Research Article Open Access

Farrah Sadre-Marandi* and Praachi Das

Extension of Caspar-Klug theory to higher order pentagonal polyhedra

https://doi.org/10.1515/cmb-2018-0001
Received July 4, 2017; accepted February 15, 2018

Abstract: Many viral capsids follow an icosahedral fullerene-like structure, creating a caged polyhedral arrangement built entirely from hexagons and pentagons. Viral capsids consist of capsid proteins, which group into clusters of six (hexamers) or five (pentamers). Although the number of hexamers per capsid varies depending on the capsid size, Caspar-Klug Theory dictates there are exactly twelve pentamers needed to form a closed capsid. However, for a significant number of viruses, including viruses of the Papovaviridae family, the theory doesn't apply. The anomaly of the Caspar-Klug Theory has raised a new question:"For which Caspar and Klug models can each hexamer be replaced with a pentamer while still following icosahedral symmetry?" This paper proposes an answer to this question by examining icosahedral viral capsid-like structures composed only of pentamers, called pentagonal polyhedra. The analysis shows that pentagonal polyhedra fall in a subclass of T, defined by $P \geq 7$ and $T = 1 \pmod 3$.

Keywords: CA protein; capsid; icosahedral symmetry; pentamer; T-number

1 Introduction

A virus is an infectious agent that invades and infects a host through the release and replication of its genome within the host cell. Advances in virology and the design of antiviral therapeutics rely strongly on understanding the viral replication cycle, the structures of the capsids, and the mechanisms that trigger their assembly and disassembly. Understanding the structure of capsids is of utmost importance not only for the design of new antiviral drugs, but also for many other applications in nanotechnology [7, 8, 11, 24].

All viruses contain at least two components: a nucleic acid genome and a protective protein shell, called the capsid. The genome of a virus is either single or double stranded DNA or RNA. This genetic material is covered by a capsid, which is made up of protein subunits called capsomeres. Depending on the overall arrangement of these capsomeres, capsids have icosahedral, helical or complex symmetry [6, 17, 23]. Irrespective of their shape and size, the underlying theme is that the capsid structure is designed to contain and protect the viral genome. The capsid interacts with the host cell during the infection process and delivers the genome to the cell for replication of the virus.

Crick and Watson observed that the size of the packaged genomic material is too small to encode for more than a limited number of different capsid proteins and argued that icosahedral symmetry constraints on viral capsid architecture are crucial to ensure the size economy of the viral genome [5]. It was later experimentally shown that over half of known viruses follow this icosahedral symmetry. The basic principles defining the formation of icosahedral viral capsids were outlined by Caspar and Klug [4] extending mathematical knowledge to biological structures. The Caspar-Klug Theory is now a major tool in modern virology, geometrically describing the locations of protein subunits in icosahedral capsids.

***Corresponding Author: Farrah Sadre-Marandi:** Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA, E-mail: sadre-marandi.1@mbi.osu.edu
Praachi Das: Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA, E-mail: das.156@osu.edu

Caspar-Klug Theory is a fundamental concept in virology with wide-ranging applications, such as image analysis of experimental data and construction of assembly models. However, for a significant number of viruses, including viruses of the Papovaviridae family, the theory doesn't apply [2]. The Caspar-Klug Theory correctly predicts the location of the capsid units, but predicts 12 pentamers (12 groups of five proteins) and 60 hexamers (60 groups of 6 proteins) rather than the known 72 pentamers. This structural puzzle has since been solved by the tiling approach presented in [27]. This anomaly of the Caspar-Klug Theory has raised a new question:"For which Caspar and Klug models can each hexamer be replaced with a pentamer while still following icosahedral symmetry?"

The rest of this paper is organized as follows. Section 2 reviews the concepts behind the geometry of viral capsids outlined by the Caspar-Klug Theory and the folding of plane nets (interchangeably referred to as a lattice) to form icosahedral capsids. In Section 3 we discuss the icosahedral exceptions to the Caspar-Klug Theory by examining the properties of pentagonal polyhedra, which are Caspar-Klug structures recreated consisting of only pentamers. Section 4 concludes the paper with remarks.

2 Geometry of icosahedral capsids

The icosahedral viral capsids have been extensively studied due to their highly symmetric nature and occurrence. It has been confirmed experimentally that more than half of the virus species are found to exhibit icosahedral capsids and their capsid structural organization are outlined by particular types of surface lattices [17]. Caspar and Klug established a theory that defines a family of surface lattices that act as a blueprint for describing this geometric structure of the icosahedral viral capsids [4].

2.1 Caspar-Klug Theory

The Caspar-Klug Theory (CK Theory) is built upon 60 identical subunits organized on the 20 triangles creating the faces of an icosahedron. These icosahedral structures exhibit rotational symmetry: 5-fold symmetry at the vertices, 2-fold through the edges and 3-fold through the center of each triangular face. Each triangular facet can be divided into three symmetrically equivalent parts, called an icosahedral asymmetric unit (IAU), by the 3-fold symmetry axis.

Quasi-Equivalence. The simplest icosahedral structure has 60 chemically identical subunits, one per IAU, interacting identically with the same environment around it. CK Theory indicated that when more than 60 subunits interact, which appears to be true for many spherical viruses, all subunits cannot have identical environments yet can still follow icosahedral symmetry with slight but regular changes in their bonding. When putting $60N$ units on the surface of a sphere, the highest order structure would arrange them in sets of N units each. Yet the members within the sets will no longer be equivalently related. This lack of equivalence between the subunits is solved by the theory of quasi-equivalence.

Crick and Watson proposed the physical principle that the same contacts between subunits are used repeatedly [5]. The assumption is that the capsid is held together by the same type of bond, though these bonds will be slightly deformed to correspond to non-symmetrically related environments. Quasi-equivalence, then, allows for the mathematical constraint of strict equivalence to be discarded while retaining the physical essentials necessary to describe the bonding on identical units. This ensures the capsomeres are arranged in a minimum energy configuration realized by quasi-symmetry packing arrangements. For example, a virus with 240 capsomeres would contain four (240/60) types of bonding and four distinct environments for the subunits. The four types would not be equivalent but quasi-equivalent because their environments are similar but not identical.

***T*-Number.** The CK Theory enumerates the possible designs for icosahedral surface lattices by mapping the unfolded 20 triangular faces of an icosahedron onto a two dimensional hexagonal lattice (shown Figure 3 left). It is assumed this lattice consists entirely of regular hexagons. Each vertex from the triangulation should

fall in the center of a hexagon. Generating a closed structure requires curving the net and converting the hexagons chosen by the triangulation vertices into pentagons. Complying with icosahedral symmetry, each pentagon must be equidistant apart. The vector between a neighboring pair of pentagons must be a lattice vector of the triangular net. This lattice vector is generated by a linear combination of the two basis vectors for the hexagonal lattice

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

where h and k are nonnegative numbers (but not both zero), shown in Figure 1. Since the pentagons are equidistant, the (h, k) pair define all possible icosahedral surface lattice designs.

The triangulation number (T-number) is mathematically defined as the squared length of each triangle edge, with full proof explained in [26]. It counts the number of symmetrically distinct but quasi-equivalent triangular facets in the triangulation per face of icosahedron.

$$|\vec{A}|^2 = h^2 + hk + k^2 =: T. \tag{2}$$

For the case of $T = 1$, the icosahedron has 60 equivalent units sorted into twelve clusters of five proteins (pentamers) with identical bonds. The larger icosahedra have groups of six proteins (hexamers) between the pentamers and can be visualized as replacing the original triangular facets with larger equilateral triangles. The number of triangle facets per face replacing the original one is the triangulation number. Each triangulation number has a different organization of bonding, as defined by quasi-equivalence. This theory derives a family of surface lattices, parameterized by T, corresponding to all possible planar embeddings of an icosahedral surface onto a hexagonal grid.

The T-number refers to the organization of the geometric figure, not necessarily the structural components of an individual virus. Quasi-equivalence specifies that the number of different environments should equal the T-number. The exceptions to this rule are discussed in Section 3. When the correspondence is direct, the T-number provides a strong descriptive tool for determining the number of subunits per viral capsid.

Euler's Formula. Although the number of hexamers per capsid varies depending on the capsid size, CK Theory dictates there are exactly twelve pentamers needed to form a closed capsid. This follows from the Euler-Poincare characteristic, χ. For a polyhedron,

$$\chi = f + v - e = 2, \tag{3}$$

where v, e, and f are the number of vertices, edges, and faces respectively.

In a polyhedron composed only of $Pent$ pentagons and Hex hexagons,

$$f = Pent + Hex. \tag{4}$$

Counting each edge while traversing each face yields

$$2e = 5Pent + 6Hex. \tag{5}$$

Three polygons must come together at each vertex. The only options are three pentagons, two pentagons and one hexagon, or one pentagon and two hexagons. In each case there are three edges terminating a vertex. Counting each edge while traversing each vertex yields

$$2e = 3v. \tag{6}$$

It follows that

$$\chi = f + v - e = (Pent + Hex) + \left(\frac{5Pent + 6Hex}{3}\right) - \left(\frac{5Pent + 6Hex}{2}\right) = 2 \Rightarrow Pent = 12. \tag{7}$$

With each capsid containing exactly 12 pentamers and $60T$ protein subunits, the number of hexamers (Hex) also relies on the T-number:

$$Hex = 10(T - 1). \tag{8}$$

2.2 Folding of Plane Nets

Beginning with a flat hexagonal lattice consisting entirely of regular hexagons, a closed structure can be generated by curving the net and converting exactly 12 hexagons to pentagons. The pentagon locations are defined by the vertices of the triangulation with a generating vector defined in Equation 1. The pentagons are formed by cutting a 60° wedge from a hexagon and rejoining the cut edges such that a convex five-sided polygon is formed, whose center no longer lies on the hexagonal plane, as shown in Figure 1 right. This is equivalent to joining five one-sixth equal parts of a hexagon.

Algorithm for CK Folding.

1. Start with a flat hexagonal lattice;
2. Triangulate 20 equilateral triangles with side length $\sqrt{T}$ such that all vertices lie in the center of a hexagon;
3. Join faces along adjacent edges, such that the three-dimensional triangulation vertices lie in the center of a pentagon and the interior angle defect of each pentagon is less than 180°; All vertices of the polyhedron will point outwards, away from the interior of the shape;
4. Join the two sides so that their vertices of intersection lies in the center of pentagons.

Convexity. It is possible to create multiple non-convex icosahedra with the 20 triangular net, where up to three vertices lie in the center of the polyhedron rather than projected onto the surface of a sphere. In this case, a interior angle defect of at least one pentagon would be greater than 180°. On the other hand, there is a unique convex folding since the curvature along the net is given only by the pentamers [16] and a folding of this type would require each pentagon to have an interior angle defect of less than 180°.

Dual Model. The convex triangulation folding gives descriptive information on the location of the pentamers and the number of hexamers between them on the triangular faces. Though, this folding does not give an accurate visual of viral capsids since it implies the pentamers are convex (Figure 1 right). An alternative visualization of viral capsids is given by the pentagonal dodecahedron, where the pentamers and hexamers are flat and the curvature is formed along the edges of the hexamers and pentamers, rather than the edges of the triangulation (Figure 3 right).

The triangular icosahedron has 12 pentagonal vertices and 20 triangular faces. Each triangle has a 3-fold axis in its center of its face, a 2-fold axis at the center of its edge, and a 5-fold axis through the 12 vertices. The related dual structure, the pentagonal dodecahedron, shares the same symmetry elements but has a complementary morphology: 5-fold axes through the pentagonal faces, 3-fold through the vertices, and 2-fold through the edges. Both of these shapes have icosahedral symmetry, as do spherical viruses, whose shape lies between these two extremes. The icosahedron can also be represented as a spherical tiling and projected onto the plane via a stereographic projection. This projection is conformal, preserving angles but not areas or lengths. Straight lines on the sphere are projected as circular arcs on the plane.

Previous models have used the icosahedron to represent viral capsid structures with an assumed equivalence to the pentagonal dodecahedron and spherical tiling. The distinction between spherical or polyhedral shape is of no fundamental importance [4], each model correctly calculates the position and number of hexagons and pentagons per (h, k) pair, though the measurement approximations are dependent on the model chosen.

For example, assuming the icosahedral folding model the surface area can be calculated by summing the area of the 20 triangles

$$SA_I = 20\left(\frac{\sqrt{3}}{4}|\vec{A}|^2\right) := 15\sqrt{3}Ta^2(nm^2), \tag{9}$$

where a multiple of $\sqrt{3}a$ gives the conversion between lattice and nanometer (nm), and a is the length of a protein [26]. A similar calculation will show that the surface area formula for the dodecahedron is

$$SA_D := 12\left(a^2\frac{\sqrt{5(5+2\sqrt{5})}}{4}\right) + 10(T-1)\left(\frac{3a^2\sqrt{3}}{2}\right)(nm^2). \tag{10}$$

It follows that $SA_I > SA_D$ for all (h, k) and a. The difference between SA_I and SA_D is exactly twelve times the difference between the areas of the pentamers in each model. For SA_I, the pentamer area is obtained by subtracting one-sixth of the area of the hexagon, while SA_D uses the area of a regular pentamer. It is important to note that similar discrepancies can occur with other common measurement formulas.

Chirality. The smallest icosahedron has 20 equilateral triangular faces defined by (1,0) or equivalently (0,1). Larger icosahedral viruses follow the icosadeltahedral structure with $20T$ facets, where T is the triangulation number. The convex folding is unique for each pair (h, k) but not unique for T. For example, $(7, 0)$ and $(5, 3)$ both give $T = 49$ but produce different capsid organizations. It should be noted that although $(7, 0)$ and $(0, 7)$ produce identical capsids, $(5, 3)$ and $(3, 5)$ produce enantiomorphous structures. This property is further explained with the addition of the P class, which is a subset of T, summarized below and described in detail in [4, 17, 25].

Let $f = \gcd(h, k)$. Then $h = h_0 f$, $k = k_0 f$, and

$$\begin{aligned} T &= h^2 + hk + k^2 && (11)\\ &= h_0^2 f^2 + h_0 k_0 f^2 + k_0^2 f^2 && (12)\\ &= (h_0^2 + h_0 k_0 + k_0^2) f^2 && (13)\\ &= P f^2 && (14) \end{aligned}$$

for

$$P = h_0^2 + h_0 k_0 + k_0^2. \tag{15}$$

All icosahedra can be described by one of three classes based on their P-number: $P = 1$, $P = 3$, or $P \geq 7$. The P class and lattice parameters h and k are also used to define the chirality, which is defined on the lattice as the angle between the basis vector $\vec{a_1}$ and the generating vector $\vec{A}$, illustrated in Figure 2. This produces a right triangle with base length $(h + k/2)$, height $(k\sqrt{3}/2)$, and hypotenuse $(\sqrt{T})$.

The chiral angle θ can be defined as

$$\cos\theta = \frac{2h + k}{2\sqrt{T}}, \quad \sin\theta = \frac{k\sqrt{3}}{2\sqrt{T}}, \quad \tan\theta = \frac{k\sqrt{3}}{2h + k}. \tag{16}$$

It is easy to see for the first two P classes, chirality is not formed. For $P = 1$, the T-numbers are based along the h-axis or k-axis ($T = h^2 = 1, 4, 9$, etc...), with $h > 0, k = 0$ or $h = 0, k > 0$ respectively. These icosahedra have lattice lines that run parallel to the edges of the triangular facet, which can be considered a higher order icosahedron, as shown in Figure 4 $(h, 0) = (0, h)$ (blue). If $h > 0$ and $k = 0$ then $\theta = 0°$ and the lattice lines correspond to the h-axis. Similarly if $h = 0$ and $k > 0$ then $\theta = 60°$ and the lattice lines correspond to the k-axis. Capsid structures with $\theta = 0°$ or $60°$ have hexamers arranged as rings, analogous to the zigzag configuration of carbon nanotubes (CNTs), shown in Figure 4 middle left.

When $P = 3$, $h = k$ and the T-numbers are based along the line that bisects the h- and k-axes ($T = 3h^2 = 3, 12, 27$, etc...). This yields lattice lines that bisect the angles between the edges of the icosahedral facet, rotating the facets by $\theta = 30°$, as shown in Figure 4 (h, h) (cyan). This class has a zigzag arrangement of hexamers analogous to the armchair configuration of CNTs [17], shown in Figure 4 top left. $P = 1$ and $P = 3$ are the only classes which produce icosahedra with planes of symmetry.

Polyhedra with $P \geq 7$ are skewed, including all other allowed T-numbers $(7, 13, 19, 21$, etc...). The lattice lines are not symmetrically disposed with respect to the icosahedral facet and do not produce uniquely defined T numbers. This third case describes the class in which (h, k) and (k, h) produce enantiomorphous capsids, existing in right- and left-handed forms, as shown in Figure 3. For $h > k > 0$ the chiral angle falls between the h-axis and the line that bisects the h– and k– axes producing a left-handed (laevo) form with $0° < \theta < 30°$, Figure 3 top. Alternatively, for $0 < h < k$, the angle falls between the bisecting line $h = k$ and the k–axis producing a right-handed (dextro) form $30° < \theta < 60°$, Figure 3 bottom. An example of a virus structure with a skewed icosahedral lattice is the widely studied bacteriophage HK97 ($T = 7d$) [19]. The properties of these three classes of P-numbers are summarized in Table 1.

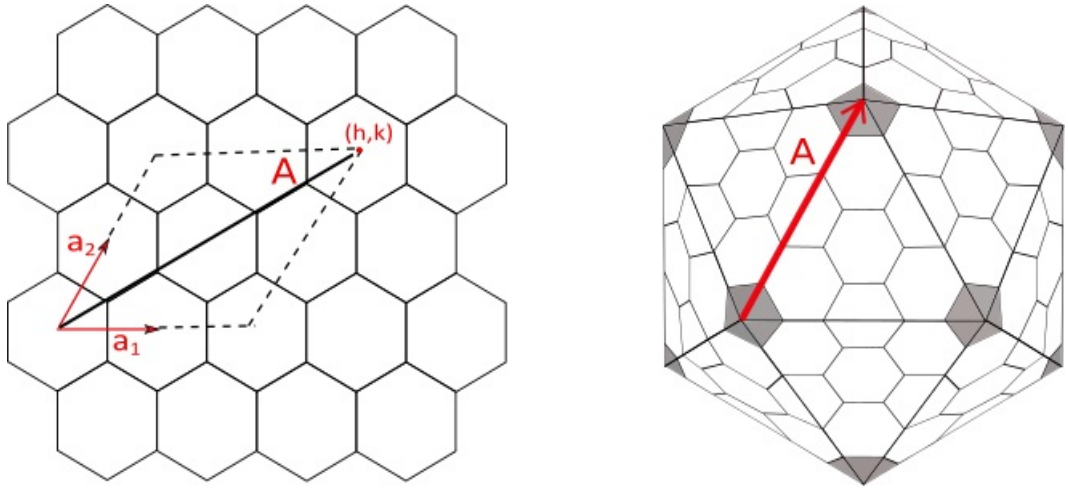

Figure 1: Two basis vectors $\vec{a}_1, \vec{a}_2$ and one generating vector $\vec{A}$ for a hexagonal lattice. In this illustration, $h = 2, k = 2$ and hence $T = h^2 + hk + k^2 = 12$. The starting and ending points of the generating vector $\vec{A}$ denote the two hexagons replaced by pentagons when the lattice is folded into an icosahedron.

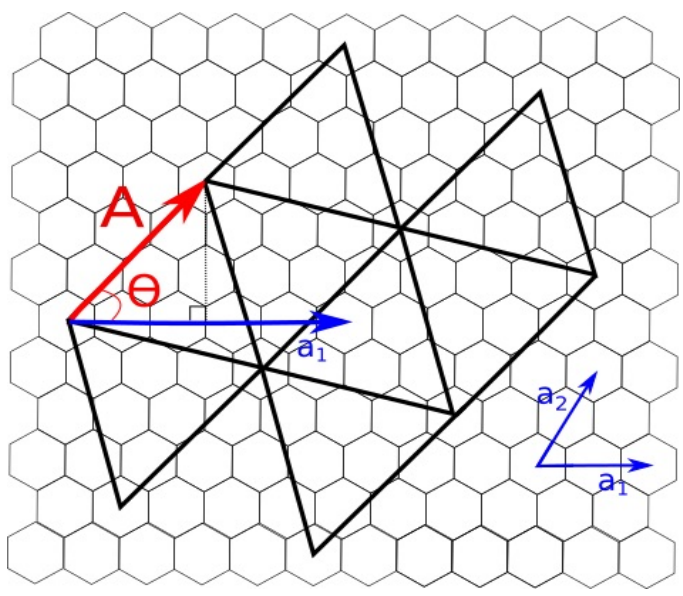

Figure 2: Chiral angle is defined as the angle between the basis vector $\vec{a}_1$ (blue) and the generating vector $\vec{A}$ (red). This produces a right triangle with base length $(h + k/2)$, height $(k\sqrt{3}/2)$ and hypotenuse $(\sqrt{T})$. Shown for $(h, k) = (1, 3)$ and $T = h^2 + hk + k^2 = 13$.

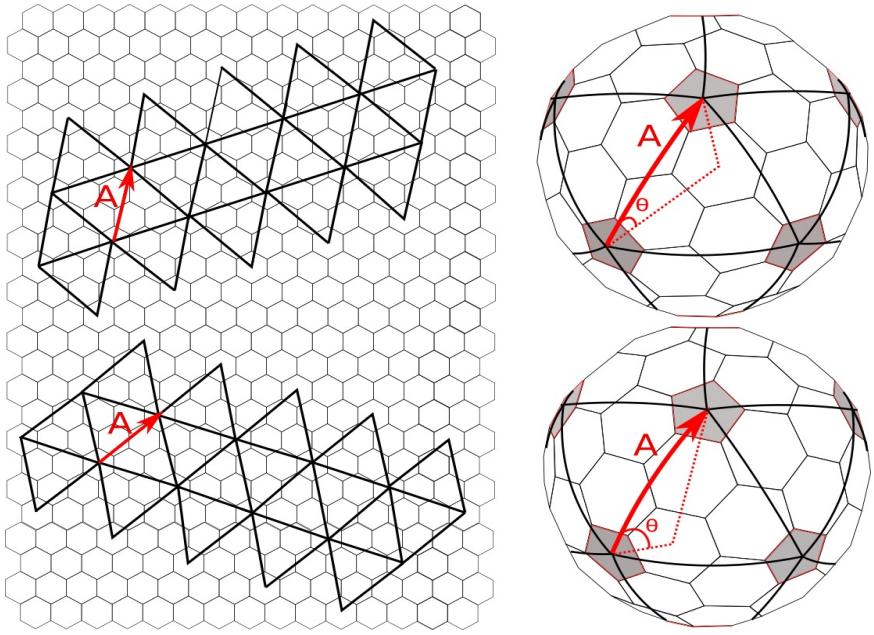

Figure 3: *Top*: A laevo lattice $(h > k)$ and folding with $(h, k) = (2, 1)$ and $T = h^2 + hk + k^2 = 7$. *Bottom*: A dextro lattice $(k > h)$ and folding with $(h, k) = (1, 2)$ and $T = h^2 + hk + k^2 = 7$. The red dotted lines indicate the chiral angle. Pentagons are shown in gray.

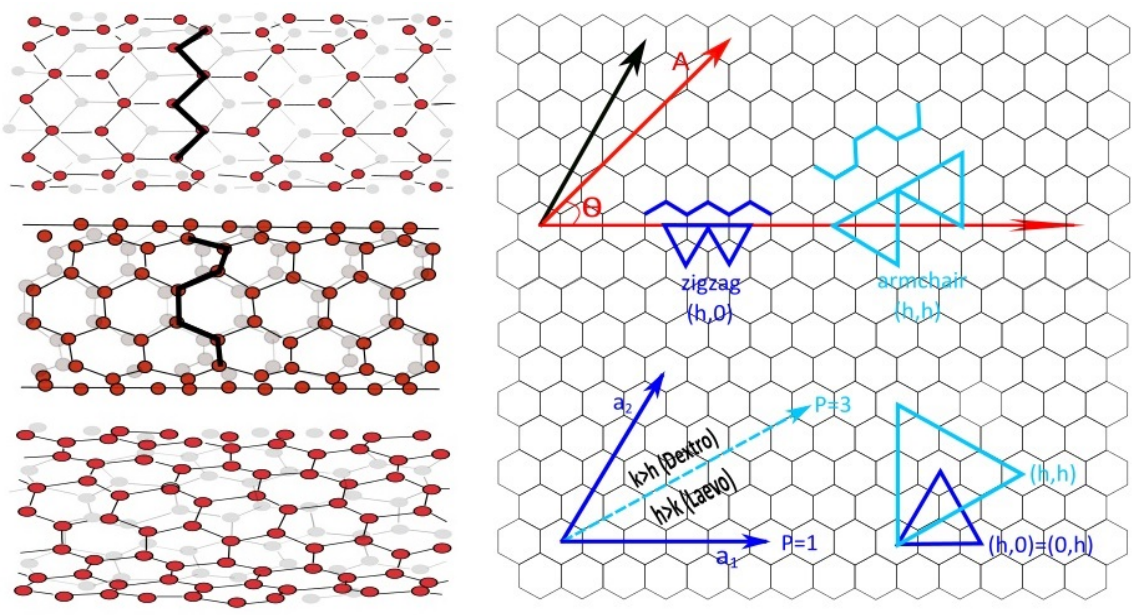

Figure 4: *Left*: Illustrations of chirality on hexagonal tubes. From top to bottom: zigzag, armchair, chiral; *Right*: Illustrations of chirality and classes of P-number on a hexagonal lattice. Top: Zigzag arrangement of constituent equilateral triangles seen for (h,0) and $\theta = 0°$ $(P = 1)$ along h-axis. Armchair arrangement seen for (h,h) with $\theta = 30°$ $(P = 3)$. Bottom left: Location of (h, k) pair based on P-number. Chiral arrangement observed for $P \neq 1$ and $P \neq 3$. Laevo configuration seen for $h > k$ where $0° < \theta < 30°$ and dextro configuration for $k > h$ where $30° < \theta < 60°$. Bottom right: Triangular facets for $P = 1$ (blue) and $P = 3$ (cyan).

Table 1: Summary of icosahedra classified by their P-number. $P = 1$ and $P = 3$ are classes which produce icosahedra with planes of symmetry and chirality is not formed. Polyhedra with $P \geq 7$ are skewed, existing in right- and left-handed forms.

Class	P-number	Generating Vector	T-number
Class I.	$P = 1$,	$(h, 0)$ or $(0, k)$,	$T = h^2$ or $T = k^2$
Class II.	$P = 3$,	(h, h),	$T = 3h^2$
Class III.	$P \geq 7$,	(h, k),	$T = f^2 P$

3 Pentagonal Polyhedra

For folding the surface of a convex polyhedra, there are two types of plane nets without mirror symmetry that maintain the nearest neighbor pattern. These two plane lattices are the square and equilateral triangle nets which have either four- or six-fold rotational symmetry. The square net can only be folded into a cube. The triangular net can be folded into an icosahedron, octahedron or tetrahedron if 5, 4, or 3 of the triangular facets join at a polyhedron vertex and are disposed symmetrically.

To fold the icosahedron, the equilateral triangular net is vital with 5-fold axes through the twelve vertices, 3-fold axes on the triangular face centers, and 2-fold axes at the edge centers. This underlying geometrical requirement forces the structure units to arrange into rings of 5 and 6 [4]. However, the clustering into pentagons and hexagons is not a geometrical necessity, but a way of providing maximum contact in a close-packed array between the structure units and the surface of the shell. For example, it is thought that the icosahedral Bluetongue virus (BTV) capsid is composed of trimers, rather than hexamers [9].

The smallest pentagonal polyhedron is the pentagonal dodecahedron, which corresponds to $T = 1$ (12 pentamers with no hexamers). It was thought this should be the only case, as illustrated by Equation 8. Though recent discovery has shown the existence of a higher order pentagonal polyhedra in viral capsids [2]. The rest of this section mathematically defines the properties of pentagonal polyhedra.

3.1 The 72 Pentamer Case

There are icosahedral viruses like those found in the papillomavirdidae and polyomaviridae families which do not follow the CK Theory predictions for their capsid organization. Although these virions seem to follow a $T = 7d$ capsid structure, they are composed of exactly 72 pentamers rather than the 12 pentamers and 60 hexamers predicted by the CK Theory [2]. This puzzle has been solved by Twarock's theory which uses a tiling approach to replace the hexagonal lattice with a Penrose-like tiling of kites and rhombi [27, 28]. This has since been expanded in [12–14], making use of the affine extensions of Lie algebras.

Tiling theory still assumes that the overall structure of the capsid is icosahedral, though the constraint of the hexagonal lattice is replaced by allowing a tiling of any (or multiple) building block(s) that preserve the overall symmetry. This theory no longer follows the principle of quasi-equivalence, but a generalized principle which states that protein subunits are located only on the corner of tiles with the same angle [27]. This ensures identical proteins occupy identical sites on the tiles.

CK Theory only considers surface lattices that are induced by hexagonal lattices, such that the pentamers are not considered as being formed before folding, but are achieved only by the folding. In other words, the formation of pentamers allows the curvature needed to construct a three-dimensional shape [16]. Yet, for the Papilloma- and Polyoma-viridae families, the capsid is constructed from a pentagonal lattice which violates Equation 5, since the edges of the pentagonal lattice are not shared like the edges in the hexagonal lattice which constructs the $T = 7d$ structure. This is simply illustrated in the $T = 1$ case, where the pentagonal lattice has "holes" or spacing (Figure 5, gray) in two-dimensions, yet removing them allows the curvature needed to construct a three-dimensional shape when folding the orange pentagons. Similarly for larger values of T, the requirement of a closed net is relaxed. This allows for "spacing" between each pentamer not formed

by the triangular icosahedral folding, in the form of a pentagonal net where the pentagons do not share edges.

For the 72 pentamer case, the $T = 7d$ structure accurately predicts the 12 pentamers formed by closure of the triangular lattice, with 60 additional pentamers along the surface in the exact location of the predicted 60 hexamers. This folding still provides icosahedral symmetry: 5-fold symmetry guaranteed by the twelve triangle vertices, with 3-fold on the triangular face centers and 2-fold on the edges maintained by a specific pentamer orientation. Further details on the orientation of the pentamers for the Papilloma- and Polyomaviridae families are explained by the tiling theory in [27, 28], explained in the following section and illustrated in Figure 6.

3.2 Higher-Order Pentamer Case

Allowing spacing of pentamers creates a new class of potential capsid structures made of only pentamers following icosahedral symmetry. These structures preserve the underlying CK structure, using a triangular net projected onto a hexagonal lattice with 5-fold axes on the vertices, 3-fold axes on the triangular faces, and 2-fold axes on the triangular edges. After folding, each hexagon is replaced by pentagons without sharing edges. This cannot be done with all T-number constructions, but only a subclass of $P \geq 7$ as summarized in Table 2 and described by the following Theorem.

Table 2: Summary of T- and P-number requirements for Penta-Hexa (polyhedra composed only of pentagons and hexagons) and Penta Polyhedra (polyhedra composed only of pentagons). Only one subclass of $P \geq 7$, with $T = 1 \pmod 3$, is able to produce Penta Polyhedra.

T-number	Penta-Hexa Polyhedra	Penta Polyhedra
$T = 0 \pmod 3$	$P = 3, P \geq 7$	impossible
$T = 1 \pmod 3$	$P = 1, P = 3, P \geq 7$	$T = 1, P \geq 7$
$T = 2 \pmod 3$	impossible	impossible

Theorem 3.1. *If each hexagon in a CK icosahedral structure can be replaced by a pentagon and maintain icosahedral symmetry, then $P \geq 7$ and $T = 1 \pmod 3$.*

First we examine the P classes in which pentagonal polyhedra are not possible. In class $P = 1$, the T numbers are based along the h-axis or k-axis ($T = h^2 = 1, 4, 9$,etc...), with $h > 0$, $k = 0$ or $h = 0$, $k > 0$, respectively. This produces lattice lines that run parallel to the edges of the triangular facet, cutting the hexagons along the edge exactly in two (shown Figure 4, bottom right $(h, 0) = (0, h)$ blue triangle). If one were to replace the hexagons with pentagons, the triangular edge would then cut a pentagon in half, causing a loss of 2-fold symmetry in the center of the edge required by the icosahedral structure. Therefore structures retaining icosahedral symmetry and following the $P = 1$ class, larger than $T = 1$, cannot be composed of only pentagons.

Class $P = 3$ is defined by the lattice pair (h, h). The triangular face of these icosahedra have a centroid of $G(0, h)$ which lies in the center of a hexagon (shown Figure 4, bottom right (h, h) cyan triangle), according to Lemma 3.2 below. Replacing each hexagon with a pentagon would place a pentagon directly in the center of the triangular face, causing a loss of 3-fold symmetry. This shows that structures following the $P = 3$ class cannot be composed of only pentagons.

Lemma 3.2. *For a CK icosahedron, $(h - k)$ is divisible by 3 iff the centroid of the triangular facet lies in the center of a hexagon on the underlying lattice.*

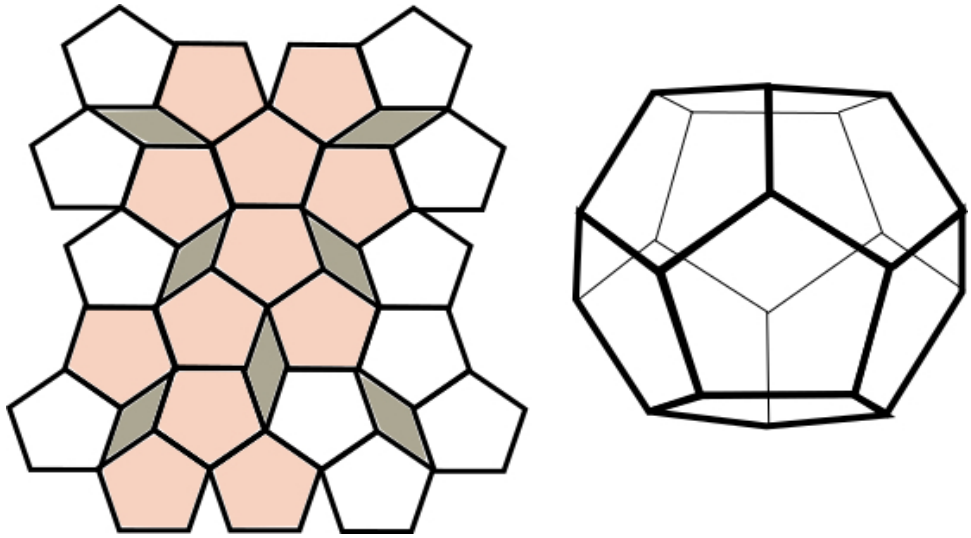

Figure 5: Illustration of a pentagonal lattice with “holes" or spacing (gray). Cutting the $T = 1$ lattice (orange) produces a three-dimensional dodecahedron.

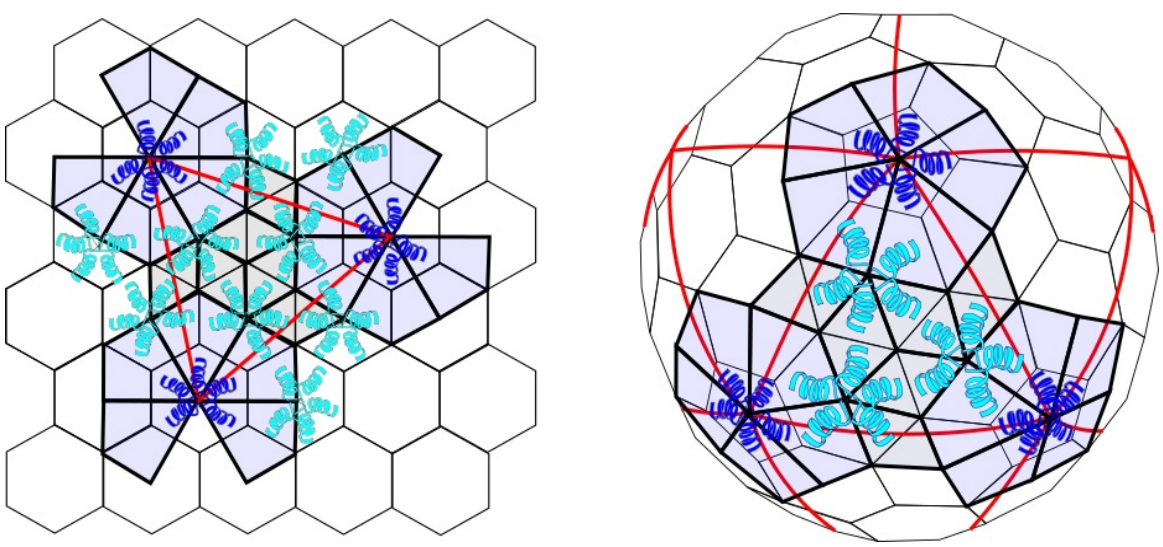

Figure 6: Illustration of the $T = 7d$ capsid structure for the Papilloma- and Polyoma-viridae families. 12 pentamers are formed by the closure of the triangular lattice (dark blue). 60 additional pentamers fall along the surface in the exact location of the Caspar-Klug predicted 60 hexamers (cyan).

Proof of Lemma 3.2. Consider an arbitrary triangular face of a CK icosahedron with vertices $A(0,0)$, $B(h,k)$ and $C(-k,h+k)$. Then the centroid of the triangular face is given by

$$G\left(\frac{h-k}{3},\frac{h+2k}{3}\right). \tag{17}$$

(i) If $(h-k)$ is divisible by 3, then so is $(h+2k)$. Then $G(m,n)$, with $m,n\in\mathbb{Z}$ not both zero, and the lattice vector from $(0,0)$ to the centroid is given by $\vec{G}=m\vec{a}_1+n\vec{a}_2$. Thus G lies in the center of a hexagon.

(ii) Let the centroid of the triangular face lie in the center of a hexagon in the underlying hexagonal lattice. Then the lattice vector from $(0,0)$ to the centroid is a linear combination of the basis vectors given by $\vec{G}=m\vec{a}_1+n\vec{a}_2$ with $m,n\in\mathbb{Z}$ not both zero, and the centroid is defined as $G(m,n)$. Thus $m=(h-k)/3$ and $n=(h+2k)/3$. But $m,n\in\mathbb{Z}$, therefore $(h-k)$ is divisible by 3. □

It remains to examine the $P\geq 7$ class. The 5-fold symmetry on each vertex is guaranteed by the construction of the icosahedron using 20 equilateral triangles, since these pentagons are achieved by the folding. The 2-fold symmetry on the edges is also possible since no edge in the triangulation of the $P\geq 7$ class can lie parallel to the lattice lines (explained in Section 2).

An underlying property of the icosahedral triangular face on a hexagonal lattice is that the triangle exhibits 3-fold rotational symmetry through the center of its face. If the center of the triangle fell on a hexagonal edge, 3-fold symmetry would be lost. This means the center of the triangle can only lie in the center of a hexagon or on a hexagonal vertex. Thus for pentagonal polyhedra, the 3-fold symmetry requirement is possible only when the center of the triangular face does not land in the center of a hexagon.

Through the discussion above, it is sufficient to prove Theorem 3.1 by the following Lemma.

Lemma 3.3. *The center of the triangular face lands on a vertex of the underlying hexagonal lattice iff* $T=1(\mod 3)$.

Remark 3.4. It is clear by

$$4T=3(h+k)^2+(h-k)^2, \tag{18}$$

that the triangulation number is either $T=0(\mod 3)$ or $T=1(\mod 3)$. Meaning, either T or $T-1$ is a multiple of 3.

Proof of Lemma 3.3. (i) Consider an arbitrary triangular face of a CK icosahedron with vertices $A(0,0)$, $B(h,k)$ and $C(-k,h+k)$. Then the centroid of the triangular face is given by Equation 17. Let the center of the triangular face be a vertex. Then by Lemma 3.2, $(h-k)\nmid 3$ and $(h-k)^2+3hk=T\nmid 3$. Therefore, if the center of the triangular face lands on a vertex of the underlying hexagonal lattice, then $T=1(\mod 3)$.

(ii) Let the centroid of the triangular face lie in the center of a hexagon in the underlying hexagonal lattice. By Lemma 3.2, (h-k) is divisible by 3. Then $(h-k)^2+3hk=T$ is also divisible by 3 and $T=0(\mod 3)$. Therefore if $T=1(\mod 3)$ then the center of the triangular face lands on a vertex of the underlying hexagonal lattice. □

Given the CK construction, the pentagonal polyhedra will have 5-fold symmetry through each triangular vertex. 2-fold symmetry on the edges is given by $P\geq 7$ and 3-fold symmetry on the triangular face is possible when the center of the triangular face lands on a vertex of the underlying hexagonal lattice, given by $T=1(\mod 3)$. To demonstrate the pentagonal polyhedra class, we consider the orientation of the pentamers of the $T=7d$ capsid structure for the Papilloma- and Polyoma-viridea families (illustrated in Figure 6).

The hexagonal lattice denotes the hexamer protein locations such that the six proteins are joined at the center while each protein points to one of the six vertices of the hexagon (See [27], Figure 5). In other words, exactly one protein lies between the basis vectors $\vec{a}_1$ and $\vec{a}_2$ on the two-dimensional lattice. Though on a pentagonal lattice, $\vec{a}_1$ would bisect the protein cluster since one of the pentamer proteins must point to the center of a hexagonal edge to maintain symmetry (for example see Figure 6 right, for orientation of a pentamer inside of a hexagon).

Proteins from neighboring pentamers, located inside of the icosahedral CK triangle, create a grid (or tiling [27]) (for example see Figure 6 left). This protein orientation ensures the maximum contact between pentamer

proteins, as described in the CK theory [4]. Since the protein orientation depends on the lattice vectors, the icosahedral CK triangle maintains 5-fold symmetry through each vertex, 2-fold symmetry on the edges and 3-fold symmetry on its face.

4 Discussion

Most viral capsids have been shown to exhibit an icosahedral fullerene-like structure, following icosahedral, helical, or complex symmetry. The Caspar-Klug Theory correctly predicts the arrangement of proteins for a majority of icosahedral capsids. However for a number of viruses following a pentagonal polyhedron structure, including viruses of the Papovaviridae family, the theory does not apply. The analysis introduced here shows that pentagonal polyhedra fall in a subclass of icosahedral viral capsid-like structures with $P \geq 7$ and $T = 1(\mod 3)$. Viral capsids have already been found that follow the pentagonal polyhedron structure with $T = 1$ and $T = 7$ [2]. It will be interesting to see if larger pentagonal polyhedron capsid structures emerge as mutations or with imminent virology discoveries. Future line of work includes investigating the assembly of pentagonal polyhedra and the physical properties of chirality required for drug design and development [3].

The CK Theory and pentagonal polyhedron structures still do not define all of the virus species found to have icosahedral symmetry. The Geminiviridae family has capsids composed of two incomplete $T = 1$ icosahedra joined together to form twinned particles [15]. The L-A virus is considered to have $T = 2$ structure composed of 22 pentamers, following a $T = 1$ icosahedral symmetry since the 12 pentons are composed of five sets of two Gag proteins [22]. There are also icosahedral viral capsids with non-identical subunits, such as the Adenovirus (AdV) capsid with four types of minor proteins classified as a $pseudo\ T = 25$ [18] and the Picornaviruses with $pseudo\ T = 3$ [25]. Many viruses have tubular, lozenge-like, or coffin shaped capsids which models for each have been proposed in [10, 17, 20, 21, 23, 26] and references therein. Models for the mature HIV-1 conical cores have been described in [23, 26].

The pentagonal polyhedron capsid-like structures presented in this manuscript, and models referenced above, could act as blueprints for novel antiviral drugs and nanotechnology applications. While viruses are mostly known for their destructive effects, they have proven to have a number of benefits as well. Understanding the structure of capsids and similar structures is of utmost importance not only for the design of new antiviral drugs, but also for applications in nanotechnology. For example, the construction of protein cages from viral proteins can be used as gene vectors, transporting generic material into cells for therapeutic purposes. The particles provide containers preventing premature degradation of drugs, which combined with the high host-cell specificity of viruses, can deliver these drugs to the specific targeted tissues [7, 8, 11, 24].

Acknowledgement: The authors thank Dr. Tony Nance at the Mathematical Biosciences Institute and Prof. Jiangguo (James) Liu at Colorado State University for numerous useful discussions. This research has been supported in part by the Mathematical Biosciences Institute and the National Science Foundation under grant DMS 1440386, and the Ohio State Undergraduate Research Scholar Award given to the second author.

Conflict of interest: The authors state no conflict of interest.

Ethical approval: The conducted research is not related to either human or animals use.

References

[1] T.S. Baker, N.H. Olsen, and S.D. Fuller, Adding the third dimension to virus life-cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol. Mol. Biol. Rev. **63**(1999), 862–922.

[2] D.M. Belnap, N.H. Olson, N.M. Cladel, W.W. Newcomb, J.C. Brown, J.W. Kreider, N.D. Christensen, and T.S. Baker, Conserved features in Papillomavirus and Polyomavirus capsids. J. Mol. Biol. **259**(1996), 249–263.

[3] W.H. Brooks, W.C. Guida, and K.G. Daniel, The significance of chirality in drug design and development. Current Top. in Med. Chem. **11**(2011), 760–770.
[4] D. Caspar and A. Klug, Physical principles in the construction of regular viruses. Cold Spring Harb. Symp. Quant. Biol. **27**(1962), 1–24.
[5] F.H.C. Crick and J.D. Watson, The structure of small viruses. Nature **177**(1956), 473–475.
[6] T. Douglas and M. Young, Virus: Making friends with old foes. Science **312**(2006), 873–875.
[7] O.M. Elrad and M.F. Hagan, Encapsulation of a polymer by an icosahedral virus. Phys. Biol. **7**(2010), 045003pp.
[8] A. de la Escosura, R. Nolte, and J. Cornelissen, Viruses and protein cages as nanocontainers and nanoreactors. J. Mater. Chem. **19**(2009), 2274–2278.
[9] J.M. Grimes, J.N. Burroughs, P. Gouet, J.M. Diprose, R. Malby, S. Ziéntara, P.P.C. Mertens, and D.I. Stuart, The atomic structure of the bluetongue virus core. Nature **395**(1998), 470–478.
[10] 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.
[11] 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.
[12] T. Keef and R. Twarock, New insights into viral architecture via affine extended symmetry groups. Comp. and Math. Meth. in Med. **9**(2008), 221–229.
[13] T. Keef, R. Twarock, and K.M. Elsawy, Blueprints for viral capsids in the family of Polyomaviridae. J. Theor. Biol. **253**(2008), 808–816.
[14] T. Keef and R. Twarock, Affine extensions of the icosahedral group with applications to the three-dimensional organization of simple viruses. J. Math. Biol. **59**(2009), 287–313.
[15] M. Krupovic, J.J. Ravantti. and D.H. Bamford, Geminiviruses: a tale of a plasmid becoming a virus. BMC Evol. Biol. **9**(2009), 1–11.
[16] J. Liu, F. Sadre-Marandi, S. Tavener, and C. Chen, Curvature concentrations on the HIV-1 capsid. Mol. Based Math. Biol. **3**(2015), 43–53.
[17] A. Luque and D. Reguera, The structure of elongated viral capsids. Biophys. J. **98**(2010), 2993–3003.
[18] C.S. Martín, Latest insights on Adenovirus structure and assembly. Viruses **4**(2012), 874–877.
[19] 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.
[20] M.F. Moody, The shape of the T-even bacteriophage head. Virology **26**(1965), 567–576.
[21] M.F. Moody, Geometry of phage head construction. J. Mol. Biol. **293**(1999), 401–433.
[22] H. Naitow, J. Tang, M. Canady, R.B. Wickner, and J.E. Johnson, L-A virus at 3.4 A resolution reveals particle architecture and mRNA decapping mechanism. Nature Letters **9**(2002), 725–728.
[23] 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.
[24] J.K. Pokorski and N.F. Steinmetz, The art of engineering viral nanoparticles. Mol. Pharma. **8**(2011), 29–43.
[25] B.V. Prasad,and M.F. Schmid, Principals of Virus Structural Organization. Adv. Exp. Med. Biol. **726**(2012), 17–47.
[26] F. Sadre-Marandi, J. Liu, S. Tavener, and C. Chen, Generating vectors for the lattice structures of tubular and conical viral capsids. Mol. Based Math. Biol. **2**(2014), 128–140.
[27] R. Twarock, A tiling approach to virus capsid assembly explaining a structural puzzle in virology. J. Theor. Biol. **226**(2004), 477–482.
[28] R. Twarock, The architecture of viral capsids based on tiling theory. J. Theor. Med. **6**(2005), 87–90.