Research Article **Open Access**

Farrah Sadre-Marandi, Jiangguo Liu*, Simon Tavener, and Chaoping Chen

Generating Vectors for the Lattice Structures of Tubular and Conical Viral Capsids

Abstract: Retrovirus capsid is a fullerene-like lattice consisting of capsid protein hexamers and pentamers. Mathematical models for the lattice structure help understand the underlying biological mechanisms in the formation of viral capsids. It is known that viral capsids could be categorized into three major types: icosahedron, tube, and cone. While the model for icosahedral capsids is established and well-received, models for tubular and conical capsids need further investigation. This paper proposes new models for the tubular and conical capsids based on an extension of the Capser-Klug quasi-equivalence theory. In particular, two and three generating vectors are used to characterize respectively the lattice structures of tubular and conical capsids. Comparison with published HIV-1 data demonstrates a good agreement of our modeling results with experimental data.

Keywords: CA protein; capsid; cone; hexamer; HIV-1; icosahedron; pentamer; tube

MSC: 92C05

DOI 10.2478/mlbmb-2014-0009
Received September 2, 2014; accepted November 8, 2014

1 Introduction

Viruses are macromolecular organisms that are constituted by infective genetic materials (DNA or RNA) and protective protein shells [8]. It is known that the viral capsid usually consists of capsid protein (CA protein) hexamers and pentamers. The binding among the CA hexamers and pentamers are further related to the hinge between the C-terminal domain (CTD) and N-terminal domain (NTD) of the CA protein [6, 12].

The CA hexamers and pentamers form a lattice structure that folds into a viral capsid. It has been observed that a viral capsid takes an icosahedral, tubular, conical or irregular shape [8, 17, 21].

The icosahedral viral capsid has been extensively studied due to the highly symmetric nature. This is the preferred geometry for viruses, since the symmetry allows 60 CA proteins or 12 pentamers to be placed on the surface in an equivalent manner, minimizing excess energy. There are many viruses whose capsids have more than 60 CA proteins, in fact around thousands of proteins. In these cases, not all of the subunits (CA proteins) can be placed in equivalent positions. Capser and Klug's theory of quasi-equivalence [7] addresses this issue by classifying icosahedral shells by similar protein neighborhoods rather than subunits. The Capser-Klug quasi-equivalence theory allows capsids with multiples of 60 subunits, indicated by the triangulation (T)

***Corresponding Author: Jiangguo Liu:** Department of Mathematics, Colorado State University, Fort Collins, CO 80523-1874, USA, E-mail: liu@math.colostate.edu
Farrah Sadre-Marandi: Department of Mathematics, Colorado State University, Fort Collins, CO 80523-1874, USA, E-mail: sadre@math.colostate.edu
Simon Tavener: Department of Mathematics, Colorado State University, Fort Collins, CO 80523-1874, USA, E-mail: tavener@math.colostate.edu
Chaoping Chen: Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523-1870, USA, E-mail: Chaoping.Chen@ColoState.EDU

number, to form with icosahedral symmetry. The quasi-equivalence is demonstrated in the experiment results reported in [30].

There are also many viruses that have tubular or lozenge-like capsids. Mature HIV-1 cores have cone-shaped capsids [1, 11]. Unlike the icosahedral viral capsids, the structures of tubular and conical capsids are not yet fully understood.

There have been models for the lattice structures of tubular and conical capsids, see [17, 21] and references therein. However, there is a flaw in the Nguyen model [21], which missed the requirement on the cone height and could result in incomplete cones if the model is inappropriately applied.

In this paper, we propose new models for tubular and conical capsids in a unified fashion based on an extension of the Capser-Klug quasi-equivalence theory. The new models are easier than the existing models. When applied to the HIV-1 (5,7)-cone (5 pentagons in the narrow end and 7 pentagons in the broad end), the capsid properties derived from our models show good agreement with published experimental data. This demonstrates the correctness and usefulness of the new models.

The rest of this paper is organized as follows. Section 2 briefly reviews the concepts of the T-number and generating vector for the icosahedral viral capsid. Section 3 presents a mathematical model for tubular viral capsids using two generating vectors and three parameters. Section 4 presents a model for conical viral capsids that uses three generating vectors and four parameters. Further details on the (5,7)- and (4,8)-cones are examined. According to the Euler theorem, there are 3 other possible cone angles for a hexagonal lattice, but models for the (1,11)-,(2,10)-, and (3,9)-cones are not investigated, since they are rarely seen in nature without overlapping in a spiral fashion [13]. Section 5 compares modeling results to published experimental data on the HIV-1 conical capsid. Section 6 concludes the paper with some remarks.

2 Generating Vector and T-number for Icosahedral Viral Capsids

This section briefly reviews the concepts of the T-number and generating vector for an icosahedral viral capsid. This will be helpful for understanding the models for tubular and conical viral capsids to be discussed in the following sections.

About half of the virus species are found to have an icosahedral capsid [17]. The geometric structure (symmetry and periodicity) of icosahedral capsids can be well characterized by the Caspar-Klug quasi-equivalence theory [7].

By the Euler theorem, for a convex polyhedron made of hexagons and pentagons, there are exactly 12 pentagons. When these 12 pentagons are evenly distributed, an icosahedron forms, which can be circumscribed into a sphere.

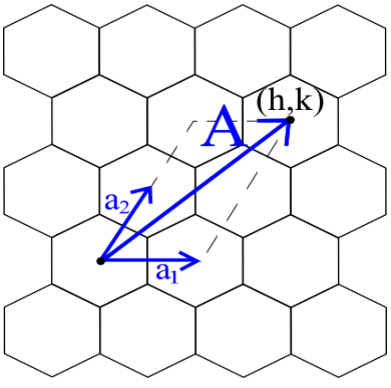

Figure 1: Two basis vectors $\vec{a}_1$, $\vec{a}_2$ and one generating vector $\vec{A}$ for a hexagonal lattice. In this illustration, $h = 1, k = 2$ and hence $T = h^2 + hk + k^2 = 7$. **Note: the two hexagons where the starting and ending points of the generating vector $\vec{A}$ reside will be replaced by two pentagons when the lattice is folded into an icosahedron.**

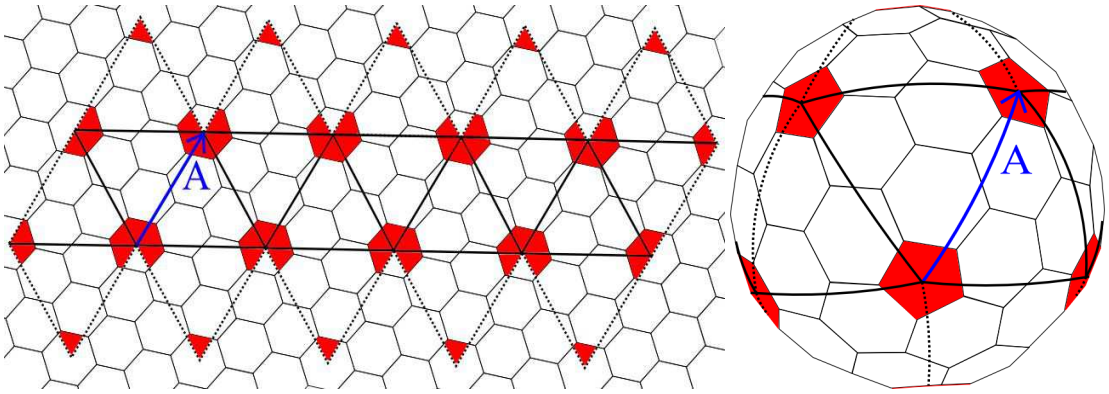

Figure 2: *Left*: A lattice with $(h, k) = (1, 2)$ and $T = h^2 + hk + k^2 = 7$. The dotted lines indicate where to cut the lattice to fold it. *Right*: The icosahedral capsid obtained from folding the lattice shown in the left panel. Pentagons are shown in red. An example for $T = 7$ is the widely studied bacteriophage HK97 [18].

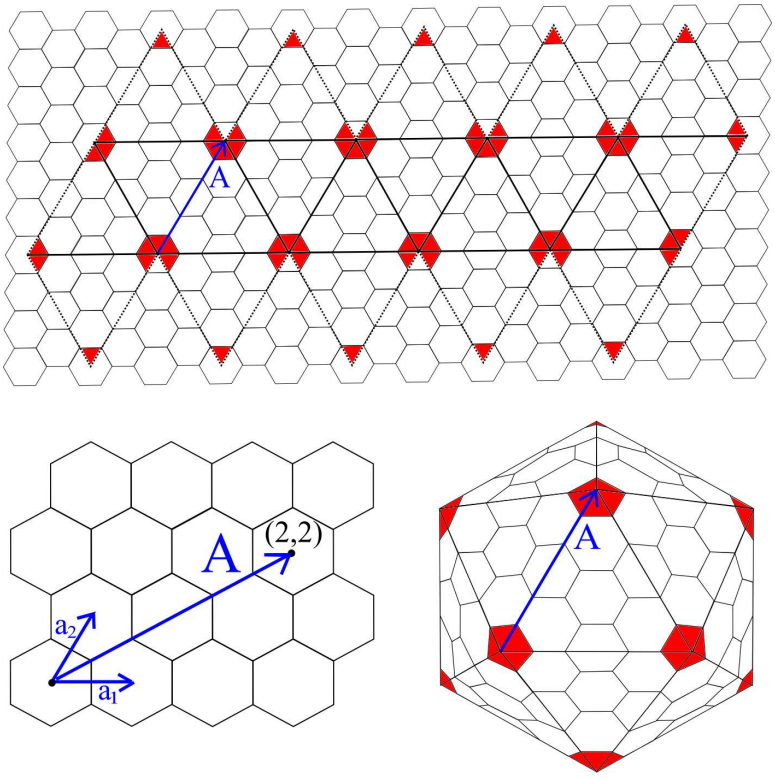

Figure 3: A commonly adopted approach for illustrating icosahedral viral capsids, see also [21] Figure 1. *Top*: The centers of the pentagons form a coarse triangular mesh; *Bottom Left*: Two basis vectors and one generating vector with $(h, k) = (2, 2)$ and $T = h^2 + hk + k^2 = 12$; *Bottom Right*: The lattice folds into an icosahedron with the pentagon centers being the triangle vertices. **Authors' note:** However, the folded pentagons and the flat triangles shown in (Bottom Right) need careful reading. Actually all pentagons are planar objects. The coarse triangles are used to locate the pentagons. The hexagons do **not** lie on the virtual triangles.

To understand the concepts of the generating vector and T-number, we start with a flat hexagonal lattice consisting entirely of identical hexagons. As shown in Figures 1 and 2, we choose the center of one hexagon as the origin and set the lengths of the basis vectors $\vec{a}_1, \vec{a}_2$ as 1. It is obvious that the angle between $\vec{a}_1$ and $\vec{a}_2$ is 60° and hence their inner product (dot product) is $\langle \vec{a}_1, \vec{a}_2 \rangle = \frac{1}{2}$. The generating vector, as shown in Figures

1 and 2, is defined as a linear combination of the two basis vectors

$$\vec{A} = h\vec{a}_1 + k\vec{a}_2, \tag{1}$$

where h, k are nonnegative numbers (but not both zero). Therefore, we have

$$|\vec{A}|^2 = \langle h\vec{a}_1 + k\vec{a}_2, h\vec{a}_1 + k\vec{a}_2\rangle = h^2\langle\vec{a}_1, \vec{a}_1\rangle + 2hk\langle\vec{a}_1, \vec{a}_2\rangle + k^2\langle\vec{a}_2, \vec{a}_2\rangle = h^2 + hk + k^2 =: T, \tag{2}$$

which is the so-called T-number. Geometrically, the T-number can be understood as the squared length of each triangle edge in the construction of the icosahedron. This relates (h, k) to the area of a single triangle by the formula: Area $= \frac{\sqrt{3}}{4}T$.

There are 20 equilateral triangles used in the construction of an icosahedron, placed symmetrically on a flat hexagonal lattice, as shown in Figure 2. The triangle size depends on the T-number, with a varying number of hexagons within. Each vertex of a triangle lands at the center of a hexagon, which is the very position of a pentagon when folded in three dimensions. The pentagon is formed by cutting a 60° wedge from a hexagon then adjoining the two cut edges. This creates a convex five-sided polygon, whose center is no longer on the hexagonal plane. Clearly, the T-number measures the distance squared between the centers of two nearby pentagons.

Figure 3 is another illustration for the icosahedral viral capsids, see also [21] Figure 1. This example has set $(h, k) = (2, 2)$ and hence $T = h^2 + hk + k^2 = 12$. Each triangle side has length $\sqrt{12}$, and the vertices lie at the centers of hexagons that will be replaced by pentagons when folded.

In structural virology, icosahedral capsids are usually described by $T(h, k)$. However, there is no guaranteed uniqueness for $T \geq 49$. For example, $(7, 0)$ and $(5, 3)$ both give $T = 49$, see [17]. To classify these virus capsids uniquely, Casper and Klug [7] proposed a reorganization of the T-number, in terms of the P classes. Any class with $P > 3$ (starting with $T = 7$) is skewed so that (h, k) generates a chiral structure mirrored by that created with (k, h) [7, 17]. Both chiral structures can be produced from this model by recreating the folding from a mirrored lattice. Further details are excluded from this paper so we may focus on the construction of the tube and conical capsids.

3 Generating Vectors for Tubular Viral Capsids

A tubular (spherocylinder) viral capsid has been observed for Cowpea Chlorotic Mottle Virus and Alfalfa Mosaic Virus, among others. As reported in [1, 4], HIV-1 core could also exhibit a tube-like capsid.

A tubular viral capsid consists of only CA protein hexamers and pentamers. The Euler theorem guarantees exactly 12 pentamers on the capsid, assuming it is a convex polyhedron. The tubular structure can be considered as cutting an icosahedron in half and extending the middle region by a hexagonal cylinder. Each end cap is a truncated icosahedron with exactly 6 pentagons and a varying number of hexagons, determined by the T-number.

The tubular model also follows the Casper-Klug quasi-equivalence theory. However, describing the lattice structure of a tubular viral capsid needs two generating vectors: one to describe the equal distance between the pentagons and the other for the varying height.

First, we define a vector $\vec{A} = h\vec{a_1} + k\vec{a_2}$ for the folding of the two end caps. This is the same as for the construction of an icosahedron. The two caps will be displaced some distance from each other, in the direction perpendicular to $\vec{A}$ by a new generating vector $\vec{B}$. The vector $\vec{B} = h\vec{b_1} + k\vec{b_2}$ is so defined that its basis vectors $\vec{b_1}$ and $\vec{b_2}$ are respectively orthogonal to $\vec{a_1}$ and $\vec{a_2}$. Specifically, we have, as shown in Figure 4 (*left panel*),

$$\vec{b_1} = \vec{a_1} - 2\vec{a_2}, \tag{3}$$

$$\vec{b_2} = 2\vec{a_1} - \vec{a_2}. \tag{4}$$

The folding template for a tubular viral capsid is shown in Figure 4. This is similar to the template for an icosahedron. The difference is exhibited in the ten triangles located at the center of the folding. They are

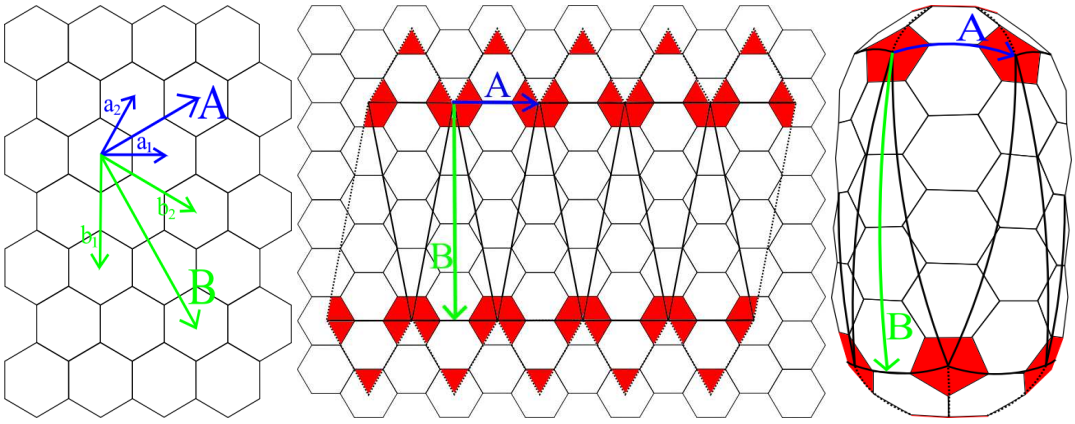

Figure 4: *Left*: Basis vectors and generating vectors. *Middle*: A lattice structure with $(h, k) = (1, 1)$, $T = h^2 + hk + k^2 = 3, \gamma = \frac{9}{6}$. The dotted lines indicate where to cut the lattice to fold it. Pentagons are shown in red. *Right*: The lattice folds into a tubular capsid.

no longer equilateral. This is due to the displacement of the end caps, creating an elongated hexagonal tube. Triangle vertices still lie at the centers of the hexagons that will be replaced by pentagons when folded in three dimensions.

In summary, the tubular folding template is constructed by two generating vectors

$$\vec{A} = h\vec{a_1} + k\vec{a_2}, \tag{5}$$
$$\vec{B} = \gamma(h\vec{b_1} + k\vec{b_2}), \tag{6}$$

where

$$\gamma = r\frac{\gcd(h,k)}{2T}, \tag{7}$$

and r is an integer. Here $\gcd(h, k)$ is the greatest common divisor of h and k. When $\gamma = 1/2$, this model reproduces an icosahedron described in the previous section.

The derivation for γ is intuitive. The only limitation on the height of a tubular capsid comes from the construction requirement that $\vec{A}$ must start and end at the center of a hexagon. There are multiple positions along the direction of $\vec{A}$ that satisfy this requirement. If h and k are not co-prime, then these positions do not create unique structures. If fact, similar positions occur along a triangle edge for every length of $\frac{\sqrt{T}}{\gcd(h,k)}$. Moving in the direction perpendicular to $\vec{A}$, there is a hexagon meeting the requirements for every integer increment of $\frac{\sqrt{3}}{2}\frac{\gcd(h,k)}{\sqrt{T}}$. Relating this to $|\vec{B}|$, we have

$$|\vec{B}| = \gamma\sqrt{3T} = r\frac{\sqrt{3}\gcd(h,k)}{2\sqrt{T}}, \tag{8}$$

and hence

$$\gamma = r\frac{\gcd(h,k)}{2T}. \tag{9}$$

The model proposed in this paper is similar to but more general than the tubular model introduced in [21]. In [21], the two generating vectors are defined as

$$\vec{A} = n(h\vec{a_1} + k\vec{a_2}), \tag{10}$$
$$\vec{B} = m(h\vec{b_1} + k\vec{b_2}), \tag{11}$$

where (h, k) are the same as those for the icosahedral capsid, (n, m) are two nonnegative integers, and b_1, b_2 are similarly defined (but have opposite directions) as in Equations (3)(4). For this model, scaling $h\vec{a_1} + k\vec{a_2}$ by

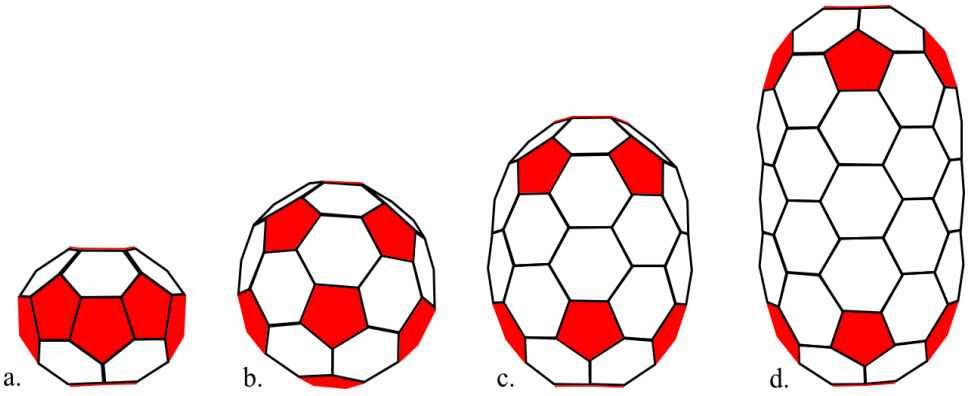

Figure 5: Tubular (spherocylindrical) capsid with $T = 3$ and varying γ. Pentagons are shown in red. From left to right: (a): $\gamma = \frac{1}{6}$, unlikely to occur in nature, but it is the smallest possible tube that can be created by this model. (b): $\gamma = \frac{3}{6}$, examples include Cowpea Chlorotic Mottle Virus or Norwalk Virus [5]. (c): $\gamma = \frac{5}{6}$, seen in bacteriophage $\phi 29$ [5, 25]. (d): $\gamma = \frac{8}{6}$, seen in Alfalfa Mosaic Virus [17].

a constant n is somewhat unnecessary, since all variations can be accounted for by varying h and k. Restricting the scaling constant in $\vec{B}$ to an integer excludes the model from covering several types of virus capsids, for instance, the occasional tubular shape of HIV-1 [1] and the bacteriophage $\phi 29$ [5, 25]. To use the model in [21] for creating the bacteriophage $\phi 29$ capsid shown Figure 5(c), n must be defined as $n = 1$ due to the end caps, and $m = \frac{5}{3}$ due to the height. This is not valid since both n and m are required to be integers.

4 Generating Vectors for Conical Capsids

A simple cone can be produced by rolling a section of a sheet around its apex and joining the two open sides. However, a cone created with a hexagonal lattice will not have infinitely many cone angles. With the hexagonal lattice, the hexagon/pentagon units along the closure line must match.

The Euler theorem implies that there are five possible cone angles for a hexagonal lattice as shown below

$$\sin(\theta/2) = 1 - P/6, \tag{12}$$

where θ is the cone angle and P is the number of pentagons located in the narrow end of the cone. The five angle values (and the corresponding P values) are $\theta = 112.9^\circ (P = 1)$, $\theta = 83.6^\circ (P = 2)$, $\theta = 60^\circ (P = 3)$, $\theta = 38.9^\circ (P = 4)$, $\theta = 19.2^\circ (P = 5)$, see [11]. For convenience, we name these cones as (1,11)-, (2,10)-, (3,9)-, (4,8)-, (5,7)-cones. In the notations for (i, j)–, i is the number of pentagons in the narrow region, j is the number of pentagons in the broad region, and $i + j = 12$. Most HIV-1 cones are in the (5,7)-pattern, but (4,8)-cones have also been observed in experiments [1, 3, 4].

As far as what has been discovered, HIV-1 is the only virus with a conical capsid, although similar phenomena have been observed in carbon nanocones [24].

4.1 Generating Vectors for the (5,7)-Cone

A (5,7)-cone has the smallest allowed cone angle formed from a hexagonal lattice. In this subsection, we consider generating vectors or a folding template for the (5,7)-cone.

For consistency, we consider a generating vector $h\vec{a_1} + k\vec{a_2}$ scaled by two nonequal integers to generate the triangles needed for the two end caps of the cone. Without loss of generality, we assume $\alpha < \beta$ are such two integers. Let $\vec{A} = \alpha(h\vec{a_1} + k\vec{a_2})$ generate the five smaller equilateral triangles needed to fold the 5 pentamers in the narrow end. A parallel vector $\vec{B} = \beta(h\vec{a_1} + k\vec{a_2})$ is used to generate the six larger identical triangles needed to create the 7 pentamers in the broad end.

These equilateral triangles are determined by the T-number and the two additional constants α and β. These triangle vertices have a slightly different meaning than those in the icosahedral and tubular models. In the icosahedral and tubular models, triangle vertices lie at the centers of hexagons that will be replaced by pentagons when folded in three dimensions. For the conical capsid, triangle vertices are not necessarily located in pentagons. To clarify, pentagons are shown in red in Figure 6.

However, more information is needed to form a closed (5,7)-cone. Since the generating vectors are defined on a hexagonal lattice, the model must be positioned correctly to ensure only hexagons and pentagons are produced during the folding. This occurs when the outer closure lines in the middle region of the lattice are parallel, as shown by the dotted lines in the far left and far right of Figure 6 (*left panel*). Without this requirement, the folding cannot close correctly. The parallel lines ensure hexamers are matched along the closure line, leading to a 0° declination that produces perfect hexamers in the middle region of the cone. This necessary addition to the Nguyen model [21] is further explained in the Discussion section.

To ensure parallel outer edges, a third generating vector, $\vec{C}$, is needed. Let $\vec{C} = q_0(h\vec{b_1} + k\vec{b_2})$, where $\vec{b_1}$ and $\vec{b_2}$ are the vectors defined in Equations (3)(4). Clearly, $\vec{C}$ is perpendicular to both $\vec{A}$ and $\vec{B}$.

We introduce a vector $\vec{D} = \rho_0(h\vec{a_1} + k\vec{b_2})$ to generate the distance between the two neighboring triangles in the bottom of the lattice that do not share a common vertex, as shown orange in Figure 6. This shall allow a parallelogram to be formed in the middle of the cone. The top and bottom sides of the parallelogram should have the same length, that is,

$$5\beta\sqrt{T} = 5\alpha\sqrt{T} + \rho_0\sqrt{T}, \tag{13}$$

and hence

$$\rho_0 = 5(\beta - \alpha). \tag{14}$$

Note that the vector length $|\vec{C}|$ can be determined in two ways: either from $q_0|h\vec{b_1} + k\vec{b_2}| = q_0\sqrt{3T}$ or from the sum of the height of the triangles generated by $\vec{B}$ and $\vec{D}$. Setting them equal yields

$$q_0\sqrt{3T} = \frac{\sqrt{3}}{2}\left(\beta\sqrt{T} + 5(\beta - \alpha)\sqrt{T}\right), \tag{15}$$

and hence

$$q_0 = \frac{1}{2}(6\beta - 5\alpha). \tag{16}$$

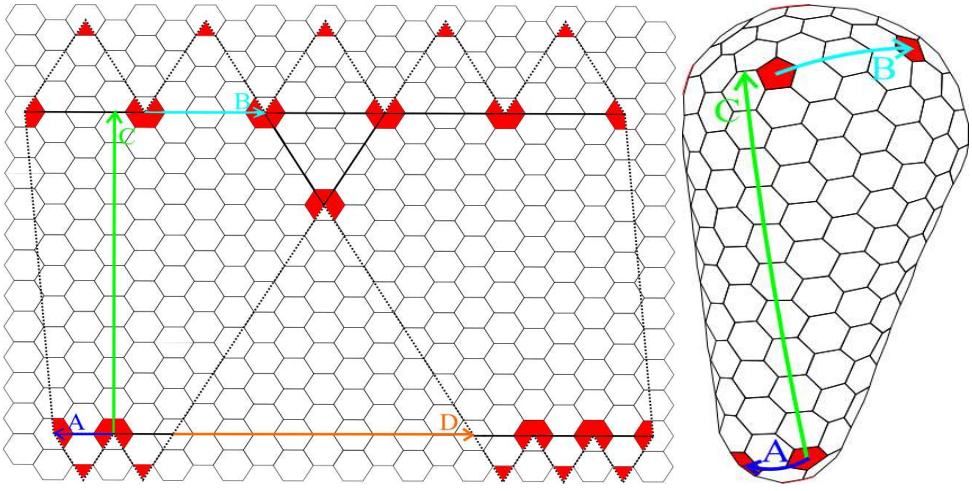

Figure 6: *Left*: Three generating vectors are needed for folding a lattice into a conical capsid: $\vec{A}$ generates the side length of the triangles in the narrow end of the cone; $\vec{B}$ generates the side length for the triangles in the broad end; $\vec{C}$ ensures the unique height needed for closure. The dotted lines indicate where to cut the lattice to fold it into a 3-dimensional cone. Pentagons are shown in red. *Right*: The lattice on the left panel with $(h, k) = (1, 1)$, $T = 3$, $(\alpha, \beta) = (1, 2)$ folds into a (5,7)-cone.

In summary, three generating vectors are needed for the folding template of a (5,7)-cone:

$$\vec{A} = \alpha(h\vec{a_1} + k\vec{a_2}), \tag{17}$$

$$\vec{B} = \beta(h\vec{a_1} + k\vec{a_2}), \tag{18}$$

$$\vec{C} = \frac{1}{2}(6\beta - 5\alpha)(h\vec{b_1} + k\vec{b_2}), \tag{19}$$

where $\alpha < \beta$ are two nonnegative integers and $\vec{b}_1$ and $\vec{b}_2$ are defined in Equations (3)(4).

Note that when $\alpha = \beta$, this cone model reproduces an icosahedron, with $T = \alpha^2\left(h^2 + hk + k^2\right)$.

Following the lattice construction, the surface area of the conical capsid is calculated by summing the areas of the three regions (the broad end, the middle region, the narrow end) on the lattice used in the folding.

$$SA = 5\left(\tfrac{\sqrt{3}}{4}|\vec{B}|^2\right) + \left(5|\vec{B}||\vec{C}| - \tfrac{\sqrt{3}}{4}|\vec{D}|^2\right) + 5\left(\tfrac{\sqrt{3}}{4}|\vec{A}|^2\right)(\text{unit}^2) =: 15T\sqrt{3}\left(2\beta^2 - \alpha^2\right)a^2\,(\text{nm}^2), \tag{20}$$

where a multiple of $\sqrt{3}a$ gives a conversion between (unit) and (nm), and a is the side length of a single hexamer in nanometers.

It is assumed that there are exactly six CA proteins in a hexamer with area $3a^2\sqrt{3}/2(\text{nm}^2)$, given the side length a. Therefore, there are $4\sqrt{3}/3a^2$ CA proteins per (nm^2). The total number of CA proteins for this conical model is given by the surface area multiplied by the number of CA proteins per area, or

$$CA = 60T\left(2\beta^2 - \alpha^2\right). \tag{21}$$

Subtracting the 60 proteins used to create the pentagons and dividing by 6 (CA proteins per hexamer) yields the total number of hexamers, N_H, on the (5,7)-conical capsid:

$$N_H = 10T\left(2\beta^2 - \alpha^2\right) - 10. \tag{22}$$

4.2 Generating Vectors for the (4,8)-Cone and Other Types Cones

Literature [2, 4, 11, 27] showed evidence that the cores of HIV-1 virus like particles (VLPs) have cone angles between 30° and 50°, which is an indication of existence of the (4,8)-conical capsid.

To maintain symmetry, the center of the narrow end should be a hexagon instead of a pentagon as seen in the previous (sub)sections. This way, four pentagons can be evenly spaced around it, The center of the narrow end is created by joining the smaller triangle tips on the lattice. For the icosahedral, tubular, and (5,7)-conical models, the narrow end has 5 triangles. When five 60° triangle tips are joined, an angular defect is produced. This results in the formation of a pentagon. However for a (4,8)-cone, six (rather than five) small triangles are used in the narrow end. When six 60° triangle tips are joined, there is no angular defect. This results in the formation of a hexagon. Similar to the (5,7)-conical model, triangle vertices are no longer guaranteed positions of pentagons when folded in three dimensions. This results in a total of 4 pentagons being created in the narrow end.

The broad end is very similar to that of the (5,7)-cone. The only difference is the addition of a 7th equilateral triangle. This allows eight pentagons to be formed in this region. Pentagon positions are shown in red in Figure 7.

Similar to the (5,7)-cone model, the (4,8)-cone model requires three generating vectors to ensure proper closure. Vector $\vec{A}$ will determine the size of the smaller triangles (in the narrow end), a parallel vector $\vec{B}$ will determine the size of the larger triangles (in the broad end), and a perpendicular vector $\vec{C}$ determines the unique height, for which the outer closure lines in the middle region of the cone are parallel.

We now explain this in detail. To enforce parallel outer edges, we define $\vec{C} = q_1(h\vec{b_1} + k\vec{b_2})$. The constant $q_1 = \frac{1}{4}(7\beta - 6\alpha)$ can be derived in a similar way to that for deriving q_0 in the (5,7)-cone model.

We use vector $\vec{D} = \rho_1(h\vec{a_1} + k\vec{b_2})$ to describe distance between two neighboring triangles that do not have a common vertex. This is similar to the vector $\vec{D}$ defined in the (5,7)-cone model. Note that in this cone model,

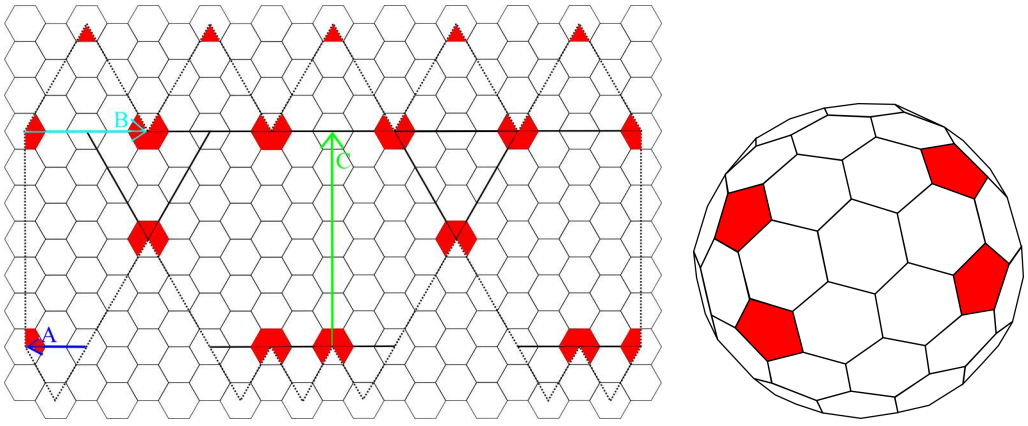

Figure 7: *Left*: Model for the (4,8)-cone using 3 generating vectors. The dotted lines indicate where to cut the lattice to fold it into a 3-dimensional cone. Pentagons are shown in red. *Right*: A view of the bottom or narrow end of the (4,8)-cone.

there are two pairs of neighboring triangles that do not share vertex, as shown in Figure 7. This shall allow a parallelogram to be formed in the middle of the cone. The top and bottom sides of the parallelogram should have the same length. Therefore, we have

$$5\beta\sqrt{T} = 6\alpha\sqrt{T} + 2\rho_1\sqrt{T}, \tag{23}$$

and hence

$$\rho_1 = \frac{1}{2}(5\beta - 6\alpha). \tag{24}$$

Similar to the (5,7)-cone model, there are two ways for expressing the length of the generating vector $\vec{C} = q_1(h\vec{b_1} + k\vec{b_2})$. Specifically, we have

$$q_1\sqrt{3T} = \frac{\sqrt{3}}{2}\left(\beta\sqrt{T} + \frac{1}{2}(5\beta - 6\alpha)\sqrt{T}\right), \tag{25}$$

and hence

$$q_1 = \frac{1}{4}(7\beta - 6\alpha). \tag{26}$$

In summary, for a (4,8)-cone, the folding template is determined by two pairs of integers (h, k) and (α, β) (β needs to be strictly even), and three generating vectors:

$$\vec{A} = \alpha(h\vec{a_1} + k\vec{a_2}), \tag{27}$$

$$\vec{B} = \beta(h\vec{a_1} + k\vec{a_2}), \tag{28}$$

$$\vec{C} = \frac{1}{4}(7\beta - 6\alpha)(h\vec{b_1} + k\vec{b_2}), \tag{29}$$

where $\vec{b}_1$ and $\vec{b}_2$ are defined in Equations (3)(4).

Based on the above (4,8)-cone model, we can derive the surface area (SA), the number of proteins (CA), and the number of hexamers (N_H) as follows.

$$SA = 5\left(\tfrac{\sqrt{3}}{4}|\vec{B}|^2\right) + \left(5|\vec{B}||\vec{C}| - \tfrac{\sqrt{3}}{2}|\vec{D}|^2\right) + 6\left(\tfrac{\sqrt{3}}{4}|\vec{A}|^2\right)(\text{unit}^2) = \tfrac{3\sqrt{3}}{8}T\left(55\beta^2 - 24\alpha^2\right)a^2\ (\text{nm}^2), \tag{30}$$

$$CA = \frac{3T}{2}\left(55\beta^2 - 24\alpha^2\right), \tag{31}$$

$$N_H = \frac{T}{4}(55\beta^2 - 24\alpha^2) - 10. \tag{32}$$

Folding Templates for the (3,9)-, (2,10)-, (1,11)-Cones. According to the Euler theorem, there are 3 other possible cone angles for a hexagonal lattice, corresponding to a narrow end with 3, 2, or 1 pentagon(s). The

$1P$ and $2P$ cones would induce higher strain due to their nonspherical shapes, thus are unlikely to form in nature [13]. The $3P$ cone, with a cone angle 60°, is the preferred cone angle for the helical cone for graphite [28]. Since these cones overlap, they would not follow the same construction rules as the isometric models investigated in this paper. Therefore, it is unnecessary to construct isometric cones for the remaining cone angles.

5 Comparison of Modeling Results to HIV-1 Data

The formulas for the surface area, number of CA proteins, and number of hexamers on the (5,7)-cone have already been established in Equations (20-22). Other common measurements for the (5,7)-cone such as the broad end diameter D_b, the narrow end diameter D_a, and the overall height H, are respectively

$$D_b = \frac{5\beta\sqrt{3T}}{\pi} a \text{ (nm)}, \tag{33}$$

$$D_a = \frac{5\alpha\sqrt{3T}}{\pi} a \text{ (nm)}, \tag{34}$$

$$H = \frac{5\sqrt{3T}}{2\pi}\left(\frac{\beta-\alpha}{\tan\theta} + (\beta+\alpha)\right) a \text{ (nm)}, \tag{35}$$

where a is the side length of a hexamer, α and β are the constants associated respectively with the scaling of $\vec{A}$ and $\vec{B}$, T is the triangulation number, and $\theta = 19.2^\circ/2$.

Equations (33)(34) can be derived from the lattice construction. Note that the circumference of the broad end is given by $5|\vec{B}| = 5\beta\sqrt{3T}a$ (nm). Similarly, for the narrow end, the circumference is $5|\vec{A}| = 5\alpha\sqrt{3T}a$ (nm).

Equation (35) follows from summing the height of the cone with the radii of both hemispheres. We use the commonly known equation for the opening angle of a right cone, that is, $\phi = 2\arctan\left(\frac{r}{h}\right)$, where r is the radius and h is the height. This leads to $H = \frac{1}{2}\left(\frac{D_b - D_a}{\tan\theta}\right) + \frac{D_b}{2} + \frac{D_a}{2}$, where $2\theta = \phi$.

Next we compare modeling results with the experimental data on HIV-1 VLPs reported in [1, 4].

Comparison with Data in [4]. In [4], it is found that most HIV-1 VLPs cores exhibit a conical shape with an average cone angle of $22.3^\circ \pm 6^\circ$, although about 7% VLPs exhibit tubular morphology and few show amorphous morphology. Measurements for the overall height, the broad end diameter, and the cone angle were performed for 267 conical cores. Among the conical capsids with a single core, [4] found that the hexamer diameter is 9.8(nm) with a 3.2(nm) spacing between repeating hexameric/pentameric units.

The spacing in the assumption for the hexagon size should be taken into consideration. This implies that a has a value $(9.8 + 3.2)/2 = 6.5$(nm). Taking into account of how small the radius is found to be in this region compared to the length of a [2], it is assumed that the pentagons are grouped closely together yet still isolated. To explain this tight grouping, we take $(h, k) = (1, 1)$ for our model defined by Equations (17-19). Similar construction of the narrow end can be found in [3, 11, 23, 26]. We then take $(\alpha, \beta) = (1, 2)$. $\alpha = 1$ also follows from the narrow end described in [3, 11, 23, 26], while $\beta = 2$ is determined from the size of the broad-end diameter in [4].

With this value for a, the model for the (5,7)-cone with $(h, k) = (1, 1)$ and $(\alpha, \beta) = (1, 2)$ produces a conical capsid with properties listed in the 3rd column of Table 1.

Table 1: Comparison of modeling results with experimental data in [4]

	Experimental data	Modeling results
Cone angle	$22.3^\circ \pm 6^\circ$	19.2°
Cone overall height	119.3(nm) ± 11(nm)	134(nm)
Broad-end diameter	60.7(nm) ± 8(nm)	62(nm)

Comparison with Data in [1]. Among the experimental data on 26 HIV-1 VLPs reported in [1] are

- 16 VLPs exhibit conical morphology;
- 3 VLPs have tubular morphology;
- The rest have irregular shape.

Among the VLPs with conical morphology, the measurements are found in the 2nd column of Table 2.

Table 2: Comparison of modeling results with experimental data in [1]

	Experimental data	Modeling results
Mean angle	20.1°	19.2°
Mean height	143(nm) (standard deviation 10.8(nm))	127(nm)
Mean surface area	21,000 (nm^2) (standard deviation 9000(nm^2))	20,612(nm^2)
Number of hexamers	206 hexamers for each capsid (1300 CA monomers)	200

To compare the experimental data in [1] to the theoretical results derived from the models proposed in this paper, we only require the side length of the hexamers and pentamers. [1] reported the surface area of the capsid with an estimated 206 hexamers. Assuming the hexamers and pentamers have the same side length a, one reaches an estimate $a = 6.1465$ (nm) for the given surface area. This value is similar to the a value found from the data in [4]. So we choose the same parameter values for our model described in Equations (17-19). Then the model for the (5,7)-cone with $(h, k) = (1, 1)$ and $(\alpha, \beta) = (1, 2)$ yields the results shown in the 3rd column of Table 2.

The comparison with the data in these two papers demonstrate good agreement of our modeling results with experimental data.

6 Discussion

Flaw in the Nguyen Model: In [21], Nguyen et al. used two generating vectors for the lattice structure of the HIV-1 (5,7) conical capsid. The two generating vectors are

$$\vec{A} = n(h\vec{a_1} + k\vec{a_2}), \tag{36}$$

$$\vec{B} = m(h\vec{a_1} + k\vec{a_2}), \tag{37}$$

where h, k are two nonnegative integers used to determine the T-number and n, m are two other nonnegative integers similar to those in our model. However, an important piece of information is missing from this model: the unique height required for closure of the cone. The required height allows the two outer middle closure lines to be parallel, which in return match the cut hexagons to form into a pentagon. If these lines are not parallel, then either a more than 60° declination (shown in yellow near the top of Figure 8 (*left panel*)) or a less than 60° declination (shown in yellow near the bottom of Figure 8 (*left panel*)) will be produced. Both cases result in incomplete pentagons formed on the capsid. Shown in Figure 8 is an example of failure in closure.

The geometric models proposed in this paper are useful for investigating the discrete curvatures and curvature concentrations on the HIV-1 conical capsids [16]. It is suggested in [6] that the asymmetry and quasi-equivalence exhibiting in tubular and conical capsids are related to the hinge between the C-terminal domain (CTD) and N-terminal domain (NTD) of the capsid protein. This information could be utilized to study the

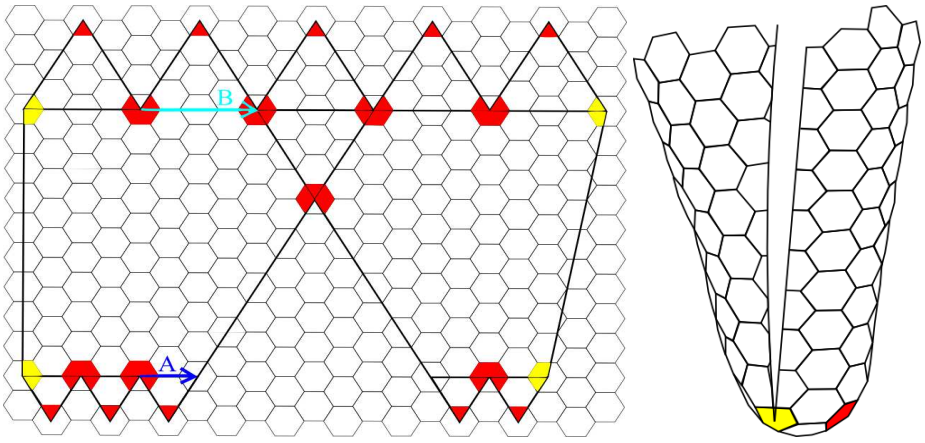

Figure 8: *Left*: An illustration of a lattice with $(h, k) = (1, 1)$, $T = 3$, $(n, m) = (1, 2)$ according to the Nguyen model with **only two generating vectors** [21], where the two generating vectors $\vec{A}$ and $\vec{B}$ are defined in Eqns (36-37). Pentamer positions are shown in red. Incomplete pentagons are shown in yellow. *Right*: The lattice folds into an incomplete (5,7)-cone, the partial hexagons along the outer edges do not match, since the unique height is not enforced in the Nguyen model.

elastic energy on the capsid and the relationship between curvature and elastic energy. This is under our investigation and will be reported in our future work.

It is also quite interesting to see that viral capsids and other protein cages can be used as containers for polymers and nano-particles to make new synthetical materials [9, 10, 15, 22].

References

[1] J. Benjamin, B.K. Ganser-Pornillos, W.F. Tivol, W.I. Sundquist, and G.J. Jensen, Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. J. Mol. Biol. **346**(2005), 577–588.

[2] J. Briggs, K. Grünewald, B. Glass, F. Förster, H-G. Kräusslich, and S.D. Fuller, The mechanism of HIV-1 core assembly: Insights from three-dimensional reconstructions of authentic virions. Structure **14**(2006), 15–20.

[3] J. Briggs and H. Krausslich, The molecular architecture of HIV. J. Mol. Biol. **410**(2011), 491–500.

[4] J. Briggs, T. Wilk, R. Welker, H-G. Kräusslich, and S.D. Fuller, Structural organization of authentic, mature HIV-1 virions and cores. EMBO **22**(2003), 1707–1715.

[5] C. Büchen-Osmond, ICTVdB - The universal virus database, Version 4, ICTVdB Management, New York, 2006.

[6] I.L. Byeon, X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, and A.M. Gronenborn, Structural convergence between cryo-EM and NMR reveals intersubunit interations critical for HIV-1 capsid function. Cell **139**(2009), 780–790.

[7] D. Caspar and A. Klug, Physical principles in the construction of regular viruses. Cold Spring Harb. Symp. Quant. Biol. **27**(1962), 1–24.

[8] T. Douglas and M. Young, Virus: Making friends with old foes. Science **312**(2006), 873–875.

[9] O.M. Elrad and M.F. Hagan, Encapsulation of a polymer by an icosahedral virus. Phys. Biol. **7**(2010), 045003.

[10] A. de la Escosura, R. Nolte, and J. Cornelissen, Viruses and protein cages as nanocontainers and nanoreactors. J. Mater. Chem. **19**(2009), 2274–2278.

[11] B.K. Ganser, S. Li, V.Y. Klishko, J.T. Finch, and W.I. Sundquist, Assembly and analysis of conical models for the HIV-1 core. Science **80**(1999), 80–83.

[12] B.K. Ganser-Pornillos, M. Yeager, and O. Pornillos, Assembly and architecture of HIV. Adv. Exp. Med. Biol. **726**(2012), 441–465.

[13] Y. Gogotsi, Nanomaterial Handbook, Taylor & Francis Group, Florida, 2006.

[14] J. Heymann, C. Butan, D. Winkler, R. Craven, and A. Steven, Irregular and semi-regular polyhedral models for Rous sarcoma virus cores. Comput. Math. Meth. Med. **9**(2008), 197–210.

[15] Y. Hu, R. Zandi, A. Anavitarte, C.M. Knobler, and W.M. Gelbart, Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size. Biophys. J. **94**(2008), 1428–1436.

[16] J. Liu, F. Sadre-Marandi, S. Tavener, and C. Chen, Curvature concentrations on the HIV-1 capsid. Preprint, Colorado State University, (2014).

[17] A. Luque and D. Reguera, The structure of elongated viral capsids. Biophys. J. **98**(2010), 2993–3003.
[18] E.R. May, J. Feng, and C.L. Brooks III, Exploring the symmetry and mechanism of virus capsid maturation via an ensemble of pathways. Biophys. J. **102**(2012), 606–612.
[19] M.F. Moody, The shape of the T-even bacteriophage head. Virology **26**(1965), 567–576.
[20] M.F. Moody, Geometry of phage head construction. J. Mol. Biol. **293**(1999), 401–433.
[21] T.T. Nguyen, R.F. Bruinsma, and W.M. Gelbart, Elasticity theory and shape transitions of viral shells. Phys. Rev. E **72**(2005), 1–19.
[22] J.K. Pokorski and N.F. Steinmetz, The art of engineering viral nanoparticles. Mol. Pharma. **8**(2011), 29–43.
[23] O. Pornillos, B.K. Ganser-Pornillos, and M. Yeager, Atomic-level modelling of the HIV capsid. Nature (Letter) **469**(2011), 424–428.
[24] A. Siber, Continuum and all-atom description of the energetics of graphene nanocones. Nanotech. **18**(2007), 1–6.
[25] Y. Tao, N.H. Olson, W. Xu, D.L. Anderson, M.G. Rossmann, and T.S. Baker, Assembly of a tailed bacterial virus and its genome release studied in there dimensions. Cell **95**(1998), 431–437.
[26] B. Turner and M. Summers, Structural biology of HIV. J. Mol. Biol. **285**(1999), 1–32.
[27] R. Welker, H. Hohenberg, U. Tessmer, C. Huckhagel, and H-G. Kräusslich, Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1. J. Virol. **74**(2000), 1168–1177.
[28] F.F. Xu, Y. Bando, and D. Golberg, The tubular conical helix of graphitic boron nitride. New J. Phys. **5**(2003), 1–16.
[29] R. Zandi, D. Reguera, R. Bruinsma, W. Gelbart, and J. Rudnick, Origin of icosahedral symmetry in viruses. PNAS **101**(2004), 15556–15560.
[30] G. Zhao, J. Perilla, E. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. Gronenborn, K. Schulten, C. Aiken, and P. Zhang, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature (Letter) **497**(2013), 642–646.
[31] www.thebacteriophages.org/chapters/0180_figure_003.htm