+\cos(\varphi)-\exp\left(-100(\varphi-108)^2\right)\right]\exp\left(-\frac{r_{N_j^{m,\alpha}N_i^{n,\alpha}}+r_{N_j^{m,\alpha}N_k^{l,\alpha}}}{a}\right). \quad (6)$$

The potential described above includes a shallow minimum at an angle of 108°, as described by the first exponential term, to allow for the possibility of forming pentamers. The parameter, a ~11 nm in Eq. 6, is used to attenuate the spatial component of the potential.

Additionally, for some simulations involving the three-site model of the NTD, an additional short-ranged, four-body intermolecular potential was used, which has the form of the planar inversion potential in the program DL_POLY (39). The functional form of one such potential is given by

$$V^{torsion}_{N_i^{n,1}C_i^nC_j^mN_j^{m,2}} = \frac{K_\omega}{2}[1-\cos(\omega)]. \quad (7)$$

The angle, ω, is the angle made by one of the CG sites relative to the plane formed by the other three sites. The potential approximates torsion due to the angle made by the plane formed by the (nonspherical) NTD-CTD interface and the CTD-CTD dimer axis, or the orientation of the triangular NTD domain relative to the interfacial NTD-CTD axis, depending on the choice of CG sites involved. This potential, $V^{torsion}_{N_i^{n,1}N_i^{n,2}C_j^mN_i^{n,3}}$, has the same functional form as Eq. 7.

With this CG force field, equilibrium MD simulations of the capsid CG lattice were performed using the DL_POLY package (40). The corresponding parameters are provided in Table 4. The capsid CG lattice was constructed from ~1470 CG capsid monomers, and different initial lattice structures were used. The MD simulations were performed under constant NVT conditions, with the temperature set to 310 K using a Berendsen thermostat (41). No periodic boundary conditions were necessary. A typical CG simulation was equilibrated for ~60 ps using velocity rescaling, followed by MD trajectory lengths of several tens of picoseconds while using a timestep of 3 fs. It should be noted that the CG simulations are significantly accelerated, so their simulation time does not correspond to a real physical timescale, which is much longer.

RESULTS

CG capsid model potentials

Because each capsid monomer, in effect, forms two interfaces corresponding to the contiguous NTD-CTD, NTD-NTD (32), and CTD-CTD binding interfaces, the minimal four-site CG model of the capsid dimer was first used with CG parameters obtained as described in the Methods section (see Tables 2 and 3).

A QHA performed on the trajectories obtained by considering center-of-mass motions of the CA helices found that the largest amplitude motions of the dimer molecules involved collective motion of individual CTD and NTDs relative to each other. Thus, the linker region connecting the capsid CTD and NTD provides considerable flexibility and enables the motion of the N-terminal domains in the dimer. The flexibility inherent in the NTD motion enables packing interactions and binding between different CA subunits, thus aiding in capsid self-assembly. This is supported by the conclusions from the CG MD simulations discussed below, which implicate variations in the additional CTD-NTD binding interface as being crucial to formation of stable, conical capsids.

Capsid shape is controlled by the CTD-NTD protein interface

The CG models described above were used to construct CG representations of the capsid lattice. To understand the

factors that determine capsid shape and stability, CG MD simulations were performed to obtain equilibrium structures of the capsid lattice, starting from different initial configurations. The results from the simulation studies are discussed below.

Low-resolution model structures

Irrespective of the initial configuration and strengths of the effective interactions, two-dimensional capsid structures remained unstable and collapsed into a disordered aggregate when only the effective CG pair potentials (Eq. 1) corresponding to CA two-body dimer interactions were considered. An example of the final structure resulting from an equilibration simulation with an initial cylindrical capsid lattice open at both ends is shown in Fig. 2 *a*. These results do not change qualitatively on adding angular intradimer interactions (i.e., the terms $V^{3body}_{N_i^n C_i^n C_j^m}$ in Eq. 1), and capsid shells continued to be unstable. These calculations imply that the effective interactions present solely due to the capsid CTD dimer at the two-site per monomer CG level are insufficient to describe capsid structural properties and cannot stabilize the capsid lattice cage at this level of CG resolution.

The CG capsid structures dramatically stabilized on adding an additional three-body interaction (Fig. 1 *b*). This interaction, $V^{3body}_{N_j^m N_i^n C_i^n}$ (Eq. 4), represents the effective CTD-NTD interfacial potential, and it constrains the NTD-NTD-CTD angle (for example, the angle labeled in the sequence $N^n_{i+1} - N^n_i - C^n_i$ in Fig. 1 *b*). The addition of this interaction was motivated by structural studies of the CA hexameric subunit (21,32,42), which show the presence of an additional hexameric NTD-CTD intermolecular binding interface contiguous with the NTD-NTD binding interface in the capsid lattice. CG MD simulations of the capsid lattice with such an effective potential included generated closed, double-walled capsid shells as the equilibrium structure (Fig. 2, *b* and *c*). These capsid shells are generally axially symmetric and resemble ellipsoidal structures. However, the surface lattice structure of these closed shells is an hexagonally close packed lattice with each CG amino NTD atom surrounded by six neighboring NTDs (i.e., they are not a p6 symmetry honeycomb lattice as believed to be dominant in the HIV-1 capsid). These results are illustrated in Fig. 2, *b–d*. The corresponding pair distribution functions for the NTD and CTD CG sites show hexagonal symmetry and are plotted in the Supporting Material.

The result from the present CG simulations (i.e., that an additional interface between CA CTDs and adjoining hexameric NTDs is required for stable, closed capsid shells) is seen in recent structural studies of the capsid lattice, which have also found an additional CTD-NTD binding interface between capsid monomers (21). The destabilization of the capsid structure in the absence of this interaction seen in our CG MD simulations also explains how small-molecule assembly inhibitors might function (9,10). It is known that these inhibitors disrupt portions of the capsid NTD, which in turn participate in this additional binding interface (9,10). Thus, inhibition can occur by disrupting this additional interaction, which stabilizes and then acts as a source of curvature for the capsid lattice structure.

To probe its role in determining capsid shell curvature, we varied the strength of the stabilizing three-body NTD-CTD interaction. The starting configuration for these CG simulations was an open capsid cylindrical lattice. Increasing the strength of the interaction resulted in closed shells of different curvatures (see Fig. 2, *b* and *c*). Stronger interactions caused stable structures that are more globular in shape (Fig. 2 *c*), whereas a weaker three-body potential gave rise to more ellipsoidal closed shells (Fig. 2 *b*). This result further supports the assertion that the NTD-CTD interaction determines the curvature of the capsid shell. However, the structures formed in these CG simulations were always symmetric at both ends of the capsid shell. Because none of the interactions used broke this symmetry, it was conjectured that cone formation could be induced if the effective three-body CG NTD-CTD interaction (Eq. 4) is modulated locally along the axis of the cylinder, so as to break cylindrical symmetry (i.e., the interaction increasingly strengthens in going from the top to the bottom of the capsid). To be more specific, the value of K_ϕ in Eq. 4 was varied by a factor-of-3 from a minimal value of 1.3 kcal/mol deg^2 from the top to the bottom of the capsid. (The resulting capsid cone-structure is shown in Fig. 2 *d*.) It seems clear that varying the strength of the interaction in a linear manner along the axis of the capsid, to break cylindrical symmetry, resulted in capsid cone formation. This result will be revisited later in the Discussion.

To additionally test the effect of the initial configuration of CG CA dimers on the nature and shape of the final capsid structure, simulations were performed with different initial configurations. These included locally disordered and approximately cylindrical structures generated from short CG simulations with the stabilizing NTD-CTD interaction turned off, as well as partially spherical, completely spherical, and open cone conformations. These calculations also resulted in closed cylindrical or ellipsoidal double-walled capsid shells with hexagonally close-packed lattices. All of these CG simulations, however, resulted in regular, axially symmetric equilibrium capsid shells. The results of these calculations can be found in the Supporting Material.

Lattice symmetry and higher-resolution CG representations

CG studies described in the previous section resulted in highly regular capsid structures with hexagonal lattice symmetry. However, in contrast to viral p6 honeycomb lattice symmetry, we always found the lattice symmetry of the CG capsid particles to be close-packed and hexagonal. Higher symmetry occurs because capsid NTDs are represented by CG particles with spherically symmetric potentials and nonspecific interactions. Within the four-site CA dimer CG model, this excess of symmetry could not be corrected

by the addition of NTD-NTD interactions (Eq. 1), which restrained the angle between adjacent NTDs to be 120°. Addition of these interactions still resulted in the formation of close-packed lattices.

As the interaction potentials are radially symmetric in the spatial coordinates, it is difficult to overcome the occurrence of higher symmetry lattices. Consequently, a more fine-grained representation of the capsid NTD was used to better approximate the NTD shape of the real protein (i.e., a somewhat higher-resolution CG model). It was found that a CG representation having at least three sites corresponding to the NTD was necessary to overcome the hexagonal close-packing symmetry (an overall eight-site CA dimer CG model). The higher-resolution model allowed for the formation of lattice structures with the correct hexagonal symmetry as discussed below. This result lends further credence to the hypothesis that capsid lattice symmetry is determined by the shape of the NTD, in addition to the interfaces that are available for binding with other proteins. The results from simulations with an eight-site CG representation of CA dimers (see Fig. 1 *d*) are discussed below.

The single-site description for the CTD of the capsid monomers was retained, whereas the NTD was represented as a triangle of three CG sites bonded together with effective harmonic potentials (Fig. 1, *c* and *d*). As before, capsid cylinders constructed out of this higher-resolution model collapsed in the absence of stabilizing CTD-NTD three-body interactions and formed two-dimensional ordered lattice structures upon adding these interactions. Even in this model the absence of a CG potential (such as in Eq. 6) to enforce constraints on the angle between hexameric NTDs results in partially close-packed configurations, wherein at least one of the NTD sites packs into a hexagonal, close-packed lattice. To offset this and induce a hexameric p6 lattice, we found the additional NTD-NTD three-body angular interaction (Eq. 6) to be essential to constrain the angle between NTD domains to be ~120°. This CG potential effectively represents the NTD helix packing constraints in the real capsid lattice, but at a significantly lower resolution.

Upon adding the NTD-NTD three-body interaction (Eq. 6), we generated lattice structures that had overall p6 lattice symmetry (see Fig. 3 and note the NTD CG site color scheme from Fig. 1 *d*). The cylinders generated by these potentials have an overall flattened shape (Fig. 3 *a*), which is because the simple NTD CG potentials that impose honeycomb symmetry also impose planarity. To correct for this, we introduced additional interactions (Eq. 7) corresponding to orientation of the triangular NTDs at the NTD-CTD interface and torsion at the CTD-CTD interfaces (see Eq. 1, Fig. 1, *c* and *d*). These torsional interactions are approximately representative of the torsion present in the atomistic capsid lattice CTD-CTD and hexameric CTD-NTD interfaces. The introduction of this interaction offsets the flattening of the CG lattice structures, and for strong dihedral interactions, it retains the p6 symmetry of the capsid lattice

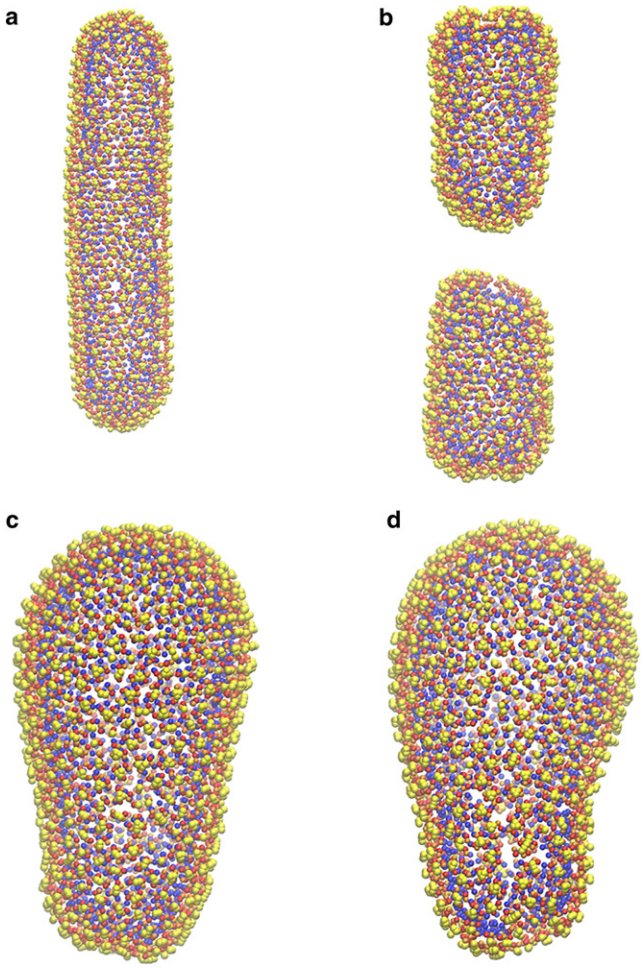

FIGURE 3 Final structures obtained from equilibrium CG simulations for eight-site dimer CA model (three-site NTD and one-site CTD). (*a*) CG cylinder with p6 hexagonal symmetry obtained from a simulation with two- and three-body CTD-NTD potentials. (*b*) Structures resulting from an addition of a CTD-CTD dihedral interaction ($K_\omega = 3.24$, Eq. 7) to enable curvature of cylinder. (*c*) Cone structure resulting from a modulation of three-body CTD-NTD potential ($K_\phi = 2.5$–7.6 kcal/mol deg^2 (Eq. 4), with a corresponding CTD-NTD orientational interaction($N_{i+1}^{n,1} : N_{i+1}^{n,2} : C_i^n : N_{i+1}^{n,3}$; K_ω = 3.24 kcal/mol, Eq. 7). Snapshots of the lattice structures of the cylinders. In panels *a* and *b* are shown in Figs. B4 and B5 of the Supporting Material. (*d*) Structure resulting from modulation of NTD-CTD interactions with CTD-CTD effective dihedral interaction included (Eq. 7).

(Fig. 3 *b* and Fig. B4 and Fig. B5 of the Supporting Material). We also used purely angular NTD-CTD-CTD interactions (Eq. 5) to examine whether the flattening could be similarly offset. Our simulation results indicate that although there is some change in curvature and the cylinder shape becomes disordered and irregular, the cross section of the cylinder remains flattened.

As before, the key three-body CTD-NTD CG interaction was also varied linearly along the axis of (rotational) symmetry to explore the possible effect of such a variation (i.e., one that gets stronger in going from top to bottom of the capsid). This resulted again in the formation of closed,

conical capsid structures (Fig. 3, *c* and *d*). Unlike with the single-site CG model for the capsid NTDs, in this case the cone formation was more robust and occurred at multiple values of the initial CG interaction strength. The lattice structure of this conical shape was disordered in some places, but regions still existed with p6 lattice symmetry.

Variation of the different CG effective interactions produces closed capsid structures with differing amounts of disorder and symmetry; however, the shape of the capsid remains relatively independent of the lattice structure. Consequently, even though an exact p6 symmetry for the viral capsid lattice is not always reproduced, these CG studies appear to have identified the role of the different net protein-protein interactions responsible for determining the shape and stability of the HIV-1 viral capsid.

DISCUSSION

The CG model calculations presented here imply that HIV-1 retroviral capsids can gain their conical structure through variations in the CTD-NTD packing interface, which break axial symmetry. Furthermore, experimental evidence indicates that the formation of asymmetric, conelike structures is influenced by the electrostatic environment in which capsid self-assembly occurs (19,27). Thus, the formation of the HIV-1 CTD-NTD packing interface may be dominated by polar interactions. As the studies presented here indicate, systematic variability in this interface would be sufficient to explain the formation of a variety of capsid shapes, including the observation of conical capsids, in the in vitro self-assembly experiments. The results also imply that related retroviruses that form symmetric capsids have more hydrophobic CTD-NTD packing interfaces in comparison to the HIV-1 capsid. This prediction is further supported by estimates of the ratio of solvent-exposed surface area (SA) occupied by nonpolar residues in the NTD portion of the CTD-NTD interface relative to the total nonpolar-solvent-exposed SA of the capsid NTDs of HIV-1 and related retroviruses (EIAV, RSV, HTLV-1, and MLV). Upon comparison to the other retroviral capsid NTDs, the HIV-1 capsid NTD domain had the smallest amount of SA covered by hydrophobic residues at the NTD-CTD interface, both in absolute terms and in relation to the total available nonpolar SA of the corresponding NTD. The MLV capsid protein was closest in this property to HIV-1, while the RSV capsid protein had the most hydrophobic surfaces. These predictions have some support from recent structural studies (21,32) of the HIV-1 CA hexamer, showing that the HIV-1 NTD-CTD interface has a strong hydrophilic and polar character. The details of the surface area estimates are provided in the Supporting Material.

The simulation studies presented here show that torsional interactions involving the CG NTD and CTD sites control the degree of curvature (or flatness) of the CG capsid lattice. This raises the intriguing possibility that torsion between capsid CTDs relative to each other along the dimer interface could influence the curvature of the HIV-1 capsid lattice. Stiffer interfaces penalize the rearrangement or assembly of CA proteins in the lattice to form pentameric defect configurations required for lattice curvature. Thus, flexible interfaces can be expected to allow a greater variety of curved lattices to form. Consequently, it may be possible that modulating torsional flexibility through mutations along the dimer or even other binding interfaces would enable the formation of capsid shells of varying sizes and curvature.

Our results further show that the lattice symmetry of the capsid is strongly influenced by the effective shape of capsid NTDs. CG representations that approximate the capsid NTD at different levels of resolution demonstrates a variety of lattice symmetries for the same basic capsid shapes. This result may have significant implications for modeling the self-assembly process of viral capsids. Because approximate representations of individual capsid protein structures will, in general, have some binding conformations differing from those for the actual capsid proteins, we find, when compared to the actual protein self-assembly pathways, that there are different structures into which CG models can combine during self-assembly. This indicates that interpreting the results of protein self-assembly simulations using effective CG models will require careful interpretation, in order to infer biologically relevant information. This work suggests that the CG models developed here could be extended to study the self-assembly kinetics of retroviral capsids. To correctly mimic the packing symmetries of capsid subunits during the self-assembly process, such studies will require the inclusion of greater specificity in CG interactions—possibly through the development of higher-resolution CG models of the individual protein interfaces. As indicated by the results, long-range electrostatic and dynamical polarization interactions are likely to be critical for determination of assembly products, and including these interactions in our presented coarse-graining approach is of considerable value in modeling capsid as well as more general biomolecular self-assembly processes.

CONCLUSIONS

Using a coarse-graining strategy, this work has examined the role of different capsid protein interactions in controlling HIV-1 capsid shape and lattice symmetry. The results suggest that the atomistic and structural factors determining capsid stability are common across different members of the family, whereas variations in overall capsid shape across different retroviruses are caused by variations in the strength of curvature-inducing binding interfaces. It is interesting to note that some of the shapes obtained from different CG simulations in this study resemble capsid shapes in other retroviruses. This may have useful implications for the development of inhibitors that can act simultaneously on

multiple retroviruses, by disrupting common mechanisms of capsid self-assembly and stability.

SUPPORTING MATERIAL

Six figures and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01563-X.

We thank the National Science Foundation Teragrid for use of clusters at the National Center for Supercomputing Applications. We thank Profs. Christopher P. Hill and Wesley I. Sundquist for many useful discussions, and also Prof. Mark Yeager and Dr. Owen Pornillos for the coordinates used in constructing the hexamer structures in Fig. 1. V.K. thanks Drs. Zhiyong Zhang, Luca Larini, and Ian Thorpe for their valuable input.

This work was funded by the National Institutes of Health (grant No. P50-GM0825445).

REFERENCES

1. Garnier, L., J. B. Bowzard, and J. W. Wills. 1998. Recent advances and remaining problems in HIV assembly. *AIDS*. 12:S5–S16.
2. Wang, C. T., and E. Barklis. 1993. Assembly, processing, and infectivity of human immunodeficiency virus type 1 Gag mutants. *J. Virol.* 67:4264–4273.
3. Hunter, E. 1994. Macromolecular interactions in the assembly of HIV and other retroviruses. *Semin. Virol.* 5:71–83.
4. Freed, E. O. 1998. HIV-1 Gag proteins: diverse functions in the virus life cycle. *Virology*. 251:1–15.
5. Li, F., and C. Wild. 2005. HIV-1 assembly and budding as targets for drug discovery. *Curr. Opin. Investig. Drugs*. 6:148–154.
6. Hillis, D. M. 2000. AIDS. Origins of HIV. *Science*. 288:1757–1759.
7. Swanstrom, R., and J. W. Wills. 1997. Synthesis, assembly and processing of viral proteins. *In* Retroviruses. J. M. Coffin, S. H. Hughes, and H. E. Varmus, editors. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. Stremlau, M., M. Perron, …, J. Sodroski. 2006. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5α restriction factor. *Proc. Natl. Acad. Sci. USA*. 103:5514–5519.
9. Tang, C., E. Loeliger, …, M. F. Summers. 2003. Antiviral inhibition of the HIV-1 capsid protein. *J. Mol. Biol.* 327:1013–1020.
10. Li, F., R. Goila-Gaur, …, C. T. Wild. 2003. PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc. Natl. Acad. Sci. USA*. 100:13555–13560.
11. Sticht, J., M. Humbert, …, H. G. Kräusslich. 2005. A peptide inhibitor of HIV-1 assembly in vitro. *Nat. Struct. Mol. Biol.* 12:671–677.
12. Kelly, B. N., S. Kyere, …, C. P. Hill. 2007. Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. *J. Mol. Biol.* 373:355–366.
13. 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.
14. Ganser-Pornillos, B. K., M. Yeager, and W. I. Sundquist. 2008. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* 18:203–217.
15. Fraenkel-Conrat, H., and R. C. Williams. 1955. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl. Acad. Sci. USA*. 41:690–698.
16. Crick, F. H., and J. D. Watson. 1956. Structure of small viruses. *Nature*. 177:473–475.
17. Caspar, D. L., and A. Klug. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27:1–24.
18. Ganser, B. K., S. Li, …, W. I. Sundquist. 1999. Assembly and analysis of conical models for the HIV-1 core. *Science*. 283:80–83.
19. Li, S., C. P. Hill, …, J. T. Finch. 2000. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature*. 407:409–413.
20. Mortuza, G. B., L. F. Haire, …, I. A. Taylor. 2004. High-resolution structure of a retroviral capsid hexameric amino-terminal domain. *Nature*. 431:481–485.
21. 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.
22. Cardone, G., J. G. Purdy, …, A. C. Steven. 2009. Visualization of a missing link in retrovirus capsid assembly. *Nature*. 457:694–698.
23. Nguyen, T. T., R. F. Bruinsma, and W. M. Gelbart. 2005. Elasticity theory and shape transitions of viral shells. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 72:051923.
24. Lidmar, J., L. Mirny, and D. R. Nelson. 2003. Virus shapes and buckling transitions in spherical shells. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 68:051910.
25. Hicks, S. D., and C. L. Henley. 2006. Irreversible growth model for virus capsid assembly. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 74:031912.
26. Nguyen, T. T., R. F. Bruinsma, and W. M. Gelbart. 2006. Continuum theory of retroviral capsids. *Phys. Rev. Lett.* 96:07810.
27. 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.
28. Campbell, S., and V. M. Vogt. 1995. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J. Virol.* 69:6487–6497.
29. 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.
30. Gamble, T. R., S. Yoo, …, C. P. Hill. 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science*. 278:849–853.
31. 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.
32. 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.
33. Karplus, M., and J. N. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules*. 14:325–332.
34. Levy, R. M., M. Karplus, …, D. Perahia. 1984. Evaluation of the configurational entropy for proteins: application to molecular dynamics. *Macromolecules*. 17:1370–1374.
35. Andricioaei, I., and M. Karplus. 2001. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115:6289–6292.
36. Zhang, Z., L. Lu, …, G. A. Voth. 2008. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys. J.* 95:5073–5083.
37. Voth, G. A., editor. 2009. Coarse-Graining of Condensed Phase and Biomolecular Systems. CRC Press/Taylor and Francis Group, Boca Raton, FL.
38. Bates, M. A., M. G. Noro, and D. Frenkel. 2002. Computer simulation of the phase behavior of a model membrane protein. *Annexin V. J. Chem. Phys.* 116:7217–7224.
39. Brooks, B. R., R. Bruccoleri, …, M. Karplus. 1983. CHARMM: a program for macromolecular energy minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
40. Smith, W., and T. R. Forester. 1996. DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph.* 14:136–141.
41. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
42. Lanman, J., T. T. Lam, …, P. E. Prevelige, Jr. 2003. Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. *J. Mol. Biol.* 325:759–772.