Optimal architectures of elongated viruses

Antoni Luque[a], Roya Zandi[b], and David Reguera[a,1]

[a]Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain; and [b]Department of Physics and Astronomy, University of California, Riverside, CA 92521

Communicated by Howard Reiss, University of California, Los Angeles, CA, January 25, 2010 (received for review June 19, 2009)

Many viruses protect their genetic material by a closed elongated protein shell. Unlike spherical viruses, the structure of these prolates is not yet well understood, and only a few of them have been fully characterized. We present the results of a simple phenomenological model, which describes the remarkable structures of prolate or bacilliform viral shells. Surprisingly, we find that the special well-defined geometry of these elongated viruses arises just as a consequence of free-energy minimization of a generic interaction between the structural units of the capsid. Hemispherical *T*-number caps centered along the 5-, 3-, and 2-fold axes with hexagonally ordered cylindrical bodies are found to be local energy minima, thus justifying their occurrence as optimal viral structures. Moreover, closed elongated viruses show a sequence of magic numbers for the end-caps, leading to strict selection rules for the length and structure of the body as well as for the number of capsomers and proteins of the capsid. The model reproduces the architecture of spherical and bacilliform viruses, both in vivo and in vitro, and constitutes an important step towards understanding viral assembly and its potential control for biological and nanotechnological applications.

assembly | polymorphism | prolate or bacilliform shells | simulation | viral capsids

Viruses are nonliving particles that infect a wide variety of organisms, from bacteria to mammals. In the course of the last century, viral studies helped the development of different scientific fields ranging from molecular and cellular biology to medicine (1). The well-defined size in the nanometer range, highly symmetrical shape, and ability to self-assemble spontaneously have led to renewed interest in understanding virus structure in the nanosciences, where several technological and biomedical applications are envisioned (2–4).

In their simplest form, viruses are composed of an infective genetic material (DNA/RNA) enclosed by a protein shell called the *capsid*, which is typically built from several copies of one or a few different proteins. In the assembled structures, capsid proteins are often clustered into multimers called *capsomers* of five (*pentamer*) or six (*hexamer*) proteins. Many viruses self-assemble and can be reconstituted in vitro from their molecular components (5–8), suggesting that the well-defined geometrical architectures of viral shells should be a consequence of free-energy minimization.

About half the viral species adopt roughly spherical structures with icosahedral symmetry, characterized by the triangulation number T (9), a structural index for viral shells. Recently, significant progress has been made in understanding the physical and geometrical principles governing the architecture of in vivo- and in vitro-reconstituted quasispherical viruses (10–16). Structures of a large number of spherical viruses have been examined in detail for several years (17)*, but the precise geometry and architecture of many nonspherical viruses, in particular those with prolate shapes, have not yet been fully understood. In this work, we focus on capped cylindrical viral capsids, which in the literature are indistinguishably referred to as prolate, bacilliform, elongated, tubular, or allantoid (18), excluding rod-like and other noncapped structures such as that of tobacco mosaic virus. Some examples of such elongated viruses are the genera ascovirus, badnavirus, barnavirus, and c2-like viruses, which infect animals, plants, fungi, and bacteria, respectively (18). Beyond their biological and pathological interest, prolate viruses are very promising candidates in nanoscience and nanotechnology as biotemplates for the fabrication of nanowires and for the selective encapsulation of materials.

In general, the few prolate viruses that have been characterized experimentally contain a cylindrical body, composed of hexamers, closed with two hemispherical caps with icosahedral symmetry centered on a 5-fold axis (19, 20). Remarkably, many in vitro experiments reveal that the body of bacilliform viruses is capable of adopting different discrete lengths (21–25), which should be intimately related to the particular architecture of the capsid. To explain the experimental observations pertinent to bacilliform capsids, Moody (25, 26) extended the ideas of Caspar and Klug (9) (CK) and presented a geometrical scheme (CKM) for the formation of elongated icosahedrally capped capsids with 5-fold axial symmetry. Nevertheless, there are several viruses that cannot be classified according to CKM model, for example, the alfalfa mosaic virus (AMV) (21, 22, 27), rice tungro virus (23), or aberrant flock house virus (FHV) (24).

Recently, Chen and Glotzer studied the optimal packing of identical hard spheres with square-well interactions on a prolate spheroid surface with convexity constraints (28). They showed that the elongated optimal structures, composed of 15, 17, 18, and 42 identical units, resemble the hypothetical structures of aberrant FHV (24) and of native $\phi 29$ (19), even though these viruses have two types of morphological units rather than one. Further, Nguyen and Brooks (29) studied the polymorphism observed in the self-assembly of viral capsids, using a simplified model for capsomers. Besides spherical T-number capsids, they obtained some elongated structures including that of the maize streak virus (MSV) (30) as well as some nonicosahedral prolates with structures similar to those of FHV and $\phi 29$.

However, the range of possible structures, the validity conditions for the Moody construction, and the physical principles justifying the appearance of prolate architectures have yet to be fully elucidated. We find that, as in the spherical case, elongated viruses adopt these precise geometries because they are free-energy minima of a very generic interaction between the capsomers. We show that the range of structures which are energetically optimal as well as the rules that control their length, helicity, and aspect ratio depend strongly on the end-caps morphology and the way the cylindrical body fits into them. In addition, our model is able to reproduce the structure of some less-explored elongated viruses. Our results can be useful for inferring viral structures using simple data from micrographs and sedimentation experiments, as well as for predicting the sizes and number of proteins of elongated capsids. Moreover, they can

Author contributions: R.Z. and D.R. designed research; A.L. performed research; A.L., R.Z., and D.R. analyzed data; and A.L., R.Z., and D.R. wrote the paper.

The authors declare no conflict of interest.

*See also the structures listed in http://viperdb.scripps.edu/.

[1]To whom correspondence should be addressed. E-mail: dreguera@ub.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0915122107/DCSupplemental.

also guide the design of artificial viruses for nanotechnological applications.

Models and Methods

Geometrical Construction of Prolate Capsids. The structural classification of spherical viruses (17) is based on the *quasi-equivalence principle* introduced by Caspar and Klug (9), which led to a geometrical model for the construction of *icosahedral shells* (*SI Text*). The main index to classify the possible icosahedral capsids is the triangulation number $T(h,k) = h^2 + hk + k^2$, with h and k nonnegative integers. This corresponds to the series $T = 1, 3, 4, 7, 9, \ldots$ and determines the number of different environments that the protein subunits have in the shell. A T-number structure has $60 \times T$ subunits and is composed of 12 pentamers and $10(T-1)$ hexamers, adding up to $10T + 2$ capsomers. Moreover, different T-structures can be grouped together in terms of *classes* that share common geometrical properties. Defining $h \equiv fh_0$ and $k \equiv fk_0$ with f the highest common factor of h and k, we can then rewrite $T(f, h_0, k_0) = Pf^2$ where the *icosahedral class*, $P(h_0, k_0) = h_0^2 + h_0k_0 + k_0^2$ (9), is a subset of T ($P = 1, 3, 7, 13, \ldots$). For instance, $T = 4$ and $T = 9$ belong to the class $P = 1$ with $f = 2$ and $f = 3$, respectively.

About a half century ago, Moody (25, 26) extended CK's concepts and presented a geometrical model (CKM), which nicely describes the structure of *prolate* capsids with 5-fold axial symmetry (see *SI Text*). According to this model, an elongated polyhedral viral capsid can be constructed from an icosahedron by increasing the height of its 10 midtriangles. Each cap of the resulting prolate is made of five triangular faces defined by the same triangulation number $T_{\text{end}}(h,k) = Pf^2$ as in the CK model. The surface of a midtriangle defines a new triangulation number, $T_{\text{mid}}(h,k,h',k') = hh' + hk' + kk' = Qf$, with $Q = h_0h' + h_0k' + k_0k'$ controlling the length of the prolate, where h' and k' are nonnegative integers. The total number of proteins in the capsid is $N_s = 30(T_{\text{end}} + T_{\text{mid}}) = 30f(Pf + Q)$, distributed in $N = 5(T_{\text{end}} + T_{\text{mid}}) + 2 = 5f(Pf + Q) + 2$ capsomers. Remarkably, Moody showed that for a fixed T_{end}, the body of a prolate with 5-fold axial symmetry can grow only in steps of $\Delta N = 5f$ hexamers, corresponding to a minimum step in Q of $\Delta Q_{\text{min}} = 1$.

Minimal Model of Prolate Capsids. Capsid proteins of distinct viruses can differ in terms of the sequence, size, and conformations of their amino acids. However, in their final structures, capsid proteins are invariably clustered into morphological units, e.g., pentamers and/or hexamers, which are very similar in all viruses. Thus, while the interaction between proteins can be very complex and species-specific, capsomers are expected to interact through a more generic and isotropic potential.

In ref. 11, we introduced a minimal model, which was able to capture the essential ingredients of the interaction between capsomers: a short-range repulsion, representing subunit conformational rigidity (to prevent protein overlapping), plus a longer-range attraction, corresponding to the driving force for capsomer aggregation. One of the simplest models fulfilling these requirements is the well-known Lennard-Jones-like potential. Indeed, we assume an effective capsomer-capsomer interaction potential described by

$$V(r) = \varepsilon_0 \left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6} \right], \qquad [1]$$

where r is the separation between capsomer centers, σ is their optimal distance, and ε_0 is the binding energy between them. The capsomer-capsomer binding energy ε_0 is taken to be $15k_BT$ (where k_B is the Boltzmann constant and T is the absolute temperature), a typical value reported from atomistic calculations of subunit binding energies (31).

Another essential feature in viral capsids is that we can find two different morphological units, which are aggregates of five (pentamers) or six (hexamers) proteins, that in general may have different "internal energy" arising from differences in folding, stretching, and contact interactions between the protein subunits. Therefore, in our model we define two types of capsomers, pentamers and hexamers, which have an energy difference ΔE. Accordingly, the Boltzmann factor $e^{-\Delta E/k_BT}$ accounts for the relative probability for a noninteracting unit to be a pentamer. Moreover, pentamers and hexamers are modeled as spherical capsomers with different sizes, and the ratio between their radii is taken to be the same as the ratio between the radii of the circles inscribed in a pentagon and hexagon of the same edge, respecting the fact that pentamers and hexamers are made of the same protein. Therefore, the ratio between the radius of an hexamer, σ_0^h, and that of a pentamer, σ_0^p, is given by $\tan(\pi/6)/\tan(\pi/5)$. The values of the parameters of the interaction between the different capsomers are taken to be $\sigma_{h-h} = 2\sigma_0^h$, $\sigma_{p-p} = 2\sigma_0^p$, and $\sigma_{h-p} = (\sigma_0^h + \sigma_0^p)$ for the equilibrium distances between hexamer-hexamer, pentamer-pentamer, and hexamer-pentamer, respectively. The binding energy was assumed to be the same ($\varepsilon_0 = 15k_BT$) in all cases. All simulation results will be reported in reduced units, i.e., in units of ε_0 for the energy and σ_0^h for lengths.

We note that with these ingredients the model in ref. 11 was able to successfully explain the appearance of icosahedral shells as well as other structures in spherical capsids observed both in vivo and in vitro.

In order to obtain the optimal structures of prolate capsids, corresponding to the free-energy minima, we assumed that our model capsomers are constrained to live on the surface of a spherocylinder and carried out a set of Monte Carlo simulations with a standard Metropolis scheme.[†]

We consider that N capsomers are placed on the surface of a spherocylinder composed of a cylindrical body of length L and radius R closed by two semispherical caps of also radius R. Initially, all N capsomers are hexamers and are placed at random positions on the spherocylindrical surface. In order to find the free-energy minimum structure, for each N at a fixed value of R, three different types of Monte Carlo moves are attempted: (i) a move to change the position of the capsomer on the surface of the spherocylinder; (ii) a move to change the capsomer size (switching between hexamers and pentamers); and (iii) a move to adjust the length L of the cylindrical body. We emphasize that this Monte Carlo protocol can explore the energy minima but it is intended to describe neither the actual kinetics nor the pathways involved in the physical process of assembly. Different relative ratios between the different types of moves were tested in order to facilitate the reaching of the optimal structure. Typically, the combination that worked best was 200:5:1 (position:size:length). The system was then allowed to equilibrate (typically for 25,000 to 125,000 MC steps), and both the average and the absolute minimum values of the free-energy after equilibration were recorded.

Results and Discussion

Two Different Morphological Units. Here, we first analyze the optimal structures for capsids composed of two different morphological units (hexamers and pentamers) having the same internal energy, i.e., $\Delta E = 0$. Later we consider the case in which $|\Delta E/k_BT| \gg 1$, corresponding to viral capsids made only by one type of morphological unit. We simulated prolate capsids with a total number of capsomers between $12 < N < 100$, and for each N we obtained the optimal values of the length and free energy per capsomer, ε, as a function of the radius R. The solid

[†]It is important to note that for large values of N, nonstandard techniques as parallel tempering or simulated annealing would be necessary to minimize trapping in local minima.

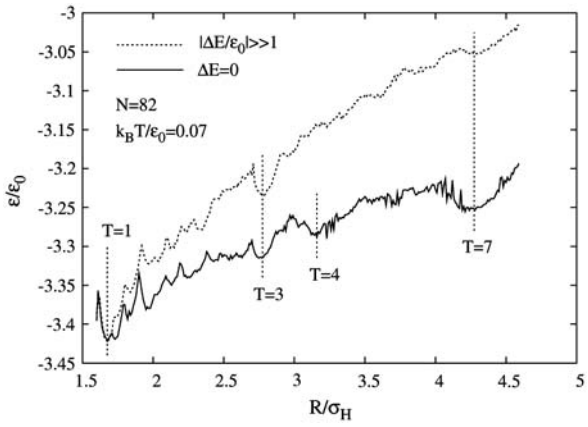

Fig. 1. Spectrum of the free energy per capsomer ε versus the radius R for a fixed number of capsomers $N = 82$. The solid line corresponds to the results for the simulation with two types of capsomers ($\Delta E/k_BT = 0$), and the dashed line is for one type of capsomer ($|\Delta E/k_BT| \gg 1$).

curve of Fig. 1 shows a particular but representative example of the spectrum $\varepsilon(R)$ for $\Delta E = 0$.

The results indicate that capsids with small radii, and thus long cylindrical bodies, are energetically favorable in the absence of bending penalties. This result is not unexpected because capsomers in the cylindrical body can be easily arranged in the locally optimal configurations, i.e., a regular hexagonal pattern with all capsomers having six nearest neighbors located at an equilibrium distance from each other. However, the capsomers in the spherical caps naturally have higher energies on average due to the existence of pentamers, which in many cases impose a significant amount of stress on the capsid (32). Hence, for a small radius, the number of capsomers on the cylindrical body increases, leading to lower total energies on average. Note that the final structure of a capsid does not necessarily correspond to the global energy minimum structure for a given N but rather corresponds to one of the different *local minima*, which end up being selected once the bending energy, spontaneous curvature, size of the encapsidated genome (33, 34) or even scaffolding mechanisms (35) are taken into account.

Fig. 1 illustrates the presence of several deep *local* minima in the energy per capsomer at particular values of the radius. These minima appear at very precise values of the radii and, more importantly, appear consistently in the free-energy landscapes of capsids with different values of N. These results suggest the existence of a discrete set of special "magic" values for the radius of the structures, which are particularly optimal from the energetic point of view.

Table 1. Optimal sequences with no energy difference between hexamers and pentamers ($\Delta E = 0$)

⟨R⟩*	Cap†	Body‡	N^0_{seq}§	ΔN¶	⟨ΔL⟩‖
1.668	1 (5-fold)	R_5	12	5	1.68
1.673	1 (3-fold)	Z_6	12	3	0.97
1.674	1 (2-fold)	skew	12	2	0.61
2.761	3 (5-fold)	Z_{10}	32	5	1.00
2.875	3 (3-fold)	R_9	32	9	1.72
3.205	4 (5-fold)	R_{10}	42	10	1.71
4.243	7 (5-fold)	skew	72	5	0.70

*Average radius of the sequence.
†Structure of the hemispherical caps. For the icosahedral ones: T_{end} (axial symmetry).
‡Pattern of hexamers in the cylindrical part.
§Number of capsomers for the spherical seed of the sequence.
¶Number of hexamers added to the body for consecutive particles in a sequence.
‖Length step obtained from the simulation.

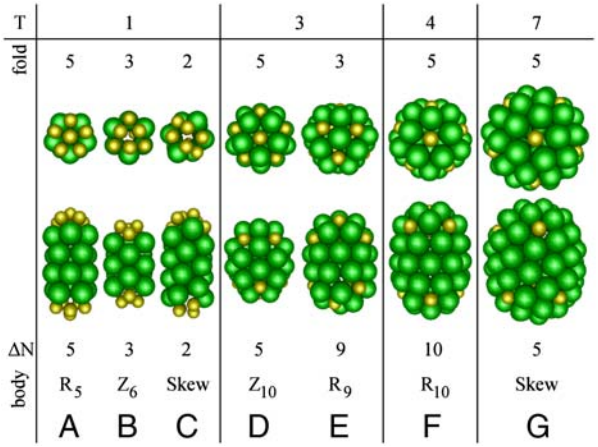

Fig. 2. Gallery of optimal prolate capsids based on icosahedral caps for $\Delta E = 0$, where hexamers are colored in green, and pentamers are colored in gold. The top images show a zenithal view of the cap, whereas the lower images show a lateral view of a representative example of the complete bacilliform. For each cap, the arrangement of hexamers in the body as well as the minimum step in the number of capsomers ΔN are also indicated.

Table 1 lists the values of the radii that correspond to the deepest local free-energy minima that appear consistently for different values of N. The careful analysis of the corresponding optimal structures at these special values of the radii indicate that twelve pentamers have always been distributed symmetrically on the caps. Note that the energetically favorable structures correspond to having T-numbered caps with cylindrical bodies made of a rolled-up hexagonal sheet of hexamers. In particular, we obtain well-defined local free-energy minima for the radii that correspond to semispherical $T_{end} = 1$, $T_{end} = 3$, $T_{end} = 4$, and $T_{end} = 7$ caps centered on a 5-fold axis, coinciding to the lowest triangulation numbers in the CKM classification. Thus, we recover the structure of ϕ29 (19), which is a $T_{end} = 3$ (5-fold) $Q = 5$ (Fig. 2*D*). We also find well-defined minima for radii corresponding to $T_{end} = 1$ and $T_{end} = 3$ with caps centered on a 3-fold axis. These two architectures do not follow the Moody's geometrical construction rules for bacilliform capsids (26) but have been suggested to appear in some plant viruses, such as the elongated AMV (21, 27) and the rice tungro bacilliform virus (RTBV) (23), respectively. In our simulations, prolates with icosahedral caps centered on a 2-fold axis were also observed but only for $T_{end} = 1$ capsids. These structures are unusually distorted, and apparently unfavorable for larger caps. We have also found other less pronounced free-energy minima associated with tubular structures made of rings or zigzag layers of hexamers closed by nonicosahedral caps, which were associated with the shallow minima in the spherical case (11).

Fig. 2 shows a gallery of the icosahedral capped prolates obtained in our simulations.‡ The bodies of these optimal structures are made of hexamers arranged either in rings of n (R_n) or zigzag (Z_n) layers (Table 1) with the exception of $T_{end} = 1$ (2-fold) and $T_{end} = 7$ (5-fold), which have a skewed body.

Not all possible icosahedral caps belong to energetically optimal elongated particles. We found that the prolates with their end-caps centered on 5-, 3-, and 2-fold axes but associated with the same T-number have different energies. Fig. 3 plots the sequences corresponding to $T_{end} = 1$ and $T_{end} = 3$. We observe that for $T_{end} = 1$ the 3-fold structure has a slightly lower energy than the 5-fold one, which might be the reason why some elongated plant viruses, like AMV (21) and aberrant FHV (24),

‡We used *gOpenMol* (36, 37) for 3D representations.

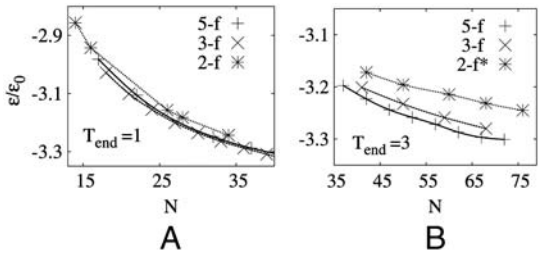

Fig. 3. (*A*) Sequence of local free-energy minima for $T_{end} = 1$ bacilliform particles centered on each of the possible axes of symmetry (the lines are just a guide for the eye). (*B*) Same plot for $T_{end} = 3$, where the 2-fold structures have been computed based on a geometrical model.

seem to have a capsid based on a 3-fold rather than the typical 5-fold construction of bacteriophages (25). Moreover, the energy of the 2-fold particles is not that high, and some of them are in fact locally optimal structures, hence we cannot disregard that some viruses follow this symmetry, at least for aberrant forms. For $T_{end} = 3$, structures associated with the caps centered on 3-fold axis still appear as locally optimal architectures for particular numbers of capsomers, despite having in general higher energies than those centered on the 5-fold axis. In fact, based on diffraction studies, it has been suggested that some plant viruses, for instance RTBV (23) and badnavirus genus (38), have this particular structure. In contrast, for $T_{end} = 3$ the 2-fold symmetry does not appear as an optimal structure in our simulations, and for $T_{end} > 3$ only the 5-fold symmetry appears to be an optimal architecture. Moreover, it turns out that for $T_{end} > 1$ the 5-fold-based structure is consistently the one having the smallest energy, followed by that of the 3-fold and the 2-fold.

Plots of the minimized energy per capsomer versus the number of capsomers at a fixed radius reveal a well-defined sequence of energy minima associated with a specific cap configuration. Fig. 4 illustrates, as an example, a plot of $\varepsilon(N)$ versus N for a fixed radius, R, that corresponds to a $T_{end} = 7$ cap centered on a 5-fold axis. The smallest optimal prolates associated with these energy minima are also shown in the figure. It is remarkable that the difference in the number of capsomers between these structures is fixed and equal to $\Delta N = 5$ for the case presented in the figure, which is in agreement with Moody's rule, i.e., $\Delta N = 5f$, for the construction of prolate viruses with caps centered on 5-fold symmetry (26), where ΔN is the smallest number of hexamers that have to be added to the cylindrical part of the capsid to grow it in the longitudinal direction. However, our results also reveal the existence of different rules associated with the prolates which are closed by T-numbered caps centered on the 3-fold or 2-fold axes. Table 1 summarizes our results for the different optimal sequences, the set of new ΔN's and the basic properties of their related structures. In particular, by generalizing the CK and

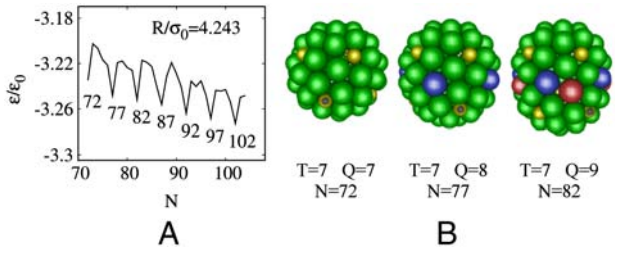

Fig. 4. (*A*) Free energy per capsomer ε versus number of capsomers N at fixed radius R, corresponding to the case $T_{end} = 7$ with two different morphological units. (*B*) Images of the structures corresponding to the first three minima of the sequence: a spherical $T = 7$ capsid and two prolate capsids with $T_{end} = 7$, and $Q = 8$ and $Q = 9$, respectively. At each step hexamers introduced in the body are colored in blue ($Q = 8$) and red ($Q = 9$). Moreover, the highlighted pentamer illustrates the rotation of the inferior cap.

Table 2. Optimal sequences for spherocylinders with one kind of capsomer ($|\Delta E/k_BT| \gg 1$)

$\langle R \rangle$*	Cap†	Body	N^0_{seq}	ΔN	$\langle \Delta L \rangle$
1.703	1 (5-fold)	R_5	12	5	1.72
1.724	1 (3-fold)	Z_6	15	3	1.01
2.593	snub cube	skew	24	3	0.69
2.789	3 (5-fold)	Z_{10}	32	5	1.01
2.912	3 (3-fold)	R_9	32	9	1.75
3.344	cubic	Z_{12}	44	6	1.03
4.235	7 (5-fold)	skew	72	5	0.65
4.244	7 (3-fold)	skew	72	3	0.39

*Columns have the same meaning as in Table 1.
†T-numbers of icosahedral caps correspond to T^*_{end} (see text).

Moody's ideas to the construction of 3-fold and 2-fold prolates, it is possible to show that the minimum step in the number of capsomers in these cases is

$$\Delta N_{3\text{-fold}} = \begin{cases} 9f & |h_0 - k_0| \propto 3 \\ 3f & \text{the rest} \end{cases}, \quad [2]$$

$$\Delta N_{2\text{-fold}} = \begin{cases} 14f & |h_0 - 2k_0| \propto 7 \\ 2f & \text{the rest} \end{cases}. \quad [3]$$

This geometrical prediction is in accordance with what we have found in our simulations and constitutes an extension of Moody's rule for any prolate capsid with icosahedral caps.

One Type of Morphological Unit. Some viruses, such as those of the polyomaviridae family or the maize strike virus (MSV), are made from only one type of capsomer (pentamers in the cited cases) and can assemble into elongated particles (30, 39). We can study the set of possible optimal structures of these capsids by increasing the energy difference ΔE between pentamers and hexamers. For spherical capsids (11), an increase of the energy difference between pentamers and hexamers leads to a dramatic reorganization of the optimal structure spectrum.

A significant reorganization of the spectrum of the optimal sequences is also observed for the spherocylindrical capsids if ΔE is set to a large value. The dashed line in Fig. 1 shows the results of the optimal energy per capsomer as a function of the radius $\varepsilon(R)$, for $N = 82$ and $|\Delta E|/k_BT \gg 1$, which corresponds to all capsomers being pentamers. Note that for the same number of capsomers N, the energy of structures composed of two types of capsomers tends to be lower than that of structures composed of one type of morphological unit. Analyzing the energy profiles for different values of N, we find caps with $T^*_{end} = 1, 3$, and 7 structures centered on both 5-fold and 3-fold symmetries.[§] No cap with $T^*_{end} = 4$ symmetry was found (see Table 2). Nonicosahedral caps with the structure of a snub cube, a chiral octahedral Archimedean solid with $N = 24$ capsomers, or with cubic symmetry are also obtained for a number of prolate structures. In fact, all our caps assume structures that have been found in the study of spherical capsids in ref. 11. The nonicosahedral structures can be a possible explanation of the aberrant structures of some viruses with one type of morphological unit. We emphasize that these structures are constructed by only one kind of capsomer and are not accounted for by Moody's geometrical model. However, we also observe that the selection rules, ΔN, for all semiicosahedral capped architectures are independent of having one or two types of capsomers in the capsid (see Tables 1 and 2).

[§]The triangulation number T^* indicates that the distribution of capsomers is equivalent to the classical T but with only one type of capsomer.

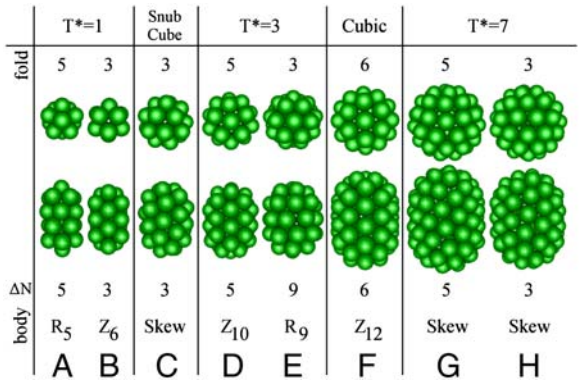

Fig. 5. Gallery of optimal representative prolate capsids with icosahedral caps when $|\Delta E|/k_BT \gg 1$, i.e, for one morphological unit, where capsomers are all pentamers. The top images show a zenithal view of the cap and the bottom images a lateral view of the complete bacilliform.

Interestingly, the polyoma virus can produce in vitro elongated particles, where pentamers occupy hexavalent positions in the cylindrical body (39). In particular, the most frequent case has approximately the radius of the native $T^* = 7$ icosahedral shell and presents a skew body, which should correspond to our result for the 5-fold or 3-fold prolate with $T^*_{\text{end}} = 7$ cap (Fig. 5 *G* and *H*). Moreover, polyoma is also able to assemble into icosahedral $T = 1$ and nonicosahedral snub cube particles (40). Since elongated particles with smaller radii have been also observed, we hypothesize that some of them could correspond to the bacilliforms $T^*_{\text{end}} = 1$ (5-fold) or (3-fold), and the elongated snub cube (see Fig. 5).

Quite remarkable is the structure of geminate viruses and the effect of genome length on the size of their capsids. For example, both MSV (30) and African cassava mosaic virus (41), which are the subject of intense research for gene delivery, adopt a spherical $T = 1$ shell, but form geminate capsids only when the full length genome is present (30). A geminate capsid is composed of 22 pentamers and its structure resembles a $T^*_{\text{end}} = 1$ (5-fold) (Fig. 5*A*).

Method of Prolate Classification. A careful study of the structures of prolate capsids obtained in our simulations indicates that, for a fixed cap radius, only one particular arrangement of hexamers in the body optimizes the total energy of the structure. Furthermore, we found that in the majority of prolates, the radii related to the optimal structure of the caps also correspond to the minima energy structures of bacilliform capsids. Note that for icosahedral caps, the class *P* and the axial symmetry (5-fold, 3-fold, or 2-fold) determine the particular pattern of hexamers in the body. The pattern, indeed, defines the sequence, i.e., the particular values of N and the minimum step ΔN compatible with the structure. Therefore, each optimal structure is characterized by a particular combination of $(R, \Delta N)$ or $(R, \Delta L)$ values. This restriction gives rise to a simple rule, which permits us to determine the structure of prolate viruses using a small set of experimentally accessible data. For instance, if due to the limitations on the dihedral angles between protein subunits, a particular virus can only assume a specific radius but with different lengths to form a prolate structure, the precise structure of the virus can be constructed using the results of our model. A particularly good example is AMV as it adopts prolate shapes with the same radius but different number of protein subunits (22), namely, $N_{\text{sub}} = 60$, 132, 150, 186, and 240, when reconstituted in vitro. The smallest capsid is a $T = 1$ structure composed of 12 pentamers. Considering that the body of the virus is formed only by hexamers, the number of proteins can easily be translated into the number of capsomers, obtaining the series $N = 12$, 24, 27, 33, and 42. According to our model, the structure of these prolate AMV particles can only be explained if the caps of the virus belong to a $T_{\text{end}} = 1$ structure centered on 3-fold symmetry axis, as it was suggested by Hull (21, 27). In a similar way, the structure of the two aberrant FHV (24) seem to follow the same architecture with $N = 15$ and $N = 18$. Furthermore, diffraction images of the elongated body of RTBV (23) suggest that this virus is a bacilliform $T_{\text{end}} = 3$ (3-fold), and the same seems to apply for the genera badnavirus (38). On the other hand, the $\text{AMV}_{\text{Top}-a-t}$ particle is made of 120 subunits and does not follow the general series of AMV, but its radius is similar to the other AMV particles. Based on our results, a plausible structure for this virus could be a $T_{\text{end}} = 1$ (5-fold) or (2-fold) with $N = 22$. According to ref. 42, this particle is elongated but slightly flattened. This structure could be a prolate centered on a 2-fold axis, which due to the distribution of pentamers in the caps, might adopt a more flattened shape (Fig. 2*C*). We emphasize that all these 3-fold and 2-fold structures do not follow the CKM geometrical model because they are not centered on the 5-fold axis. On the other hand, the existence of these well-defined rules for ΔN (see Tables 1 and 2) suggest that, starting from a particular virus, it should be possible to control the length of a perfectly closed capsid using, for instance, the right assembly conditions or different lengths of genetic material or polyelectrolyte. This possibility would be especially appealing for applications in encapsulation, nanopatterning, and the fabrication of nanowires.

Conclusions

According to our findings, the icosahedrally capped prolates centered on 5-fold axial symmetry proposed by Moody are indeed free-energy minima structures. In addition, we extended Moody's geometrical construction rules to include bacilliform viruses with icosahedral caps centered on 3-fold and, exceptionally, on 2-fold axes. We also explored other nonicosahedral structures that compete energetically with the *T*-number capped structures, especially for one type of capsomer.

Prolate viruses show a sequence of magic numbers for the caps, thus leading to strict selection rules for the length and structure of the body, as well as for the number of capsomers N assumed by each structure. We have been able to present a simple rule for the classification of prolates which could be helpful for the experimental determination of structures, for instance directly from electron microscopy images. Our results shed some light on the structure of several elongated viruses, such as ϕ29, AMV, aberrant FHV, RTBV, badnavirus, polyoma tubular particles, and geminate viruses, many of which have not yet been fully characterized. Moreover, the structural insights provided by the results of this work open the door to the design and control of the structure and dimensions of artificial viral capsids which can be extremely useful as nanotemplates and customized containers in nanotechnological applications.

ACKNOWLEDGMENTS. The authors are grateful to Tara Armand, Jan Wedekind, and Professors Charles Knobler and Umar Mohideen for critically reading the manuscript. We acknowledge support from the National Science Foundation (Grant DMR-06-45668), the Spanish Ministry of Science and Innovation (FIS2008-04386), and the Generalitat de Catalunya and European Social Fund (FI2009-B1-96 and I3).

1. Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2004) *Principles of Virology* (ASM Press, Washington, DC).
2. Douglas T, Young M (2006) Viruses: Making friends with old foes. *Science* 312:873–875.
3. Lee S-W, Mao C, Flynn CE, Belcher AM (2002) Ordering of quantum dots using genetically engineered viruses. *Science* 296:892–895.
4. Comellas-Aragonès M, et al. (2007) A virus-based single-enzyme nanoreactor. *Nat Nanotechnol* 2:635–639.
5. Adolph KW, Butler PJ (1976) Assembly of a spherical plant virus. *Philos Trans R Soc London B* 276:113–122.
6. Xie Z, Hendrix RW (1995) Assembly in vitro of bacteriophage HK97 proheads. *J Mol Biol* 253:74–85.

7. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI (1999) Assembly and analysis of conical models for the HIV-1 core. *Science* 283:80–83.
8. Lavelle L, et al. (2009) Phase diagram of self-assembled viral capsid protein polymorphs. *J Phys Chem B* 113:3813–3819.
9. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp Quant Biol* 27:1–24.
10. Bruinsma RF, Gelbart WM, Reguera D, Rudnick J, Zandi R (2003) Viral self-assembly as a thermodynamic process. *Phys Rev Lett* 90:248101.
11. Zandi R, Reguera D, Bruinsma RF, Gelbart WM, Rudnick J (2004) Origin of icosahedral symmetry in viruses. *Proc Natl Acad Sci USA* 101:15556–15560.
12. Zlotnick A (2003) Are weak protein-protein interactions the general rule in capsid assembly?. *Virology* 315:269–274.
13. Keef T, Taormina A, Twarock R (2005) Assembly models for Papovaviridae based on tiling theory. *Phys Biol* 2:175–188.
14. Šiber A, Podgornik R (2007) Role of electrostatic interactions in the assembly of empty spherical viral capsids. *Phys Rev E* 76:061906.
15. Hagan MF (2008) Controlling viral capsid assembly with templating. *Phys Rev E* 77:051904.
16. Rapaport DC (2008) Role of reversibility in viral capsid growth: A paradigm for self-assembly. *Phys Rev Lett* 101:186101.
17. Baker TS, Olson NH, Fuller SD (1999) Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol R* 63:862–922.
18. (2002) *The Springer Index of Viruses*, ed Tidona CA (Springer-Verlag, Berlin Heidelberg).
19. Tao Y, et al. (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* 95:431–437.
20. Fokine A, et al. (2004) Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci USA* 101:6003–6008.
21. Hull R, Hills GJ, Markham R (1969) Studies on alfalfa mosaic virus. *Virology* 37:416–428.
22. Heijtink RA, Houwing CJ, Jaspars EMJ (1977) Molecular weights of particles and RNAs of alfalfa mosaic virus. Number of subunits in protein capsids. *Biochemistry* 16:4684–4693.
23. Hull R (1996) Molecular biology of rice tungro virus. *Annu Rev Phytopathol* 34:275–297.
24. Dong XF, Natarajan P, Tihova M, Johnson JE, Schneemann A (1998) Particle polymorphism caused by deletion of a peptide molecular switch in a quasiequivalent icosahedral virus. *J Virol* 72:6024–6033.
25. Moody MF (1999) Geometry of phage head construction. *J Mol Biol* 293:401–433.
26. Moody MF (1965) The shape of the T-even bacteriophage head. *Virology* 26:567–576.
27. Hull R (1976) The structure of tubular viruses. *Adv Virus Res* 20:1–32.
28. Chen T, Glotzer SC (2007) Simulation studies of a phenomenological model for elongated virus capsid formation. *Phys Rev E* 75:051504.
29. Nguyen HD, Brooks CL III (2008) Generalized structural polymorphism in self-assembled viral particles. *Nano Lett* 8:4574–4581.
30. Zhang W, et al. (2001) Structure of the maize streak virus geminate particle. *Virology* 279:471–477.
31. Reddy VS, et al. (1998) Energetics of quasiequivalence: Computational analysis of protein-protein interactions in icosahedral viruses. *Biophys J* 74:546–558.
32. Zandi R, Reguera D (2005) Mechanical properties of viral capsids. *Phys Rev E* 72:021917.
33. Hu Y, Zandi R, Anavitarte A, Knobler CM, Gelbart WM (2008) Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size. *Biophys J* 94:1428–1436.
34. Zandi R, van der Schoot P (2009) Size regulation of ssRNA viruses. *Biophys J* 96:9–20.
35. Thuman-Commike PA, Greene B, Malinski JA, King J, Chiu W (1998) Role of the scaffolding protein in P22 procapsid size determination suggested by $T = 4$ and $T = 7$ procapsid structures. *Biophys J* 74:559–568.
36. Laaksonen L (1992) A graphics program for the analysis and display of molecular dynamics trajectories. *J Mol Graphics* 10:33–34.
37. Bergman DL, Laaksonen L, Laaksonen A (1997) Visualization of solvation structures in liquid mixtures. *J Mol Graph Model* 15:301–306.
38. Hull R, Lockhart BE, Reddy DVR, Schoelz JE (2000) Family Caulimoviridae. *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*, eds van Regenmortel MHV, et al. (Academic Press, San Diego), pp 335–347.
39. Baker TS, Caspar DLD, Murakami WT (1983) Polyoma virus "hexamer" tubes consist of paired pentamers. *Nature* 303:446–448.
40. Salunke DM, Caspar DLD, Garcea RL (1989) Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys J* 56:887–900.
41. Frischmuth T, Ringel M, Kocher C (2001) The size of encapsidated single-stranded DNA determines the multiplicity of African cassava mosaic virus particles. *J Gen Virol* 82:673–676.
42. Cusack S, Oostergetel GT, Krijgsman PC, Mellema JE (1983) Structure of the top_{a-t} component of alfalfa mosaic virus. A non-icosahedral virion. *J Mol Biol* 171:139–155.