The Structure of Elongated Viral Capsids

Antoni Luque* and David Reguera
Departament de Física Fonamental, Facultat de Física, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

ABSTRACT There are many viruses whose genetic material is protected by a closed elongated protein shell. Unlike spherical viruses, the structure and construction principles of these elongated capsids are not fully known. In this article, we have developed a general geometrical model to describe the structure of prolate or bacilliform capsids. We show that only a limited set of tubular architectures can be built closed by hemispherical icosahedral caps. In particular, the length and number of proteins adopt a very special set of discrete values dictated by the axial symmetry (fivefold, threefold, or twofold) and the triangulation number of the caps. The results are supported by experimental observations and simulations of simplified physical models. This work brings about a general classification of elongated viruses that will help to predict their structure, and to design viral cages with tailored geometrical properties for biomedical and nanotechnological applications.

INTRODUCTION

Viruses are submicroscopic organisms constituted, in their simplest form, by an infective genetic material (DNA or RNA) and a protective protein shell (the capsid) (1). The remarkable structural and mechanical properties of viral capsids have been a subject of increasing interest in the fields of biomedicine and nanotechnology in recent years (2–5).

In general, each virus has a well-defined wild-type shell that can be rodlike, quasispherical, bacilliform, or conical. However, by controlling the environmental conditions, e.g., the pH and salt concentration, many viruses can self-assemble in vitro in different shapes (6–10). The formation of these common and well-defined capsid architectures is essentially a consequence of a general physical principle: the free energy minimization of weak interactions among identical units (11–13).

The size of viruses, on the order of tens to hundreds of nanometers, restricts the amount of information that can be coded in the viral genome. Therefore, capsids are typically built from multiple copies of one or a few similar small proteins for the sake of genetic economy (14). These subunits interact with each other and self-assemble into a regular hollow shell. In two dimensions, the hexagonal lattice or its dual, i.e., the triangular one, maximize the packing and the number of interactions of identical units. Starting from them, it is possible to build all basic capsid shapes. The open helical tube characteristic of rodlike viruses can be obtained by simply wrapping the lattice. To construct closed shells, one needs to introduce 12 pentagonal defects (15). If they are evenly distributed, one gets the polyhedral shell with icosahedral symmetry of quasispherical viruses. Prolate or bacilliform capsids can be made by wrapping the lattice into an helical tube and closing each of its ends with six defects. Finally, conical viruses are obtained by making a closed tube with a different number of defects at both ends.

The physical and geometrical principles leading to the formation of rodlike and quasispherical icosahedral viruses are now relatively well understood. However, many viruses, including some bacteriophages, such φ29 or T4, and several fungus, plant, and animal viruses, e.g., in the genera barnavirus, badnavirus, and ascovirus, respectively (16), have a prolate capsid whose geometrical construction is not so well understood.

Recent works have shed some light on the structure of prolate viruses. In particular, Nguyen et al. (17) compared the energy of spherical, tubular, and conical viral shells using continuum elasticity arguments. In addition, some simulations performed in the literature (18,19) have obtained elongated shells from physical models of different degrees of complexity. Moreover, by using a simple model of capsomer-capsomer interaction that successfully explained the structure of spherical viruses (12), we have shown that the optimal structures for closed elongated viruses are, in general, hexagonally ordered tubes closed by hemispherical caps with icosahedral symmetry (13).

The main goal of this work is to describe the geometrical principles that lead to the construction of such bacilliform viral capsids. We will focus our attention on closed elongated viruses, which are also labeled in the literature as prolate, bacilliform, elongated, tubular, or allantoid (16), leaving aside specifically open-ended rodlike viruses such as tobacco mosaic virus. Our work is based on the ideas introduced by Caspar and Klug (20) and further extended by Moody (21), and establishes a general geometrical framework to describe icosahedral spherical capsids as well as icosahedrally capped bacilliform shells. The choice of these particular structures is not arbitrary and has been justified on energetic grounds (13).

The importance of this geometrical description is that it enumerates and characterizes the set of structures that can be built. We find that prolate capsids adopt a discretized

Submitted November 2, 2009, and accepted for publication February 26, 2010.

*Correspondence: luque@ffn.ub.es or tonisantolaria@gmail.com

Editor: Reinhard Lipowsky.

© 2010 by the Biophysical Society
0006-3495/10/06/2993/11 $2.00

doi: 10.1016/j.bpj.2010.02.051

set of lengths, radii, and numbers of proteins. The knowledge of these selection rules can be very useful to infer the structure of a virus from simple experimental data as well as for nanotechnological applications that rely on a proper control of the dimensions and architecture of viral capsids.

Icosahedral capsids

Roughly half of all viral species have a quasispherical capsid with icosahedral symmetry. Their construction rules were introduced in 1962 by Caspar and Klug (20), based on the idea of quasiequivalence (Fig. 1). In general, the optimal way to build a capsid made of identical units is to arrange them in a regular polyhedron where all of them sit in identical environments and share the same interactions with each other. However, the largest regular polyhedron that can be built fulfilling this strict requirement of equivalence is an icosahedron containing 60 proteins (three on each of its 20 faces), which corresponds to the structure of the smallest quasispherical viruses. The best alternative to build larger capsids is to pack these identical proteins in a limited number of quasiequivalent positions optimizing their mutual interactions. Caspar and Klug (20) showed that this leads necessarily to icosahedral symmetry as the most efficient design. In these capsids proteins can be geometrically clustered in two types of morphological units: pentamers, which are five proteins aggregated around each vertex of the icosahedron, and hexamers, which are clusters of six proteins evenly distributed on the faces and edges of the capsid.

Starting from a flat hexagonal lattice or equivalently its dual, the triangular one, there is a limited number of ways to create a closed shell with icosahedral symmetry. Essentially, one has to replace 12 evenly-distributed hexamers by the 12 pentamers required by Euler's theorem to make a closed surface (15). The different ways to accomplish that correspond to different triangulation numbers (T) that classify the quasispherical icosahedral viruses.

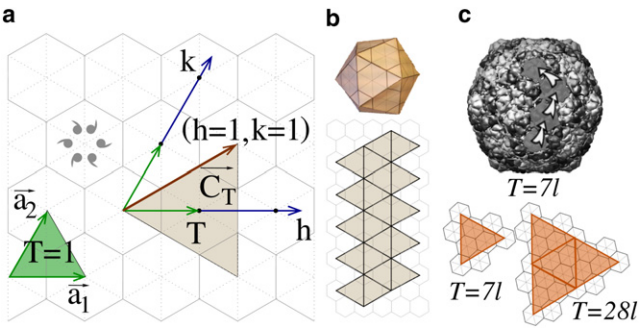

FIGURE 1 (*Color online*) (*a*) Basic elements of the Caspar and Klug construction. The shaded face is a $T = 3$ ($h = k = 1$). The icosahedral shell is built by 20 of these triangles. (*b*) Flat icosahedral template for a $T = 3$ virus (*bottom*) and the resulting folded capsid (*top*). (*c*) (*Top*) Example of a $T = 7l$ capsid corresponding to bacteriophage HK97 (22). Arrows indicate the steps along the hexagonal lattice ($h = 2$,$k = 1$) from one pentamer to the next. (*Bottom*) Two triangular faces from the class $P = 7l$.

The T-number is the area of a triangular face of the icosahedral shell defined by the vector that joins two adjacent pentamers in the lattice, namely

$$\vec{C}_T = h\vec{a}_1 + k\vec{a}_2 \equiv (h, k), \tag{1}$$

where (h, k) are nonnegative integers that give the number of steps to connect two nearest pentamers along the principal directions of the hexagonal lattice, i.e., $\vec{a}_1$ and $\vec{a}_2$ (Fig. 1 *a*).

The smallest triangular face is defined by (1, 0) or equivalently (0, 1), and has an area $S_0 = |\vec{a}_1 \times \vec{a}_2|/2$. T is the number of these basic triangles contained in a face of the resulting icosahedra, i.e., the area of the equilateral triangle defined by $\vec{C}_T$ divided by S_0. Using the elementary properties of the hexagonal lattice listed in Supporting Material A, one obtains

$$T = h^2 + hk + k^2 = Pf^2. \tag{2}$$

As h and k are nonnegative integers, T follows a particular series of ''magic'' numbers, i.e., $T = 1, 3, 4, 7, 9, 12, 13, \ldots$, which also correspond to the number of quasiequivalent locations in the shell (20,23).

Because an elementary $T = 1$ triangle accommodates three proteins and the resulting icosahedron is built by 20 T-faces, the total number of proteins in the capsid is

$$N_{\text{sub}} = 60T. \tag{3}$$

Every structure has $N_P = 12$ pentamers, accounting for 60 proteins, and the remaining subunits are distributed in $N_H = 10(T - 1)$ hexamers. Therefore the total number of capsomers in the capsid is

$$N = 10T + 2. \tag{4}$$

In modern structural virology, icosahedral capsids are always described by $T(h, k)$. However, the triangulation number is not always unique; e.g., for some $T \geq 49$, more than one pair (h, k) share the same T, e.g., (7, 0) and (5, 3) give $T = 49$. Therefore, Caspar and Klug (20) proposed also a reorganization of the T-structures in terms of classes P (not to be confused with the pseudotriangulation number used to label icosahedral capsids made of chemically different proteins) with common geometrical properties. Defining f as the greatest common divisor of h and k, i.e., $f = gcd(h, k)$, we can rewrite $h = fh_0$ and $k = fk_0$. Thus, from Eq. 2, we obtain

$$P = h_0^2 + h_0k_0 + k_0^2. \tag{5}$$

The class P is a subset of T ($P = 1, 3, 7, 13, \ldots$) and groups those icosahedral capsids that have an analogous distribution of hexamers (Fig. 1 *c*). Shells with (h, 0) or (0, k), i.e., $T = 1, 4, 9, \ldots$, belong to the class $P = 1$; those characterized by (h, h), i.e., $T = 3, 12, 27, \ldots$ belong to $P = 3$; and any class $P > 3$ is skewed, so that (h, k) generates a chiral structure specular to the shell generated by (k, h). To distinguish these situations we use the labels l, laevo or left-handed, for $h > k$,

and d, dextro or right-handed, for $h < k$. As an example, a $T = 7l$ capsid is plotted in Fig. 1 *c*.

For a given $P = T(h_0, k_0)$, the value of f^2 defines the number of P-triangles necessary to tile the $T(h, k)$-face (see Eq. 2). For instance, the face of a $T = 28l$ has four times ($f = 2$) the distribution of proteins of a $T = P = 7l$ triangle (Fig. 1 *c*). Shells belonging to the same class have an analogous arrangement of proteins, and so they should show similar physical properties, as evidenced, for instance, in the distribution of local stresses (24). Furthermore, the class P plays an important role in the geometrical properties of prolate capsids.

It is appropriate to make some clarifications regarding the Caspar and Klug model that will also apply to our extension for prolate shells. The Caspar and Klug construction only determines the point symmetry and is compatible with different clustering of the proteins. Thus, capsids with $20T$ trimers, $30T$ dimers, or $60T$ monomers are also possible. Moreover, the subunit in the Caspar and Klug model is not necessarily a single protein. For instance, the bluetongue virus core (25) has 120 proteins leading to a forbidden $T = 2$, which violates the model of Caspar and Klug. However, in terms of dimers the capsid contains 60 units organized as in a $T = 1$ shell.

On the other hand, not all quasispherical viruses strictly comply with the Caspar and Klug model. In their native form, polyoma and papilloma viruses are built only with pentamers arranged in a $T = 7$ capsid (26). Polyomavirus can also be reconstituted in vitro in a quasispherical, but nonicosahedral, structure that resembles a snub cube (26), which is completely outside the Caspar and Klug model. In this context, we must point out that both Caspar and Klug's capsids and the exceptions mentioned above have been found to be free energy minima of protein aggregates (12) and can also be explained using a tiling approach (27,28).

PROLATE OR BACILLIFORM CAPSIDS: A GENERALIZED GEOMETRICAL MODEL

A significant number of viral species have a closed elongated capsid whose precise structure is not so well characterized. In the late 1960s, Moody (21,29) extended the ideas of Caspar and Klug and described the construction of prolate capsids based on the elongation of an icosahedron along a fivefold axis of symmetry (Fig. 2).

The structure of some bacteriophages complies with Moody's model, as it has been confirmed by cryo-electron microscopy (cryo-EM) reconstructions (30,31). However, several bacilliform plant viruses, such the alfalfa mosaic virus (AMV), seem to have threefold rather than fivefold axial symmetry (32). Hull (33) and Moody (21) put forward the hypothesis that bacilliform viruses could have a tubular body closed by half icosahedral caps cut in different ways, but, except for the fivefold case, they did not describe precisely the geometrical rules to construct them.

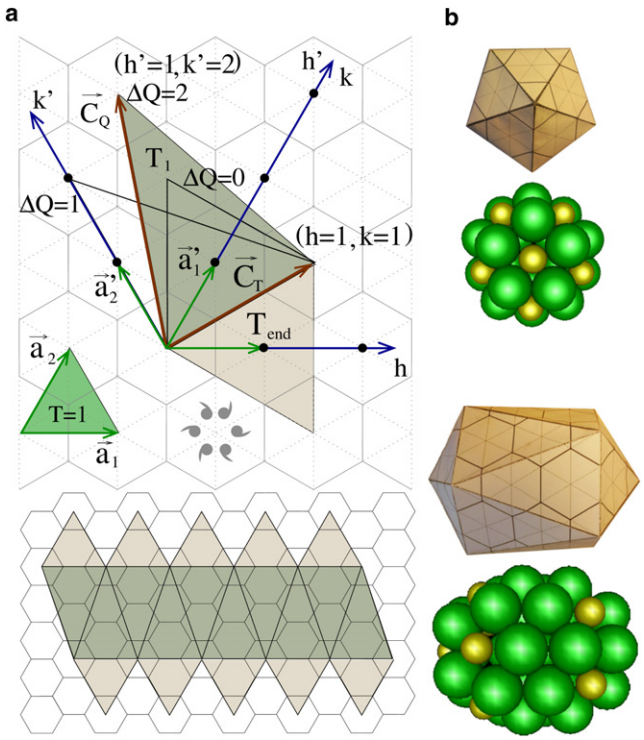

FIGURE 2 (*a*) (*Top*) Illustration of Moody's geometrical model for fivefold prolate capsids. (*Bottom*) Complete flat design of a $T_{end} = 3$ and $T_{mid} = Q = 5$ prolate capsid, which corresponds to the shell of a $\phi 29$ (30). (*b*) Zenithal (*top*) and lateral (*bottom*) views of the folded structure of a $T_{end} = 3$ and $T_{mid} = Q = 5$ prolate capsid. Below each view, there is a ping-pong model representation of the same capsid, where hexamers are colored in green and pentamers in gold.

In this section, we generalize Caspar and Klug and Moody's models to build bacilliform shells by elongating an icosahedron along all its different axes of symmetry: fivefold, threefold, and twofold. This model sets the basis for the geometrical characterization of prolate capsids, leading also to general rules that dictate the total number of proteins in any bacilliform shell as well as the number of subunits necessary to increase its length.

To make an elongated capsid starting from an icosahedron, its 20 triangular faces must be distributed among a tubular body and two equivalent caps, with the requirement of keeping six vertexes, i.e., pentamers, regularly arranged, in each cap.

This leads to a different scenario for each of the three possible axial symmetries (Fig. 3). In the fivefold case (Fig. 3 *a*), each cap is made by five triangular faces, whereas in the threefold and twofold situations the caps are made by four faces (Fig. 3, *b* and *c*). In the twofold case, the two specular possibilities (shown in Figs. 3, *c* and *d*) are valid to define the cap.

The number of faces in the body is then obtained by subtracting from the 20 faces of an icosahedron the number of triangles involved in the caps. Thus in the fivefold case the body has 10 triangular faces, and in the threefold and twofold

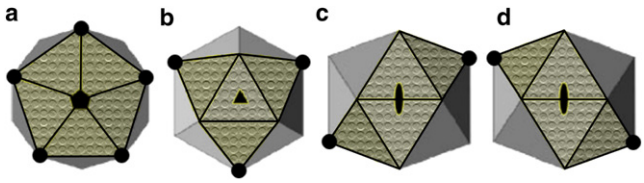

FIGURE 3 The three axes of symmetry of an icosahedron: fivefold (*a*), threefold (*b*), and twofold (*c* and *d*). The patterned triangles emphasize the end-faces that constitute the cap of the elongated structure. The solid dots highlight the vertexes that define the rim of the caps. In panels *c* and *d*, we show that the construction of the twofold prolate is intrinsically skewed and has two possibilities.

situations it has 12. But in general, not all these triangular body faces will be equal. Symmetry arguments allow us to calculate how many nonequivalent triangles we need to construct the body of these prolates. Any elongated capsid has two types of symmetry: the axial one, i.e., fivefold, threefold, or twofold, and the equivalency between caps, i.e., a twofold axis in the middle of the body perpendicular to the axis of the capsid. The product of the two symmetries defines the symmetry number, which sets the number of body-faces that must be equivalent. Hence, by dividing the total faces in the body by the symmetry number, we obtain the number of nonequivalent body-triangles. In the fivefold case the body has 10 faces, and the symmetry number is 10. Thus we only need one kind of triangle to characterize the tubular part (Fig. 2). Threefold and twofold prolates have 12 body-faces, and symmetry numbers six and four, respectively. Therefore, we need two different body-triangles in the threefold case (Fig. 4), and three in the twofold one (Fig. 5).

All triangles in both caps are equilateral and equal, and are determined by the cap vector $\vec{C}_T$, (Eq. 1). As in the Caspar and Klug model, this vector defines the triangulation number of the caps $T_{\text{end}} \equiv T$ (Eq. 2), and fixes the radius of the prolate. To describe the triangular faces of the elongated body we need a second vector $\vec{C}_Q$, which connects a pentamer in one cap to the closest one in the opposite cap, and it is given by

$$\vec{C}_Q = h'\vec{a}_1' + k'\vec{a}_2' \equiv \left(h', k'\right)', \tag{6}$$

where $(h', k')'$ are integers from a second pair of hexagonal coordinates, rotated counterclockwise 60° with respect to the original ones (Fig. 2 and Supporting Material A). Even though it is not strictly necessary to define a new pair of axes, this representation is more convenient, because for $h' = h$ and $k' = k$ we recover, in the body, the equilateral triangle that defines the face of an icosahedron.

Fivefold prolates

This case was studied by Moody (21,29), and it is the simplest situation because the body is made by 10 copies of the same midtriangle (Fig. 2). This triangular body-face is defined straightforwardly by $\vec{C}_T$ and $\vec{C}_Q$ (Eqs. 1 and 6). Its normalized surface, i.e., $|\vec{C}_T \times \vec{C}_Q|/2S_0$, defines a new triangulation number

$$T_1 = hh' + hk' + kk' \equiv Q_1 f, \tag{7}$$

which Moody labeled as T_{mid}. This T_1-number can also be expressed in terms of $f = gcd(h, k)$, with

$$Q_1 = h_0 h' + h_0 k' + k_0 k', \tag{8}$$

which we rename as $Q_{5F} \equiv Q_1$ in the fivefold case.

For $(h' = h, k' = k)$ we obtain $T_1 = T_{\text{end}}$ and $Q_{5F}^0 = Pf$, thus recovering a Caspar and Klug's icosahedral shell. The elongation with respect to the spherical capsid is then characterized by $\Delta Q \equiv Q - Q^0$. It is geometrically possible to build structures with $T_1 < T_{\text{end}}$, i.e., $\Delta Q < 0$, but this oblate capsid have not been found experimentally (29).

The sum of triangulation numbers of all faces gives the total surface of the bacilliform capsid. In this case there are 10 end-triangles in the caps and 10 midtriangles in the body. The total number of proteins in the capsid is then

$$N_{\text{sub}}^{5F} = 3(10\,T_{\text{end}} + 10\,T_1) = 30f(Pf + Q_{5F}). \tag{9}$$

As in icosahedral capsids, the 12 pentamers of a prolate require 60 proteins. Thus the number of hexamers is $N_{\text{H}}^{5F} = 5(T_{\text{end}} + T_1) - 10$, and the total number of capsomers is

$$N_{5F} = 5(T_{\text{end}} + T_1) + 2 = 5f(Pf + Q_{5F}) + 2. \tag{10}$$

The value of T_{end} controls the radius of the structure. If we fix it, i.e., P and f are constant, the different values of $Q_{5F}(h', k')$ in Eq. 10 give the possible lengths of the prolate in terms of number of capsomers. As h' and k' are integers, the number of capsomers and proteins in the body of a prolate can only adopt a discrete set of values. One can prove using Bezout's identity (Supporting Material B) that the minimum step possible in Q_{5F} is

$$\Delta Q_{5F}^{\min} = 1. \tag{11}$$

Thus, unlike T_{end}, Q_{5F} can be any nonnegative integer. Combining Eqs. 10 and 11, the minimum length step of a prolate in terms of capsomer numbers is

$$\Delta N_{5F}^{\min} = 5f\Delta Q_{5F}^{\min} = 5f \tag{12}$$

or $\Delta N_{\text{sub}}^{5F} = 30f$ in terms of subunits. Hence, prolates based on $f = 1$ caps, e.g., $T_{\text{end}} = 1$ or $T_{\text{end}} = 3$, have lengths discretized in steps of five capsomers, i.e., 30 proteins. However, those based on $f = 2$, e.g., $T_{\text{end}} = 4$ or $T_{\text{end}} = 12$ caps, must add multiples of 10 capsomers, i.e., 60 proteins, to enlarge the structure.

There are some examples of fivefold prolate viruses, especially among bacteriophages. For instance, φ29 is a $T_{\text{end}} = 3$, $Q_{5F} = 5$ (30), and bacteriophage T4 has $T_{\text{end}} = 13l$ and $Q_{5F} = 20$ (31).

Threefold prolates

Prolates can also be made by the elongation of an icosahedron along one of its threefold axes. To build this structure one needs two different types of body-triangles (Fig. 4), as discussed before. The first triangle is the same T_1 used in the fivefold situation. The second triangle is defined by the body vector $\vec{C}_Q$ and a 120° counterclockwise rotation of the cap vector

$$\vec{C}_T^{120^o} = (-h-k)\vec{a}_1 + h\vec{a}_2 \equiv (-h-k, h). \quad (13)$$

This is because these two nonequivalent body-faces and three end-triangles from the cap must share a common vertex at the origin, defining a pentamer. Therefore, to close the structure properly, we must introduce a 60° wedge in the plane between the second triangle and the adjacent end-face (Fig. 4). The normalized surface of this second triangle is then $|\vec{C}_Q \times \vec{C}_T^{120^o}|/2S_0$, which defines a new triangulation number

$$T_2 = hh' + kh' + kk' \equiv Q_2 f, \quad (14)$$

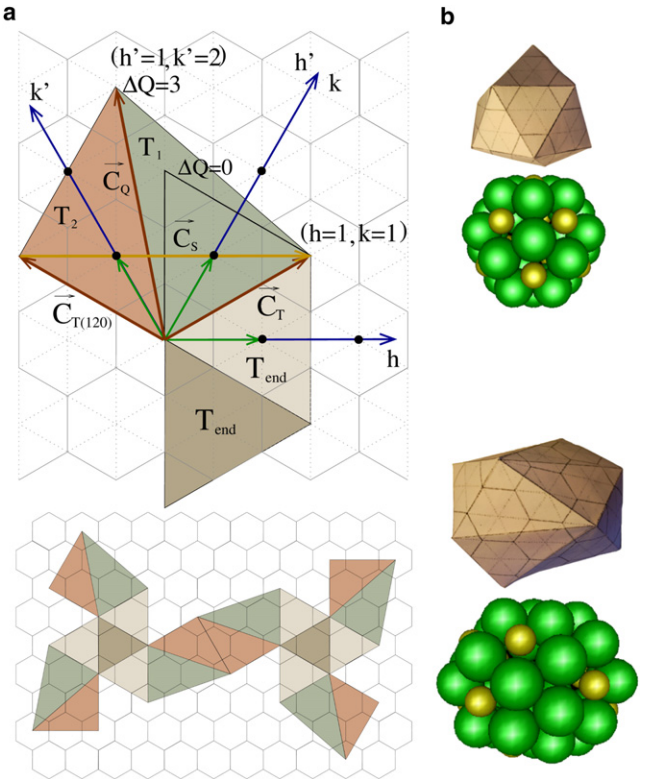

FIGURE 4 (*a*) (*Top*) Basic elements to build a prolate based on hemispherical icosahedral caps centered on a threefold axis. The vector $\vec{C}_s = \vec{C}_h/3$ (in *yellow*) joins to consecutive pentamers along the rim of the cap. (*Bottom image*) Complete flat design of the prolate with the eight end-triangles and the 12 body-triangles. (*b*) Zenithal (*top image*) and lateral (*bottom image*) view of the resulting folded structure, along with its ping-pong model representation. The case illustrated in this figure corresponds to a $T_{end} = 3$ and $Q_{3F} = 9$.

where

$$Q_2 = h_0 h' + k_0 h' + k_0 k'. \quad (15)$$

To characterize the threefold body we sum up the contributions of the two nonequivalent midtriangles, yielding

$$Q_{3F} \equiv Q_1 + Q_2 = h_0\left(2h' + k'\right) + k_0\left(2k' + h'\right), \quad (16)$$

which for the spherical case reduces to $Q^0_{3F} = 2Pf$. Note that the value of Q_{3F} for the isometric particle starts at $Q^0_{3F} = 2\,T_{end}/f$.

The surface of the capsid determines the total number of subunits as in the fivefold case. Now we have eight end-triangles with T_{end}, six midtriangles with T_1, and another six with T_2, which leads to

$$N^{3F}_{sub} = 3(8\,T_{end} + 6\,T_1 + 6\,T_2) = 6f(4Pf + 3Q_{3F}). \quad (17)$$

Again, the 12 pentamers of the prolate account for 60 proteins, so the number of hexamers is $N^{3F}_H = 4\,T_{end} + 3(T_1 + T_2) - 10$, and the total number of capsomers is

$$N_{3F} = 4T_{end} + 3(T_1 + T_2) + 2 = f(4Pf + 3Q_{3F}) + 2. \quad (18)$$

As before, the value of Q_{3F} controls the length of the shell and can only adopt a discrete set of values. In this case the minimum step possible in Q_{3F} is (Supporting Material B)

$$\Delta Q^{min}_{3F} = \begin{cases} 3 & \text{if } |h_0 - k_0| \propto 3 \\ 1 & \text{the rest} \end{cases}. \quad (19)$$

Therefore, the possible lengths of threefold prolates increase at discrete steps of capsomers:

$$\Delta N^{min}_{3F} = 3f\Delta Q^{min}_{3F} = \begin{cases} 9f & \text{if } |h_0 - k_0| \propto 3 \\ 3f & \text{the rest} \end{cases}. \quad (20)$$

Accordingly, there are two different situations depending on the value of $|h_0 - k_0|$. In particular, for the class $P = 1$ ($h_0 = 1, k_0 = 0$) we get $\Delta N_{3F} = 3f$, hence the possible lengths of a $T_{end} = 1$ ($f = 1$) capped shell are discretized by steps of at least $\Delta N^{min}_{3F} = 3$ capsomers, i.e., 18 proteins. This growing law agrees with the results obtained for AMV (34) and supports its classification as a threefold $T_{end} = 1$ bacilliform particle (see Applications: Structural Characterization of Prolate Viruses, below). On the other hand, a $P = 3$ ($h_0 = k_0 = 1$) prolate has $\Delta N^{min}_{3F} = 9f$. Thus, the lengths of a $T_{end} = 3$ ($f = 1$) capsid correspond to multiples of $\Delta N^{min}_{3F} = 9$ capsomers. In fact, rice tungro bacilliform virus (RTBV) has been suggested to be a threefold $T_{end} = 3$ prolate (35).

Twofold prolates

In this case, the body is determined by three nonequivalent midtriangles (Fig. 5). The first body-face is again the T_1-triangle. The second midtriangle is the T_2-triangle introduced in the threefold case. The third midtriangle is

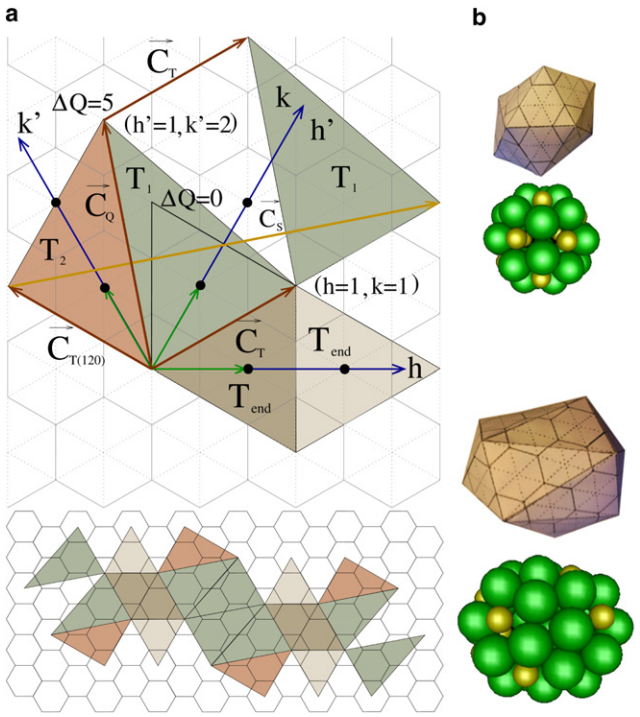

FIGURE 5 (*a*) (*Top*) Basic elements required to build a prolate capsid with twofold axial symmetry. The vector $\vec{C}_s = \vec{C}_h/2$ (in *yellow*) joins to consecutive pentamers in the rim of the cap. (*Bottom*) Complete flat design of the prolate with the eight end-triangles and the 12 body-triangles. (*b*) Zenithal (*top image*) and lateral (*bottom image*) view of the resulting folded structure, along with its ping-pong model representation. The case illustrated in this figure corresponds to a $T_{end} = 3$ and $Q_{5F} = 14$.

a translation by $\vec{C}_T$ of the first one and has the same triangulation number T_1, but they are not equivalent because is not possible to relate each other by applying only the symmetries of a twofold prolate. To take into account the three nonequivalent midtriangles, it is convenient to define the body in terms of

$$Q_{2F} \equiv 2Q_1 + Q_2 = h_0\left(3h' + 2k'\right) + k_0\left(3k' + h'\right), \quad (21)$$

which for the case of a spherical structure reduces to $Q^0_{2F} = 3Pf = 3\,T_{end}/f$.

We can compute the total number of proteins of the prolate, as we did for the fivefold and threefold cases. Now we have eight end-triangles in the caps with T_{end}, and four of each of the three midtriangles in the body, where two of them share the same T_1 number. Thus, we obtain

$$N^{2F}_{sub} = 3(8\,T_{end} + 8\,T_1 + 4\,T_2) = 12f(2Pf + Q_{2F}). \quad (22)$$

The twofold prolate has 12 pentamers, $N^{2F}_H = 4\,T_{end} + 2(2\,T_1 + T_2) - 10$ hexamers, and a total number of capsomers

$$\begin{aligned} N_{2F} &= 4T_{end} + 2(2T_1 + T_2) + 2 \\ &= 2f(2Pf + Q_{2F}) + 2. \end{aligned} \quad (23)$$

Again, the value of Q_{2F} determines the length of the prolate, which can grow at discretized steps of (Supporting Material B)

$$\Delta Q^{min}_{2F} = \begin{cases} 7 & \text{if } |h_0 - 2k_0| \propto 7 \\ 1 & \text{the rest} \end{cases}, \quad (24)$$

or, in terms of capsomers,

$$\Delta N^{min}_{2F} = 2f\Delta Q_{2F} = \begin{cases} 14f & \text{if } |h_0 - 2k_0| \propto 7 \\ 2f & \text{the rest.} \end{cases} \quad (25)$$

We have then two different cases depending on the value of h_0 and k_0. In particular for the class $P = 1$ ($h_0 = 1, k_0 = 0$) or $P = 3$ ($h_0 = k_0 = 1$) the growing law is $\Delta N^{min}_{2F} = 2f$, hence the possible number of capsomers of a $T_{end} = 1$ or a $T_{end} = 3$ capped shell is discretized at intervals of two capsomers. On the other hand, for $P = 7l$ with ($h_0 = 2, k_0 = 1$) the minimum step is $\Delta N^{min}_{2F} = 14f$. Note that for twofold prolates, the chirality is important. For instance, for the specular case ($h_0 = 1, k_0 = 2$), we obtain $\Delta N^{min}_{2F} = 2f$. Thus, a shell based on a $T = 7l$ can have different lengths separated by steps of 14 capsomers, whereas for a $T = 7d$ the minimum step is two hexamers.

Additionally, prolates with twofold axial symmetry can be made in two different and nonequivalent ways, depending on how we chose the triangles of the caps (Fig. 3, *c* and *d*). In this work we have chosen the distribution shown in Fig. 3 *c*, which we can call twofold dextro. A similar construction can be made with the laevo selection of end-cap triangles (Fig. 3 *d*) obtaining the same growing rules, but interchanging h_0 by k_0 in Eq. 25.

We are not aware of any prolate virus which is known for sure to have a twofold construction. However, it is possible that an aberrant particle of the AMV (36) could be the case (discussed below in Applications: Structural Characterization of Prolate Viruses).

A tubular description

The generalized model of elongated capsids introduced above allows us to enumerate all possible icosahedral prolates. In this section, we will describe a procedure to compute, for any icosahedrally capped shell, the radius, the length, and the position of the capsomers in the body. This geometrical characterization of the resulting capsids was carried out neither in Caspar and Klug nor in Moody's model, but it turns out to be quite helpful for both recognition and design of viral shells.

The tubular body of an elongated virus can be built by rolling up an hexagonal sheet (Fig. 6), much in the same way as with carbon nanotubes (37). However, for prolate viral capsids, only the subset of tubes closed by icosahedral caps with fivefold, threefold, or twofold axial symmetry is valid.

This procedure involves an approximation, because one assumes that the surface of the resulting cylinder is the

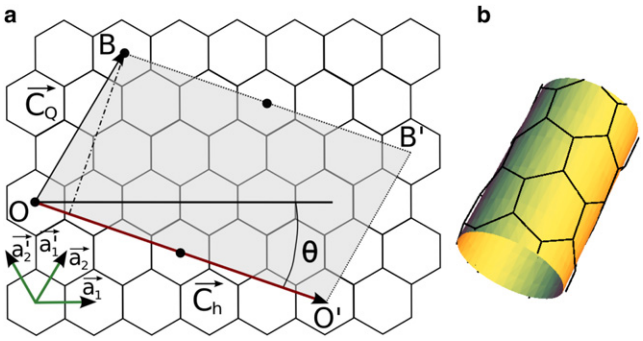

FIGURE 6 (*a*) The unrolled tubular body of a prolate virus (*shaded area*) shown on the honeycomb lattice. The solid dots indicate the location of the pentamers in the rim. (*b*) Tubular body obtained by rolling up the shaded area in the direction of $\vec{C}_h$ so that O meets O' and B meets B'. The example corresponds to a $T_{\text{end}} = 1$ and $Q_{2F} = 9$ twofold prolate, with $h = 1$, $k = 0$, and $h' = 3$, $k' = 0$.

same as that of the flat lattice, which implies that hexamers will be bent and stretched in the tube. However, it can be shown that it is a very good approximation (see Supporting Material H).

Radius of the tube

In the tubular approach (Fig. 6), the circumference of the tube is determined by the chiral vector

$$\vec{C}_h = m\vec{a}_1 + n\vec{a}_2 \equiv (m, n), \tag{26}$$

which belongs to the hexagonal lattice, i.e., m and n are integers, and connects all pentamers along the rim of the cap. Hence, $\vec{C}_h$ is related to the cap vector $\vec{C}_T$, but differently for each axial symmetry.

The fivefold case is particularly simple. The rim of the tube is delimited by the five vertexes of the icosahedron, i.e., pentamers, that lie on a plane perpendicular to the axis of the prolate (Fig. 3 *a*). Therefore, by unrolling the body of the capsid, it is easy to see that the circumference of the tube is just made by five times the cap vector $\vec{C}_T$ (Eq. 1):

$$\vec{C}_h^{5F} = 5\vec{C}_T = 5f(h_0, k_0). \tag{27}$$

Hence, for $P = 1$ we obtain $\vec{C}_T^{5F} = (5f, 0)$, and for $P = 3$ we have $\vec{C}_T^{5F} = (5f, 5f)$.

In the threefold case, the circumference of the cap is defined by three nonconsecutive vertexes (pentamers) in a section perpendicular to the axis (Fig. 3 *b*). The vector that connects two of these vertexes lying on the rim is plotted in Fig. 4 and, in terms of the cap vector, is given by $\vec{C}_T - \vec{C}_T^{120^\circ}$. Thus the chiral vector is just obtained by summing up three times this vector

$$\vec{C}_h^{3F} = 3\left(\vec{C}_T - \vec{C}_T^{120^\circ}\right) = 3f(2h_0 + k_0, k_0 - h_0). \tag{28}$$

Similarly, in the twofold case, the circumference of the tube is defined by the two pentamers that are farther apart and lie on a section perpendicular to the axis (Fig. 3 *c*). As we can see in Fig. 5, the rim vector that connects the two pentamers in the unrolled body is $2\vec{C}_T - \vec{C}_T^{120^\circ}$. Hence, the chiral vector is made by two times the rim vector, i.e.,

$$\vec{C}_h^{2F} = 2\left(2\vec{C}_T - \vec{C}_T^{120^\circ}\right) = 2f(3h_0 + k_0, 2k_0 - h_0). \tag{29}$$

For symmetry reasons, the specular construction defined in Fig. 3 *d* leads to the same results, but permuting h_0 and k_0.

The radius of the resulting tube is, in all cases,

$$R = \frac{|\vec{C}_h|}{2\pi}, \tag{30}$$

and the particular expressions for each symmetry are listed in Table S1. From these results, we also observe that for a given T_{end}

$$R_{5F} \lesssim R_{3F} \lesssim R_{2F}, \tag{31}$$

hence, the fivefold prolate has the smallest radius, followed by the threefold and finally the twofold structure.

Chiral angle and distribution of hexamers in the body

The distribution of hexamers in the body of a prolate is determined by the chiral angle θ, which is the angle between $\vec{C}_h$ and the vector $\vec{a}_1$ of the hexagonal lattice (Fig. 6). Accordingly,

$$\cos(\theta) = \frac{\vec{C}_h \cdot \vec{a}_1}{|\vec{C}_h|} = \frac{2m + n}{2\sqrt{m^2 + mn + n^2}}, \tag{32}$$

where, due to the symmetry of the hexagonal lattice, θ is defined between 0° and 60°.

There are two special situations. For $n = 0$ ($\theta = 0^\circ$) or $m = 0$ ($\theta = 60^\circ$), i hexamers in the body are arranged in rings, R_i, whereas for $m = n$ ($\theta = 30^\circ$), i hexamers are distributed in zigzag rows, Z_i (rings and zigzag layers correspond to zigzag and armchair structures, respectively, in carbon nanotubes).

The components of the chiral vector $\vec{C}_h$ depend on the axial symmetry. Therefore, ring and zigzag bodies are associated to different classes P for each axial symmetry (Table S2). For instance, rings are obtained for $P = 1$ fivefold, $P = 3$ threefold, and $P = 7l$ twofold, whereas zigzag appear for $P = 1$ threefold, $P = 3$ fivefold, and $P = 21d$ twofold.

Smallest length step and particle length

The chiral vector defines the radius of the tube, but its height is controlled by the body vector $\vec{C}_Q$, which connects two pentamers in different caps (Fig. 6). Hence, the length of the tubular part of the capsid is given by the perpendicular projection of $\vec{C}_Q$ onto $\vec{C}_h$, namely,

$$L = \frac{|\vec{C}_h \times \vec{C}_Q|}{|\vec{C}_h|} = \frac{a^2\sqrt{3}}{2}\frac{Q}{|\vec{C}_h|}sf, \quad (33)$$

where the value of s can be 5, 3, or 2 in consonance with the axial symmetry. Because Q is discretized by $\Delta Q^{\min}$, the possible lengths of a prolate are discretized by

$$\Delta L^{\min} = \frac{a^2\sqrt{3}}{2}\frac{\Delta Q^{\min}}{|\vec{C}_h|}sf. \quad (34)$$

This is directly related to the elongation of the body

$$\Delta L \equiv \Delta L^{\min}\left(Q - Q^0\right) \equiv \Delta L^{\min}\Delta Q, \quad (35)$$

which becomes zero for a spherical capsid, when $Q = Q^0$. In Table S1 we list the values of these properties for each axial symmetry.

Another relevant geometrical property is the aspect ratio, which can be defined as the total length of the prolate divided by its width, i.e.,

$$ar = \frac{2R + \Delta L}{2R} = 1 + \frac{\Delta L}{2R}. \quad (36)$$

Finally, the position of all capsomers in the tube can also be evaluated, as shown in Supporting Material D.

Degeneracy and number of relative orientations

Some prolate structures can be built in different ways. Here the term ''degenerate'' refers to architectures that have the same T_{end}, axial symmetry, radius, and length, but differ in the relative orientation of pentamers in both caps. As shown in Supporting Material E, the degeneracy, i.e., the number of different structures is

$$D = f\Delta Q^{\min}. \quad (37)$$

Therefore, all elongated structures based on icosahedral caps with $f > 1$ or $\Delta Q^{\min} > 1$ are degenerate. In the fivefold case, we have $\Delta Q^{\min}_{5F} = 1$; therefore structures with $f = 1$, e.g., $T_{\text{end}} = 1, 3, 7, 13\ldots$, have always a unique prolate capsid, whereas for $f > 1$, e.g., $T_{\text{end}} = 4, 9, 12$, there are f shells with different relative orientation between the pentamers of the two caps. We stress that this occurs even for the spherical case when $T_{\text{end}} = T_1$. For instance, a $T_{\text{end}} = 4$ ($f = 2$) can adopt two spherical configurations (Fig. S2) but only one has full icosahedral symmetry. In fact, we have found that for $T = 4$ the nonicosahedral structure is feasible and has the same free energy as the icosahedral one in simulations of a simple physical model for spherical capsids (13).

Interestingly, elongated structures with threefold and twofold axial symmetry can be degenerate even for $f = 1$, but only for classes P with $\Delta Q^{\min} > 1$. For example, there are three possible relative orientations between the caps of a $T_{\text{end}} = 3$ ($\Delta Q^{\min} = 3$) for a given length. Once again, this holds even for the nonelongated capsids, leading to three possible structures for a spherical $T = 3$ capsid, one of them with complete icosahedral symmetry and the other two just keeping the threefold axial symmetry. The same happens, for instance, with structures based on $T_{\text{end}} = 7d$ with twofold axial symmetry and the choice of caps of Fig. 3 *c*, where for a fixed length there are seven possible choices for the relative orientation between the caps, even for the nonelongated case.

RESULTS

The main results of this section are summarized in Table S1 and Table S2, listing the expressions of the relevant quantities for each axial symmetry (Table S1) along with their specific values for the smallest classes P in the Caspar and Klug classification (Table S2).

Applications: structural characterization of prolate viruses

In this section, we will use different viruses to illustrate how the insights gained in this work can be useful for characterization purposes.

Bacteriophage T4

Bacteriophage T4 is one of the few prolate viruses whose capsid structure has been determined at high resolution (31), and we will illustrate that is possible to infer its structure using a few experimental data.

In particular, we will use the experimental diameter of the capsid, $2R_{\text{exp}} = 86 \pm 3$ nm, and the distance between hexamers in the body, $a^{\text{exp}} = 14 \pm 2$ nm, both obtained from Fokine et al. (31). Inserting these data in the formulas of the radius for the different symmetries listed in Table S1, we obtain $T_{5F} = 14.9 \pm 1.6$, $T_{3F} = 13.8 \pm 1.5$, and $T_{2F} = 13.3 \pm 1.4$ as potential values for the triangulation number of the cap. Because $T_{\text{end}} = 14$ or 15 are not valid results, the triangulation number should be either $T_{\text{end}} = 13$ or 16.

Moreover, the shell is composed by 167 capsomers: 155 hexamers made of 930 copies of *gp*23*, 11 pentamers made of *gp*24*, and an effective pentamer corresponding to the *gp*20 connector. Taking into account the growing laws for the different T_{end}-caps proposed (Table S3), we observe that only $T_{\text{end}} = 13$ with fivefold axial symmetry leads to a capsid with 167 capsomers. The number of hexamers involved in the elongation is 35 because the icosahedral shell has 132 capsomers, thus from Eq. 12 we get $\Delta Q_{5F} = 7$. Taking into account that in the spherical case $Q^0_{5F} = 13$, our analysis suggests that the structure of bacteriophage *T*4 is a fivefold prolate with $T_{\text{end}} = 13$ and $Q_{5F} = 20$, which is in fact the structure resolved in the cryo-EM reconstruction. We can also compute the aspect ratio using Eq. 36, obtaining a value of 1.3 in agreement with the experimental value, 1.4 ± 0.2.

Therefore, using three simple inputs, i.e., diameter, distance between hexamers, and total number of proteins, it is possible to infer the structure of the virus. These data can be obtained from different experimental techniques, e.g., electron micrographs, optical diffraction, or sedimentation, but unfortunately seem not to be available for most viruses.

Alfalfa mosaic virus (AMV)

AMV is a well-studied plant virus and adopts different lengths depending on the amount of genetic material encapsidated (32,34) (Table S4). The number of protein subunits in the in vitro reconstituted capsids have been determined from their molecular weights and correspond to $N_{\rm sub} = 60$, 132, 150, 186, and 240 (34). Only the smallest capsid has been reconstructed by x-ray (38) and is a spherical $T = 1$ composed of 12 pentamers. The elongated particles have a diameter similar to that of the icosahedral one, hence they should have $T_{\rm end} = 1$ caps. Assuming that the body of the prolate particles is formed by hexamers, the number of proteins can easily be translated into number of capsomers, obtaining the series $N = 12, 24, 27, 33$, and 42. Therefore, a multiple of at least three hexamers is added in every step. According to our model, this can only be explained if AMV elongated particles adopt a structure $T_{\rm end} = 1$ centered on a threefold symmetry axis (Table S3 and Table S4). The same architecture was already proposed in the literature (32,33) based on optical diffraction and Geodestix models (Geodestix, Spokane, WA).

The predictions of our model are not only useful to infer the structure, but can also be used to extract other geometrical and structural information. For instance, the architecture proposed has a body made of hexamers arranged in a zigzag pattern Z_6, and the minimum step of $\Delta N_{3F} = 3$ capsomers corresponds to an increment in length of $\Delta L_{3f} = a/2$ (Table S2). Experimentally, it is known that each step of 18 subunits, i.e., three hexamers, increases the length by 4.34 nm (34). Hence the distance between hexamers should be $a \simeq 8.68$ nm, which is in agreement with diffraction analysis (32), and from that one can estimate, for instance, the typical size of a capsomer or a coat protein.

AMV also makes an aberrant elongated particle that contains 120 protein subunits, and does not follow the sequence discussed above (36). Assuming that the central body is built of hexamers, this number of proteins corresponds to $N = 22$ capsomers. In addition, the particle has again a similar radius suggesting that it is based on $T_{\rm end} = 1$ caps. Hence, in the framework of our model there are two possible capsids for this aberrant particle (Table S3): a $T_{\rm end} = 1$ bacilliform shell centered on a fivefold axis with $Q_{5F} = 2$, and a $T_{\rm end} = 1$ prolate centered on a twofold axis with $Q_{2F} = 5$. However, Cusack et al. (36) suggests that the particles show an oblate shape. In that case the twofold situation seems a better candidate, as twofold structures are quite distorted and could lead to deformed shapes. In any case, experimentally, it is not clear whether this aberrant particle is polymorphic.

Rice tungro bacilliform virus

The structure of this bacilliform virus has not been fully determined yet. The diameter of the tubular part, $2R_{\rm exp} = 30 \pm 3$ nm, has been obtained from EM micrographs, and diffraction experiments suggest that the distance between hexamers in the body is $a_{\rm exp} = 10 \pm 2$ nm, and they are arranged in rings (35). Geometrically, for each axial symmetry, there is only one class P having bodies made of hexamer rings: $P = 1$, $P = 3$, and $P = 7l$ for five-, three-, and twofold symmetries, respectively (Table S2). We can use the experimental estimate of $a_{\rm exp}$ to calculate what would be the radius of the cap for each of these possibilities. The result is, respectively,

$$2R_{5F}^{P=1} = f(16 \pm 2)\ \text{nm},$$

$$2R_{3F}^{P=3} = f(28 \pm 3)\ \text{nm},$$

and

$$2R_{2F}^{P=7} = f(44 \pm 9)\ \text{nm},$$

Thus, comparing with the experimental value $2R_{\rm exp}$, RTBV is either a $T_{\rm end} = 4$ (fivefold) or a $T_{\rm end} = 3$ (threefold) structure with a body made of rings of 10 or 9 hexamers, respectively. The lack of further experimental information does not allow us to discriminate between both possibilities. However, we can use our model to predict what would be the expected geometrical properties of the virus in each case. If RTBV is based on a $T_{\rm end} = 4$ (fivefold), its number of subunits should follow the law $N_{\rm sub}^{5F} = 240 + (n \times 60)$, and its total length should be $L(n) = 2R + n\Delta L$, where $\Delta L = 8.7 \pm 1.7$, which is the same for both architectures because they have ring-bodies. Experimentally, the length of the predominant particle is $L_{\rm exp} = 130 \pm 3$ nm. Therefore, we obtain a value of $n = 11 \pm 3$ for the number of steps, that taking into account that $Q_{5F}^{0} = 2$, leads to $Q_{5F} = 13 \pm 3$. Thus, the structure would have $N = 152 \pm 30$ capsomers or $N_{\rm sub} = 900 \pm 180$ proteins. Analogously, if RTBV is based on a $T_{\rm end} = 3$ (threefold) architecture, we would obtain a structure characterized by $N = 130 \pm 30$, $Q_{3\rm F} = 39 \pm 9$, and $N_{\rm sub}^{3F} = 180 + (n \times 54) = 770 \pm 160$ protein subunits. Note that, simply by knowing the total number of proteins or the molecular weight of the capsid, one could know which is the right structure.

Hull (35) proposed that RTBV is an elongated particle based on $T_{\rm end} = 3$ (threefold), which is one of the solutions of our analysis. However, from the experimental data used above we cannot reject the $T_{\rm end} = 4$ (fivefold) architecture.

CONCLUSIONS

In this work we have presented a geometrical model that establishes the architectural principles that control the construction of spherical and prolate viruses with icosahedral symmetry. Closed elongated viruses can be constructed by

the elongation of an icosahedron along a fivefold, threefold, or twofold axis of symmetry. Interestingly, there is a finite set of possibilities to do it, and that leads to discretization rules for the length and number of proteins. These rules are determined by the axial symmetry and the T number of the cap. Moreover, our analysis leads to expressions for the radius and length of prolates, as well as the arrangement of capsomers in the tubular body.

The model accounts for the Caspar and Klug structures (20) and the fivefold prolates described by Moody (21,29) as a special case, adding new geometrical details. More importantly, we have shown that it is possible to construct quasispherical capsids that can be conceived as two hemispherical caps rotated around one of their symmetry axes. These degenerate viruses do not have complete icosahedral symmetry, but are spherical structures that could compete and interfere in the assembly of viral particles. In fact, preliminary results suggest that, in terms of free energy, these structures are, in some cases, equal in stability to the normal icosahedral structures. In general, elongated viruses with $f > 1$ or $\Delta Q^{\min} > 1$ can also have more than one structure with the same length but differing on the relative orientation between the caps. This can add an extra complication to the experimental reconstruction of prolate structures.

It is worth mentioning that using a very simple model of interaction between capsomers, we have found that these icosahedral prolate structures are indeed free energy minima, thus justifying their possible occurrence in nature (13). However, not all of them seem to be equivalent in energy or even energetically optimal, which might be the reason why some structures, especially those based on twofold axial symmetry, seem hard to be observed in native viruses.

On the other hand, there are viruses, like polyomavirus, that are able to adopt elongated structures built exclusively by pentamers (39,40). Strictly, these structures do not follow the geometrical model described above. However, Luque et al. (13) showed that hexagonally ordered tubes closed by icosahedral caps and made only by one type of capsomer are energy minima and follow the same selection rules predicted by our model.

The results of this work open the door to a simple characterization of elongated viruses using a few parameters, e.g., subunit's size or number, particle dimensions or chirality of the body, which can be obtained from different standard experimental techniques, such as electrophoresis, electron microscopy and electron or x-ray diffraction.

The fact that prolate viruses can adopt different lengths suggest that in principle it should be possible to control it by using the proper assembly conditions and/or using, for instance, different lengths of genetic or nongenetic materials. This possibility would facilitate the design of artificial viral capsids in applications such as nanopatterning or nanotemplating. The structural information provided by the geometrical principles laid out in this work could be potentially very helpful in this task.

SUPPORTING MATERIAL

Four figures, four tables, and 30 equations in eight sections are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00332-2.

The authors are grateful to J. R. Caston and I. Fita for critically reading the manuscript.

We acknowledge support from the Spanish "Ministerio de Ciencia e Innovación" (grant No. FIS2008-04386 and I3 program) and the FI program (Generalitat de Catalunya and European Social Fund).

REFERENCES

1. Flint, S. J., L. W. Enquist, …, A. M. Skalka. 2004. Principles of Virology. ASM Press, Washington, DC.
2. Douglas, T., and M. Young. 2006. Viruses: making friends with old foes. *Science*. 312:873–875.
3. Lee, S.-W., C. Mao, …, A. M. Belcher. 2002. Ordering of quantum dots using genetically engineered viruses. *Science*. 296:892–895.
4. Hu, Y., R. Zandi, …, W. M. Gelbart. 2008. Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. *Biophys. J.* 94:1428–1436.
5. Sun, J., C. DuFort, …, B. Dragnea. 2007. Core-controlled polymorphism in virus-like particles. *Proc. Natl. Acad. Sci. USA*. 104:1354–1359.
6. Adolph, K. W., and P. J. Butler. 1976. Assembly of a spherical plant virus. *Philos. Trans. R. Soc. Lond., B*. 276:113–122.
7. Bancroft, J. B., C. E. Bracker, and G. W. Wagner. 1969. Structures derived from cowpea chlorotic mottle and brome mosaic virus protein. *Virology*. 38:324–335.
8. Lavelle, L., M. Gingery, …, J. Ruiz-Garcia. 2009. Phase diagram of self-assembled viral capsid protein polymorphs. *J. Chem. Phys.* 113:3813–3819.
9. Ganser, B. K., S. Li, …, W. I. Sundquist. 1999. Assembly and analysis of conical models for the HIV-1 core. *Science*. 283:80–83.
10. Heymann, J. B., C. Butan, …, A. C. Steven. 2008. Irregular and semi-regular polyhedral models for Rous sarcoma virus cores. *Comp. Math. Methods Med.* 9:197–210.
11. Zlotnick, A. 2003. Are weak protein-protein interactions the general rule in capsid assembly? *Virology*. 315:269–274.
12. Zandi, R., D. Reguera, …, J. Rudnick. 2004. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA*. 101: 15556–15560.
13. Luque, A., R. Zandi, and D. Reguera. 2010. Optimal architectures of elongated viruses. *Proc. Natl. Acad. Sci. USA*. 107:5323–5328.
14. Crick, F. H. C., and J. D. Watson. 1956. Structure of small viruses. *Nature*. 173:473–475.
15. Coxeter, H. S. M. 1989. Introduction to Geometry, 2nd Ed. Wiley, New York.
16. Tidona, C. A. and G. Darai (Eds.). 2002. The Springer Index of Viruses. Springer-Verlag, Berlin, Heidelberg.
17. 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.
18. Chen, T., and S. C. Glotzer. 2007. Simulation studies of a phenomenological model for elongated virus capsid formation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 75:051504.
19. Nguyen, H. D., and C. L. Brooks, 3rd. 2008. Generalized structural polymorphism in self-assembled viral particles. *Nano Lett.* 8: 4574–4581.
20. Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27:1–24.
21. Moody, M. F. 1965. The shape of the T-even bacteriophage head. *Virology*. 26:567–576.

22. Carrillo-Tripp, M., C. M. Shepherd, …, V. S. Reddy. 2009. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.* 37(Database issue):D436–D442.

23. Kerner, R. 2008. Classification and evolutionary trends of icosahedral viral capsids. *Comp. Math. Methods Med.* 9:175–181.

24. Zandi, R., and D. Reguera. 2005. Mechanical properties of viral capsids. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 72:021917.

25. Grimes, J. M., J. N. Burroughs, …, D. I. Stuart. 1998. The atomic structure of the bluetongue virus core. *Nature.* 395:470–478.

26. Salunke, D. M., D. L. D. Caspar, and R. L. Garcea. 1989. Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* 56:887–900.

27. Twarock, R. 2004. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* 226:477–482.

28. Twarock, R. 2005. Mathematical models for tubular structures in the family of *Papovaviridae*. *Bull. Math. Biol.* 67:973–987.

29. Moody, M. F. 1999. Geometry of phage head construction. *J. Mol. Biol.* 293:401–433.

30. Tao, Y., N. H. Olson, …, T. S. Baker. 1998. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell.* 95:431–437.

31. Fokine, A., P. R. Chipman, …, M. G. Rossmann. 2004. Molecular architecture of the prolate head of bacteriophage T4. *Proc. Natl. Acad. Sci. USA.* 101:6003–6008.

32. Hull, R., G. J. Hills, and R. Markham. 1969. Studies on alfalfa mosaic virus. II. The structure of the virus components. *Virology.* 37:416–428.

33. Hull, R. 1976. The structure of tubular viruses. *Adv. Virus Res.* 20:1–32.

34. Heijtink, R. A., C. J. Houwing, and E. M. J. Jaspars. 1977. Molecular weights of particles and RNAs of alfalfa mosaic virus. Number of subunits in protein capsids. *Biochemistry.* 16:4684–4693.

35. Hull, R. 1996. Molecular biology of rice tungro viruses. *Annu. Rev. Phytopathol.* 34:275–297.

36. Cusack, S., G. T. Oostergetel, …, J. E. Mellema. 1983. Structure of the top $_{A\text{-}T}$ component of alfalfa mosaic virus. A non-icosahedral virion. *J. Mol. Biol.* 171:139–155.

37. Saito, R., M. S. Dresselhaus, and G. Dresselhaus. 1998. Physical Properties of Carbon Nanotubes. Imperial College Press, London, UK.

38. Kumar, A., V. S. Reddy, …, J. E. Johnson. 1997. The structure of alfalfa mosaic virus capsid protein assembled as a T=1 icosahedral particle at 4.0-Å resolution. *J. Virol.* 71:7911–7916.

39. Kiselev, N. A., and A. Klug. 1969. The structure of viruses of the papilloma-polyoma type. V. Tubular variants built of pentamers. *J. Mol. Biol.* 40:155–171.

40. Baker, T. S., D. L. D. Caspar, and W. T. Murakami. 1983. Polyoma virus 'hexamer' tubes consist of paired pentamers. *Nature.* 303: 446–448.