Continuum Theory of Retroviral Capsids

T. T. Nguyen,[1,2] R. F. Bruinsma,[1] and W. M. Gelbart[2]

[1]*Department of Physics and Astronomy, University of California at Los Angeles, 475 Portola Plaza, Los Angeles, California 90095, USA*

[2]*Department of Chemistry and Biochemistry, University of California at Los Angeles, 607 Young Drive East, Los Angeles, California 90095, USA*

(Received 10 September 2005; published 21 February 2006)

We present a self-assembly phase diagram for the shape of retroviral capsids, based on continuum elasticity theory. The spontaneous curvature of the capsid proteins drives a weakly first-order transition from spherical to spherocylindrical shapes. The conical capsid shape which characterizes the HIV-1 retrovirus is never stable under unconstrained energy minimization. Only under conditions of fixed volume and/or fixed spanning length can the conical shape be a minimum energy structure. Our results indicate that, unlike the capsids of small viruses, retrovirus capsids are not uniquely determined by the molecular structure of the constituent proteins but depend in an essential way on physical constraints present during assembly.

DOI: 10.1103/PhysRevLett.96.078102 PACS numbers: 87.15.Nn, 61.50.Ah, 81.16.Dn, 87.16.Dg

The life cycle of an HIV-1 retrovirus begins when RNA viral genome molecules inside an infected host cell are enclosed by an *immature*, roughly spherical, shell of ''Gag'' proteins [1]. The ''nucleocapsid'' (NC) portion of this Gag protein binds to the RNA molecules, driving genome packaging, while the ''matrix'' (MA) region binds to the host cell membrane, leading to envelopment of the protein shell by a lipid bilayer, followed by escape (''budding'') of the virus from the cell. Upon budding, the Gag protein undergoes self-cleavage into three separate portions —NC, CA, and MA—resulting in disassembly of the immature capsid and release of the middle ''capsid protein'' (CA) portion. The CA proteins self-assemble into the *mature* capsid that contains two copies of the viral genome, along with some cellular RNA and one copy of the reverse transcriptase protein [2]. A retroviral capsid—still enveloped by the lipid bilayer—has a typical size of about 100 nm. Most retroviral genera have spherical capsids, type *D* and beta retroviruses are spherocylindrical, while the lentiviruses (e.g., HIV-1) are *conical* [3].

Electron microscopy of single-molecule-thick sheets of CA proteins reveals a basic structural motif that consists of hexameric CA rings, with a diameter of about 10 nm and local p6 symmetry, hooked together by dimerized *C*-terminal domains belonging to adjacent hexamers [4]. Spherical, spherocylindrical, and conical retroviral capsids all share this organizational principle. The different shapes are distinguished [5] by the distribution of the 12 protein *pentamers*, which—according to Euler's theorem—have to be inserted into the hexagonal lattice for it to be able to form a closed shell. For spherical shells, these 12 pentamers must be equidistant; for spherocylindrical shells, each of the two caps contains six pentamers, while for conical shells the larger cap must contain seven (or more) pentamers and the smaller cap five (or fewer).

The three canonical shapes can all be reproduced under *in vitro* conditions by direct self-assembly of purified HIV-1 CA protein and its mutants [3], hence necessarily without the enclosing membrane. This would appear to indicate that the three capsid shapes represent competing free energy minima of CA protein aggregates, with the molecular structure of the capsid protein determining capsid shape, as is indeed the case for many self-assembling smaller viruses. *In vivo* studies indicate that, unlike capsids of small viruses, HIV-1 capsids have a broad distribution of sizes or shapes [6].

Establishing a connection between variants of the CA protein structure and capsid shape by an all-atom molecular-dynamic simulation would be a forbidding task in view of the enormous size of the capsids (about 1500 CA proteins) and incomplete information regarding the detailed subunit interactions. In this Letter, we apply continuum elasticity theory to thermodynamically determine a shape phase diagram for retroviruses. (Alternatively, interpretation of spherical capsid assembly also has been based on consideration of growth kinetics [7]. However, no such interpretation for conical capsids has been proposed.) It was shown by Lidmar, Mirny, and Nelson (LMN) [8] that the continuum elasticity theory for hexagonal shells can account for global features of larger icosahedral viral capsids, such as the ''buckling transition'' by which faceting arises in capsids upon increase in their size and/or their ratio of stretching to bending moduli. We have extended this approach to treat *nonicosahedral* shapes, specifically the spherocylinder and cone, by including the possibility of a nonzero spontaneous curvature for the 2D lattice of capsid protein [9]. The various sizes and shapes of viral capsids could then, in principle, be understood by relating the phenomenological moduli of continuum theory to the structural variants of the CA protein.

Continuum elasticity theory assigns an energy $\mathcal{H}$ to a hexagonal sheet that is the sum of an "in-plane" stretching energy $\mathcal{H}_s$ and an "out-of-plane" bending energy $\mathcal{H}_B$. We evaluated $\mathcal{H}$ numerically by approximating the capsid shape as a closed triangular net containing 12 sites of fivefold coordinations. The in-plane elastic energy $\mathcal{H}_s$ is given as the pairwise sum of harmonic interaction potentials between nearest neighbors nodes i and j of the net:

$$\mathcal{H}_s = \frac{\varepsilon}{2}\sum_{ij}(|\vec{r}_i - \vec{r}_j| - a)^2. \quad (1)$$

Here a is the equilibrium spacing of the harmonic potential and ε the spring constant. This equilibrium spacing should *not* be viewed as the spacing between CA proteins but merely as a short distance cutoff for the numerical computation of the continuum elastic energy. The spring constant is related to the (2D) *Young's modulus* of the continuum theory by $Y = 2\varepsilon/\sqrt{3}$ [LMN]. The out-of-plane bending energy $\mathcal{H}_B$ is given as a pairwise interaction between the normal directions $\hat{n}_I$ of neighboring triangles of the net:

$$\mathcal{H}_B = k\sum_{IJ}[1 - \cos(\theta_{IJ} - \theta_0)]. \quad (2)$$

Here θ_{IJ} is the dihedral angle between the normal directions $\hat{n}_I$ and $\hat{n}_J$ of two adjacent triangles I and J. The energy scale k is related to the bending or *Helfrich modulus* of continuum theory by $\kappa = \sqrt{3}k/2$ [LMN]. By comparing measured and computed icosahedral capsid shapes, LMN found that $Y/\kappa \simeq$ nm^{-2}, while nanoindentation studies [10] of viral capsids yield values for κ in the range of 10–100k_BT. The preferred dihedral angle θ_0 is related to the *spontaneous curvature* of continuum theory by $C_0 = 2\theta_0/\sqrt{3}a$; nonzero C_0 is formally associated with the absence of an "up-down" mirror symmetry of the CA proteins [11] to the lowest order. Because capsid proteins arrange in a hexagonal lattice, to the lowest order, we can assume the capsid shell to be effectively isotropic in the in-plane directions [12].

The starting shapes of the simulation studies were determined by *isometric construction*, i.e., shapes constructed by folding *inextensible* templates cut from a hexagonal lattice into spherical, spherocylindrical, or conical shells. The isometric construction provides the standard Caspar-Klug (CK) classification of icosahedral viruses [13] (Fig. 1) and can be extended to spherocylinders and cones [9] (Fig. 2). As shown in Fig. 2, a right isometric cylindrical shell can be classified by four positive integers m, n, h, and k with the m/n ratio proportional to the aspect (length-to-width) ratio. The number of nodes on such a shell is $10mn(h^2 + k^2 + hk) + 2$. A right 5–7 isometric cone is also classified by four integers m, n, h, and k, now with m/n proportional to the ratio of the diameters of the two caps. The number of nodes on such a shell is $10(2m^2 - n^2)(h^2 + k^2 + hk) + 2$. By choosing a large enough number of nodes on the triangular net, one can simulate the corresponding spherocylinder and conical continuum shells with high accuracy. Earlier studies [8,9] show that the continuum limit is reached when the number of nodes exceeds 10 000. In this limit, the shell elastic energy does not depend on the specific values for $(m, n|h, k)$ but depends only on two dimensionless parameters: the *Föppl–von Kármán* (FvK) number $\gamma = YS/\kappa$, specifying the ratio of stretching and bending energies, and the dimensionless spontaneous curvature $\alpha = 2\theta_0 S^{1/2}/\sqrt{3}a$. In our simulation, for simplicity, we fix $h = 1$ and $k = 0$. For the reference icosahedral shell, we choose $m = n = 55$ corresponding to a triangular net of 30 252 nodes.

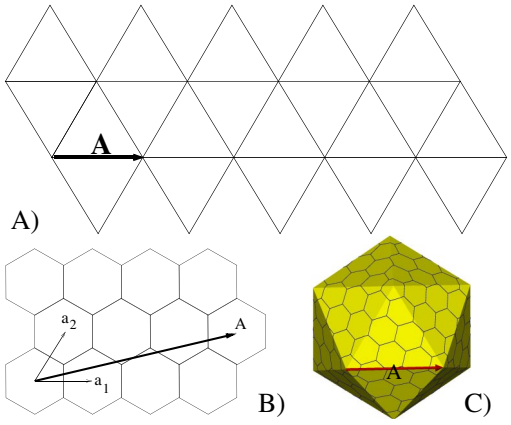

FIG. 1 (color online). CK isometric construction of a spherical shell. (a) Folding template for an icosahedron. The template is indexed by the lattice vector $\vec{A} = h\vec{a}_1 + k\vec{a}_2$ of a hexagonal lattice with basis vectors $\vec{a}_1$ and $\vec{a}_2$. (b) The case $h = 3$ and $k = 1$. (c) The corresponding icosahedron.

Using these isometric shell shapes, we minimized the elastic energy with respect to shell shape by the conjugate gradient method [14]. The capsid surface area S was maintained fixed, but the enclosed volume was allowed to vary freely since empty viral capsids produced by *in vitro* self-assembly are not under osmotic pressure. For

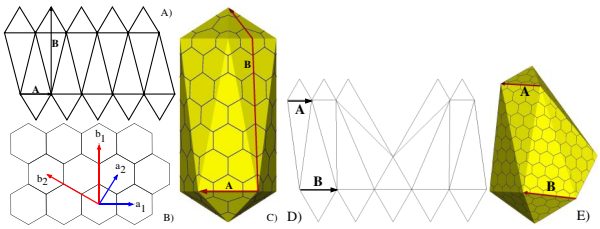

FIG. 2 (color online). Nonicosahedral isometric shells. (a) Folding template of an isometric spherocylinder. The template is indexed by two perpendicular vectors $\vec{A} = n(h\vec{a}_1 + k\vec{a}_2)$ and $\vec{B} = m(h\vec{b}_1 + k\vec{b}_2)$, with $m > n$ two positive integers. (b) The four vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{b}_1$, and $\vec{b}_2$. (c) Isometric spherocylinder with $(m, n, h, k) \equiv (4, 2, 1, 0)$. (d) Folding template of an isometric 5–7 cone. The template is indexed by two parallel vectors $\vec{A} = n(h\vec{a}_1 + k\vec{a}_2)$ and $\vec{B} = m(h\vec{a}_1 + k\vec{a}_2)$. (e) Isometric 5–7 cone with $(m, n, h, k) \equiv (4, 3, 1, 0)$.

a given shape family, conical or spherocylinderical, we calculate the capsid energy for a different aspect ratio m/n and choose the shape with the optimum aspect ratio whose energy is lowest. For a given set of parameters (α, γ), we compare the optimum cone with that of the optimum spherocylinder and the reference sphere to determine the thermodynamically most stable shape. The resulting shape phasediagram is shown in Fig. 3.

A first-order transition line separates icosahedral from spherocylindrical shells while a *buckling threshold* at $\gamma_B \simeq 3000$ separates (approximately) spherical from (approximately) polyhedral shells. The FvK number of a retrovirus such as HIV-1 is of the order of 2×10^4 for $Y/\kappa \simeq$ nm^{-2} (see Fig. 3), significantly above the buckling threshold. Microscopy studies of conical retroviral capsids reveal that they are indeed distinctly polyhedral [6,11,15]. Interestingly, for FvK numbers in this range, the shape transition is either weakly first-order or continuous. A more detailed study of the sphere-spherocylinder shape transition can be found in our previous work (Ref. [9]).

The shape phase diagram Fig. 3 would account for the many retrovirus genera with spherical and spherocylindrical shapes, if we assume that the spontaneous curvature parameter α for the various CA proteins lies in the range of 3–7. The near continuous nature of the transition also would be consistent with the polymorphism observed in CA self-assembly. A baffling prediction of the continuum theory is, however, that lentivirus conical shells never constitute a minimum of the elastic energy. For FvK numbers of the order of 10^4, the elastic energy of a conical shell is about 5κ higher than that of a spherocylindrical shell, which would mean that, under conditions of thermodynamic self-assembly, the ratio of cones to spherocylinder populations would be very small.

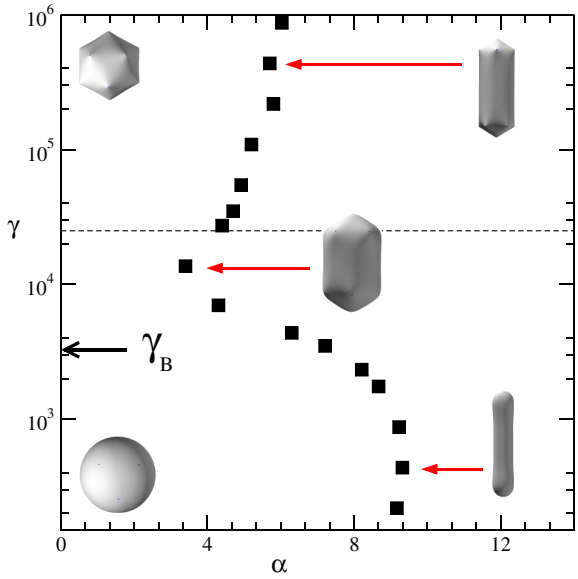

FIG. 3 (color online). Shape phase diagram at fixed area S. The vertical axis is the FvK number $\gamma = YS/\kappa$, and the horizontal axis the dimensionless spontaneous curvature $\alpha = 2\theta_0 S^{1/2}/\sqrt{3}a$. The buckling threshold $\gamma_B \simeq 3000$ separates spherical from polyhedral shells. The black squares show the phase boundary obtained from numerical simulation. On the right side of this boundary, the spherocylinder shape has the lowest energy. Three such shapes close to the transition line are shown for small, intermediate, and large γ (all shapes have the same area but they are scaled to fit the allotted space). The FvK number of HIV-1 capsids is of the order of 2×10^4 (dotted line).

Self-assembly under *in vitro* conditions of HIV-1 "wild-type" CA proteins indeed produces long cylindrical shells with only a small fraction of cones. The cone-to-cylinder ratio is, however, greatly enhanced (to about 2/3) when the CA proteins are fused to the NC (i.e., RNA binding) part of the Gag protein with assembly taking place in the presence of viral RNA molecules [15]. In this case, RNA genome molecules are trapped inside the capsid, which places a *volume constraint* on the shell. Finally, under *in vivo* conditions, assembly produces a majority fraction of cones and a minority fraction of spherocylinders. In that case, a second constraint is operative: Electron micrographs [6] of *in vivo* assembled capsids indicate that the *spanning length* of both capsid types is limited by the diameter of the surrounding spherical membrane. These two constraints are not independent: Spherocylinders and cones of the same area and volume also happen to have the same spanning length to lowest order in the cone aperture angle θ.

Figure 4 shows a shape phase diagram under fixed volume conditions. The FvK number was chosen to be

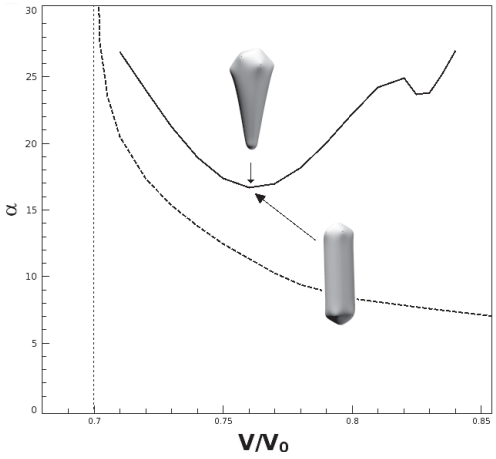

FIG. 4. Transition from spherocylinder to cone at fixed area S and fixed volume V. The horizontal axis is the ratio of the capsid volume to that of a reference icosahedral shell V_o. The vertical axis is the dimensionless spontaneous curvature. The FvK number is fixed at 25 000. At the indicated transition point, the m/n ratio equals 98/31 for the spherocylinder and 40/13 for the cone; for larger (smaller) values of V/V_o, the m/n ratio at the transition is smaller (larger). The dotted line shows the transition line predicted by the LMN analytical theory.

25 000. The vertical axis is again the dimensionless spontaneous curvature α, while the horizontal axis is the ratio V/V_0 of the shell volume to that of the reference icosahedral shell. For each shape (spherocylinder and cone), we determined the m/n value corresponding to a given V/V_0 and calculated the energy as a function of α.

With increasing α, we now encounter a (discontinuous) transition from the spherocylinder shape to the cone shape for reduced volumes between 0.70 and 0.85. The maximum near $V/V_o = 0.84$ appears be related to a buckling transition of the smaller endcap. LMN proposed an approximate analytical theory for the elastic energy of icosahedral shells by adding the elastic energy of 12 disclination ''defects'' to a ''background'' curvature energy. Their theory can be adapted to the present case, and it indeed predicts a transition from the spherocylinder to the cone shape at fixed volume, although at somewhat lower values of the spontaneous curvature (see dotted line in Fig. 4). The predicted and calculated m/n ratios agree very well on the other hand.

The driving mechanism of the transition is, according to the theory, the reduction of the curvature energy cost of equal-sized end caps by breakup into smaller and larger sizes. A rather similar instability is encountered in the phase diagram of closed lipid bilayers with fixed vesicle area and enclosed volume (but zero FvK number) from a prolate ellipsoidal shape to a pearlike shape [16]. This instability is a precursor of the well-known budding transition of lipid vesicles [17].

Comparison of minimum energy shell shapes calculated along the transition line of Fig. 4 with those observed for HIV-1 capsids is illuminating. Micrographs of HIV-1 cores resemble isometric cones with aperture angles θ close to that of a 5–7 cone [$\arcsin(1/6) \approx 19^\circ$] [6]. For reduced volumes between 0.75 to 0.85, calculated m/n ratios along the transition line ranged from 3.16 to 2.27 for the spherocylinder and from 3.07 to 1.68 for the cone. For conical HIV-1 capsids [6], the measured m/n range is from 1.5 to 2.0, and, for spherocylindrical capsids, the m/n ratio is about 2. Next, the dimensionless spontaneous curvature is of the order of 20 near the transition line—where HIV-1 capsids are most likely located—which would correspond to a preferred radius (reciprocal spontaneous curvature) of about 7 nm and, thereby, imply the radius of cylindrical capsids. Actual CA cylinders, produced by self-assembly, have a radius of about 20 nm [4].

The central result of this Letter is that unconstrained energy minimization predicts a weakly first-order transition from spherical to spherocylindrical capsid shapes. It does not yield any conditions under which the conical shape of the lentiviruses constitute an energy minimum. Rather, conical capsids are stabilized by imposing *assembly constraints* on the capsid. Assembly constraints arise physically from the size of the packaged genome and of the enveloping membrane. Our finding indicates that the molecular structure of the capsid proteins only partially ''codes'' for the capsid shape. Unlike the precisely patterned icosahedral capsids of small viruses, the CA hexameric ring architecture provides retroviruses with a flexible basic unit that can be assembled into different structures depending on constraints. To test this concept as directly as possible, one could carry out *in vitro* self-assembly experiments involving CA/NC fusion proteins in which mutations leading to larger spontaneous curvature (i.e., smaller radii for the cylinders) should correlate with enhancement of the population fraction of conical shells.

We thank B. Ganser-Pornillos and D. Nelson for helpful discussions, and we also acknowledge support by the NSF under DMR Grant No. 0404507 and CHE No. 0400363.

[1] H. R. Gelderblom, P. G. Bauer, M. Özel, P. Höglund, M. Niedrig, H. Renz, B. Morath, P. Lundquist, Å. Nilsson, J. Mattow, C. Grund, and G. Pauli, in *Membrane Interactions of HIV*, edited by R. C. Aloia and C. C. Curtain (Wiley-Liss, New York, 1992), pp. 33–54.
[2] T. Wilk, I. Gross, B. E. Gowen, T. Rutten, F. de Haas, R. Welker, H. G. Krausslich, P. Boulanger, and S. D. Fuller, J. Virol. **75**, 759 (2001).
[3] B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, and W. I. Sundquist, J. Virol. **78**, 2545 (2004).
[4] S. Li, C. P. Hill, W. I. Sunquist, and J. T. Finch, Nature (London) **407**, 409 (2000).
[5] M. Ge and K. Sattler, Chem. Phys. Lett. **220**, 192 (1994).
[6] J. Benjamin, B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist, and G. J. Jensen, J. Mol. Biol. **346**, 577 (2005).
[7] D. Endres and A. Zlotnick, Biophys. J. **83**, 1217 (2002).
[8] J. Lidmar, L. Mirny, and D. R. Nelson, Phys. Rev. E **68**, 051910 (2003).
[9] T. T. Nguyen, R. F. Bruinsma, and W. M. Gelbart, Phys. Rev. E **72**, 051923 (2005).
[10] I. Ivanovska, P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. L. Wuite, Proc. Natl. Acad. Sci. U.S.A. **101**, 7600 (2004); J. P. Michel, I. L. Ivanovska, M. M. Gibbons, W. S. Klug, C. M. Knobler, C. F. Schmidt, and G. J. L. Wuite (to be published).
[11] T. R. Gamble, S. Yoo, F. F. Vajdos, U. K. von Schwedler, D. K. Worthylake, H. Wang, J. P. McCutcheon, W. I. Sundquist, and C. P. Hill, Science **278**, 849 (1997).
[12] In-plane anisotropy effects will be important for smaller viral shells with radii comparable to the capsomer size.
[13] D. L. D. Caspar and A. Klug, Cold Spring Harbor Symposia on Quantitative Biology **27**, 1 (1962).
[14] H. S. Seung and D. R. Nelson, Phys. Rev. A **38**, 1005 (1988).
[15] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, Science **283**, 80 (1999).
[16] U. Seifert, K. Berndl, and R. Lipowsky, Phys. Rev. A **44**, 1182 (1991).
[17] H. G. Dobereiner, J. Kas, D. Noppl, I. Sprenger, and E. Sackmann, Biophys. J. **65**, 1396 (1993).