# Discrete groups and internal symmetries of icosahedral viral capsids

Richard KERNER*

∗ Address : LPTMC, Université Paris-VI - CNRS UMR 7600 ,
Tour 24, 4-ème , Boite 121, 4 Place Jussieu, 75005 Paris, France
and Newton Institute, 20 Clarkson Rd., Cambridge, UK.
Tel.: +33 1 44 27 72 98, Fax: +33 1 44 27 51 00,
e-mail : richard.kerner@upmc.fr

Abstract: A classification of all possible icosahedral viral capsids is proposed. It takes into account the diversity of hexamers' compositions, leading to definite capsid size. We show how the self-organization of observed capsids during their production results from definite symmetries of constituting hexamers. The division of all icosahedral capsids into four symmetry classes is given. New subclasses implementing the action of symmetry groups $Z_2$, $Z_3$ and $S_3$ are found and described. They concern special cases of highly symmetric capsids whose $T = p^2 + pq + q^2$-number is of particular type corresponding to the cases $(p, 0)$ or $(p, p)$.

AMS classification numbers: 92Bxx, 92Cxx, 60Cxx

*Keywords:* Viral capsid growth; Self-organized agglomeration; Symmetry

# 1 Introduction

## 1.1 Capsid viruses

One of the outstanding features of nature is its extraordinary economy of means. Nature seems to be very parsimonious, no matter whether the economy concerns some fundamental laws processes, or more complicated phenomena including life and evolution.

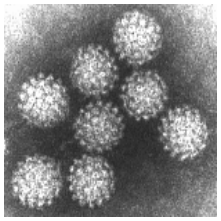
In the inanimated world this parsimony can be found in the *least action principle* and the minimization of free energy or free enthalpy in thermodynamical equilibrium. In biology one can often observe extraordinary efficiency of organisms which adapt themselves to various conditions by minimizing their energy loss and maximizing their survival probability.

One of the important ways to ensure minimal or maximal value of essential parameters is the use of symmetries. It is not a coincident that the least surface area and the greatest volume of a convex three-dimensional body is attained by the most symmetric one, which is the perfect sphere. In living organisms one is often amazed by the utter economy and efficiency, by perfect hydrodynamical properties of the fish and the aerodynamics of birds' wings.

There is also an fantastic ability to pack a maximal amount of information into smallest volumes available, which is done by the DNA dense packing in chromosomes. This spectacular ability to store information in an optimal way can be observed in viruses, in particular in the so-called *capsid viruses*, especially in their *icosahedral capsid* variety. Capsids play an essential role in viruses' survival, protecting the $DNA$, which is the acting part of the virus, from external dangers like chemical attacks or solar radiation. The capsids' sizes are optimally adapted to the size of the $DNA$ to be packed within, and when a mutation occurs leading to a longer $DNA$ chain, the size of the capsid must follow in order to be able to contain it. The information ruling capsids' construction and build-up is contained in the $DNA$ or both in $DNA$ and $RNA$, depending on virus type. It is encoded in the particular type of coat proteins, and can be reduced using the possibilities offered by the high degree of symmmetry displayed by the icosahedral shape.

In what follows, we present the generalization of the model of assembly schemes of viral capsid symmetries successfully applied to icosahedral capsids in our previous papers ([21], [2], [3], [4]) based on the analysis of the protein content of the elementary building blocks, the five- and six-fold capsomers displaying different internal symmetries.

A very large class of viruses build protection shells called *capsids*, made of special *coat proteins* and produced during the reproduction cycle inside an infected cell.



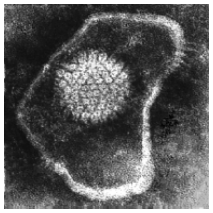


Figure 1: Papilloma viruses A herpes virus with its tegument

During infection, the capsid is left behind, while the $DNA$ strain is injected into cell's nucleus. Later on, the $DNA$ multiplies itself using the genetic material of the cell; parallelly, special *coat proteins* are synthetized, with which new capsids are constructed inside the infected cell. Then the newly produced complete $DNA$ strains are packed into the empty capsids, and newly born viruses leave the cell, infecting its neighbors.

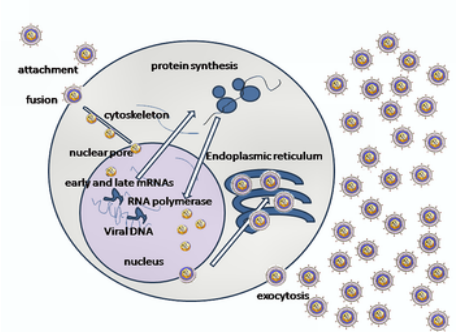



Figure 2: The stages of viral reproduction of the HSV virus.

A better knowledge of capsids' structure and symmetries can be helpful in understanding evolutionary trends and kinship between different virus species. The building schemes for capsids made of coat proteins are ruled by and encoded in the RNA and DNA molecules.

## 1.2 Triangular number

The *triangular number* $T$ is determined by two non-negative integers, $(p, q)$, via the simple formula:

$$T(p, q) = p^2 + pq + q^2.$$

All possible icosahedral structures made of 12 perfect pentagons and an appropriate number of perfect hexagons have been found by Coxeter in the middle of last century. Coxeter's classification was based on the notion of an elementary triangle, defined as follows.

Given two numbers $(p, q)$ and a perfect hexagonal grid, we start by picking up a hexagon and transform it into a perfect pentagon. Then make $p$ steps to the right, then $q$ steps at the angle of $120^o$, and plac the second pentagon there. Then repeat the same operation once more, and get the *elementary triangle*. Here are the examples of how this prescription works:

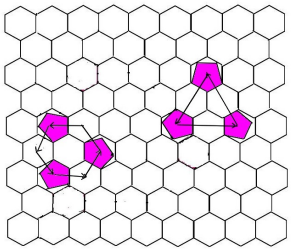


Figure 3: $(p, q) = (1, 1)$; $T = 3$, and $(p, q) = (2, 0)$; $T = 4$.

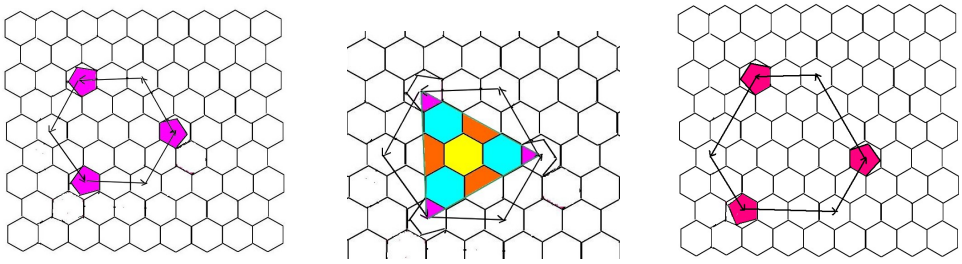


Figure 4: $(p, q) = (2, 2); T = 12,$, with its triangle; $(p, q) = (3, 2), T = 19$.

Let us show the schematic representations of icosahedral capsids. Due to the symmetry $T(p, q) = T(q, p)$, it is enough to consider pairs with $p \geq q$. The

smallest icosahedral capsids correspond to triangular numbers $T = 1,\ T = 3$ and $T = 4$.

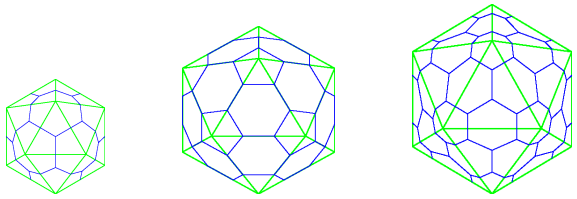


Figure 5: The smallest capsids: $T = 1,\ T = 3,\ T = 4$

The next three icosahedral capsids are generated by triangular numbers $T = 7,\ T = 9$ and $T = 12$.

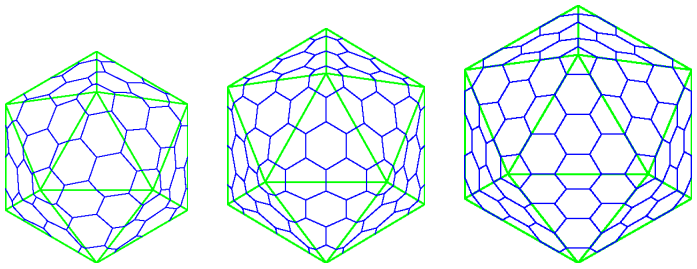


Figure 6: Capsids with $T = 7,\ T = 9$ and $T = 12$.

It is also useful to observe that the total number $N_6$ of hexamers in a capsid with a given $T$-number is given by the formula $N_6 = 10(T - 1)$.

# 2 Agglomeration

## 2.1 Random versus programmed agglomeration

The growth of icosahedral capsids via random agglomeration of capsomers seems highly unprobable. Were it really so, the final yield would be close to zero (something like $2^{-23} \simeq 10^{-8}$. This is so because in a random agglomeration process the error rate at each elementary step consisting in adding a new capsomer would be close to 50%, as it is shown in the next figure (7). The observed efficiency of capsid construction from campsomers produced in infected cells is close to 100%

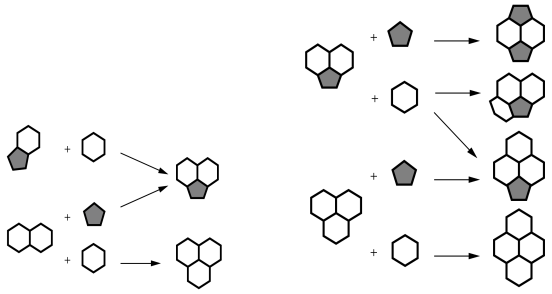


Figure 7: Random agglomeration of 5- and 6-sided capsomers

The most natural conclusion is that instead of random agglomeration of capsomers, what takes place is a very strict assembling process, with exclusive sticking rules. At least in the species whose triangular number is not too high, up to *Adenoviridae* ($T = 25$): this is the most plausible explanation of the observed *full use of capsomers*, with almost no waste.

For very large icosahedral capsids the agglomeration is less successful, and displays less ordered character In this case the driving forces for assembly in a particularly symmetric way are most probably of entropic and energetic nature.

As already stated above, the information about icosahedral capsid's structure is encoded in its *basic triangle*, twenty identical copies of which form an icosahedron. Consider the simplest case (besides a dodecahedron, with no hexamers at all, observed in *Microviridae*). The $T = 3$ capsid is quite common, namely in *Cowpea* and in the group *Paroviridae*.

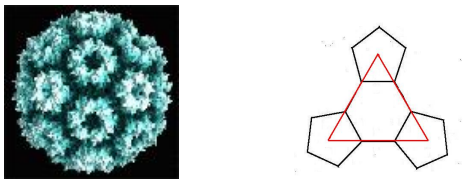


Figure 8: Basic triangle of the $T = 3$ icosahedral capsid

## 2.2 Capsomer differentiation

Sometimes the capsids are assembled with dimers and trimers, but the resulting pattern is the same, as shown in the Figure (9): In the *Adenovirus*

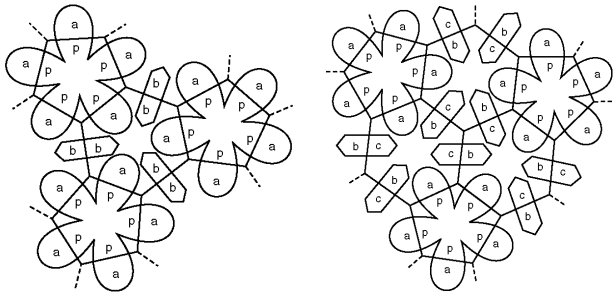


Figure 9: $T = 3$ and $T = 4$ capsids realized with dimers and trimers.

capsid hexamers occupying different positions in the capsid must differ from each other. There are *four* different types, as dictated by symmetry of the capsid, plus the twelve pentagons.

We therefore must conclude that hexamer differentiation is necessary in order to ensure the right agglomeration scheme. Had hexamers all their sides equivalent, nothing would stop the formation of undesirable clusters which cannot lead to the correct construction of the *Adenovirus* capsid, as shown in the figure below:

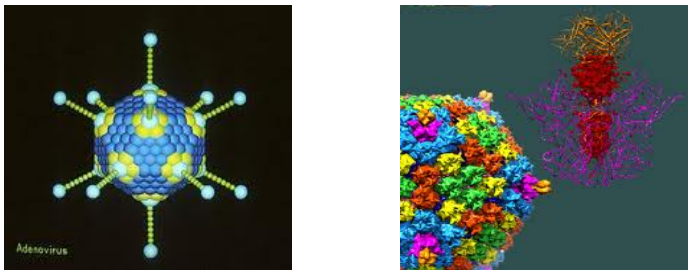



Figure 10: The *adenovirus*, $T = 25$; *Capsomer differentiation*

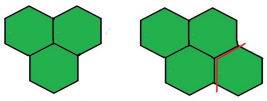


Figure 11: *allowed* *forbidden*

# 3 Symmetries

## 3.1 The simplest discrete groups

Capsomers, which are the building blocks from which capsids are assembled, display various internal symmetries due to differentiation of coat proteins forming them. Pentamers and hexamers can be made of one, two, three or more different proteins.

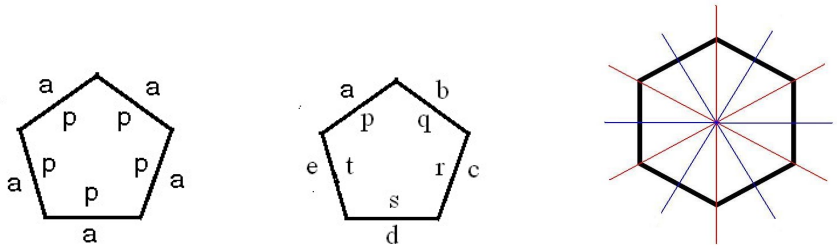



Figure 12: Being a prime number, 5 is divisible only by 1 and itself, giving rise to two admissible configurations: all sides identical, or all sides different. A hexagon displays a much higher symmetry.

The internal symmetries of capsomers can be analyzed with the help of simplest discrete groups, known as *permutation groups*. Denoted by $S_n$, they consist of all permutation operations acting on any set containing $n$ items. The dimension of an $S_n$ group is therefore equal to $n!$. Cyclic permutations of $n$ elements form an $n$-dimensional subgroup of $S_n$ denoted by $Z_n$.

The $S_2$ group contains only two elements, the identity keeping two items unchanged, and the only non-trivial permutation of two items, $(ab) \rightarrow (ba)$. This permutation is cyclic, so the $S_2$ group coincides with its $Z_2$ subgroup. The simplest representations of the $Z_2$ group are realised via its actions on the complex numbers, $\mathbf{C}^1$. Three different inversions can be introduced, each of them generating a different representation of $Z_2$ in the complex plane $\mathbf{C}^1$:

*i*) the sign inversion, $z \rightarrow -z$;

*ii*) complex conjugation, $z \rightarrow \bar{z}$;

*iii*) the combination of both, $z \rightarrow -\bar{z}$.

One should not forget about the fourth possibility, the trivial representation attributing the identity transformation to the two elements of the group, including the non-trivial one:

*iv*) the identity transformation, $z \rightarrow z$.

The $Z_2$ group can be implemented on a plane with two different actions: We shall denote the first realization by $Z_2^I$, and the second by $Z_2^R$.

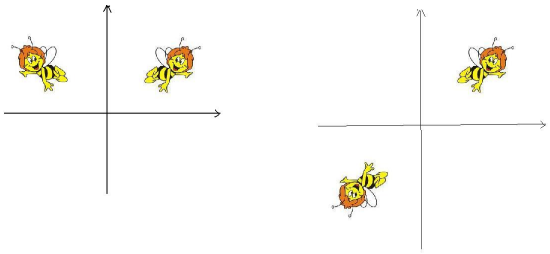


Figure 13: Inversion Rotation by $180^o$

Two simple discrete groups next in row after $Z_2$ are of particular interest to us: The symmetric $S_3$ group and its cyclic subgroup $Z_3$.

The symmetric group $S_3$ containing all permutations of three different elements is a special case among all symmetry groups $S_N$. It is exceptional because it is the first in the row to be non-abelian, and the last one that possesses a faithful representation in the complex plane $\mathbf{C}^1$.

It contains six elements, and can be generated with only two elements, corresponding to one cyclic and one odd permutation, e.g. $(abc) \to (bca)$, and $(abc) \to (cba)$. All permutations can be represented as different operations on complex numbers as follows.

The cyclic group $Z_3$ has a natural representation on the complex plane. Let us denote the primitive third root of unity by $j = e^{2\pi i/3}$. Let the permutation $(abc) \to (bca)$ be represented by multiplication by $j$. Then the three cyclic permutations can be represented via multiplication by $j$, $j^2$ and $j^3 = 1$ (the identity), corresponding to the rotation by $120^o$, $240^o$ and $360^o$ (the identity transformation, equivalent to the rotation by $0^o$).

Odd permutations must be represented by idempotents, i.e. by operations whose square is the identity operation. We can make the following choice: let the odd permutation $(abc) \to (cba)$ be represented by the complex conjugation $z \to \bar{z}$, or the reflection with respect to thge real axis.

Then the six $S_3$ symmetry transformations contain the identity, two rotations, one by $120^o$, another one by $240^o$, and three reflections, in the $x$-axis, in the $j$-axis and in the $j^2$-axis. The $Z_3$ subgroup contains only the three rotations, as shown in the following figure (14):

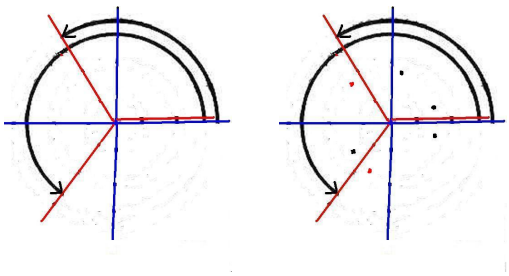


Figure 14: Rotations by $120^o$ and $240^o$ (left) and the full $S_3$ group.

## 3.2 Internal symmetries of hexamers

The symmetry of a hexamer is dictated by the place it occupies in the icosahedral capsid. It may lie on one of the symmetry axes, which may be two- or three-fold. A three-fold symmetry is realized by hexamers with the *ababab* coat proteinscheme, while a 2-fold symmetry can be realized in many ways, e.g. with *abcabc* scheme.

Let us proceed to the analysis of discrete symmetries that can occur in various hexamers. We shall start with the least symmetric one, which has totally differentiated sides according to the scheme $(abcdef)$, then proceed to the more symmetric cases. The $abcdef$ scheme is shown below: The next

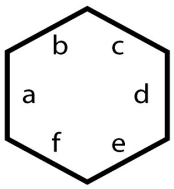



Figure 15: The hexamer with totally differentiatesd sides - six different proteins.

example is provided by a hexamer admitting only one symmetry $Z_2^R$. Its sides are labeled according to the scheme $(abcabc)$. Another simplest example is a hexamer admitting only one symmetry $Z_2^I$. Its sides are labeled according to the scheme $(bdffdb)$. A third possibility of hexamer admitting only one symmetry $Z_2^I$ is given below. Its sides are labeled according to the scheme $(abcdcb)$. More elaborated hexamers schemes can admit two $Z_2^I$ symmetries. In such a case hexamer sides are labeled according to the scheme $(bbccbb)$. Next example is provided by a hexamer admitting the $Z_3$ symmetry. Its sides are labeled according to the scheme $(ababab)$. With this in mind, we can

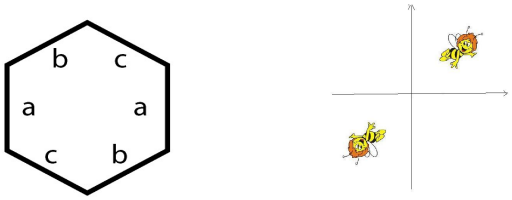



Figure 16: Partially differentiated sides - three different proteins.


Figure 17: Partially differentiated sides - three different proteins.

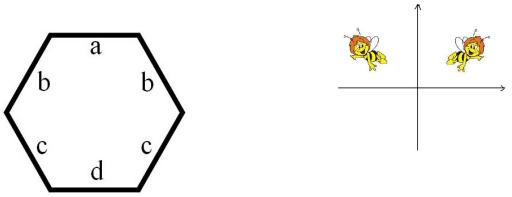



Figure 18: Partially differentiated sides - four different proteins.

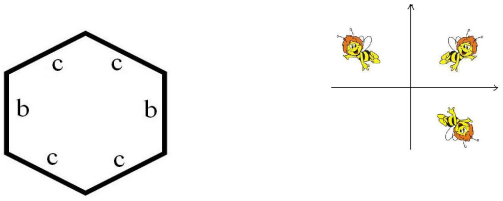
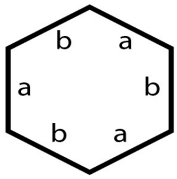



Figure 19: Partially differentiated sides - two different proteins.

explore the symmetries of capsids constructed with various hexamer types.



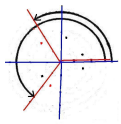


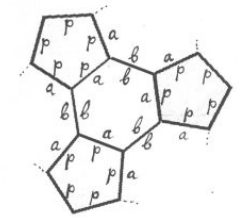
Figure 20: Partially differentiated sides - two different proteins.

# 4 Classification

## 4.1 Affinity matrices

In what follows, we shall mark the five pentagon forming proteins with letter "p", and the sides of chosen hexamers that stick to pentamers' sides with letter "a".

The information concerning the agglomeration scheme can be encoded in a corresponding *affinity matrix*, which is a square table whose lines and columns are labeled with different letters denoting all different protein types. In the intersection of lines and columns we put 1 if the corresponding protein types stick together, and 0 if the corresponding agglomeration is forbidden. This can be also interpreted as a matrix of probabilities, 1 for the 100% probability of sticking together, and 0 when sticking together is totally excluded. The first example is provided by agglomerating pentamers with only one type of hexamers, containing only two types of coat proteins: The next



| | p | a | b |
|---|---|---|---|
| p | 0 | 1 | 0 |
| a | 1 | 0 | 0 |
| b | 0 | 0 | 1 |

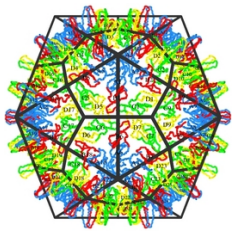
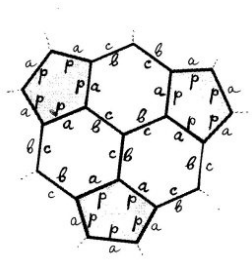


Figure 21: The assembly scheme and affinity matrix of a $T = 3$ capsid; a real example: the human *Hepatitis* virus.

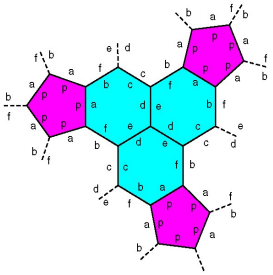
example is given by a bit more diversified hexamers containing threee types of coat proteins: The next $T$-number is 7. It admits two isomers, left and



| | p | a | b | c |
|---|---|---|---|---|
| p | 0 | 1 | 0 | 0 |
| a | 1 | 0 | 0 | 0 |
| b | 0 | 0 | 0 | 1 |
| c | 0 | 0 | 1 | 0 |

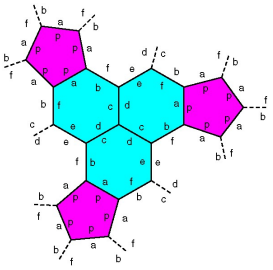
Figure 22: The assembly scheme and affinity matrix of a $T = 4$ capsid

right, and is found in *Papiloviridae*. Here is the left-oriented isomer.



| | p | a | b | c | d | e | f |
|---|---|---|---|---|---|---|---|
| p | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| a | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| b | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| c | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| d | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| e | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| f | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

Figure 23: The assembly scheme and affinity matrix of a $T = 7$ capsid (left)



| | p | a | b | c | d | e | f |
|---|---|---|---|---|---|---|---|
| p | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| a | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| b | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| c | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| d | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| e | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| f | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

Figure 24: The assembly scheme and affinity matrix of a $T = 7$ capsid (right)

New hexamers are needed to continue the game. For larger capsids, in which the rate of pentamers is lower, one cannot obtain proper assembling rules unless more than one type of hexamer is present, out of which only one is allowed to agglomerate with pentamers. In the case of two different hexamer types one obtains either the $T = 9$ capsid, or, with one totally differentiated

hexamer and two hexamers, one of the (*ababab*) type and another of the (*abcabc*) type, the $T = 12$ capsid.

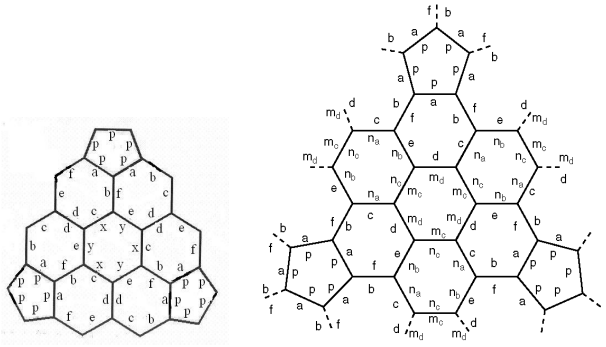


Figure 25: The assembly schemes of $T = 9$ and $T = 12$ capsids

To get the $T = 25$ adenovirus capsid, one must introduce no less than *four* different hexamers, all of the being of the (*abcdef*) type, out of which only one type can agglomerate with pentamers. As 25 is a square of prime number, each of these four types contains 6 different sides (proteins).

Counting in the unique protein found in pentamers, we get the result that the triangular number defines at the same time the number of different proteins participating in the construction.

The *"affinity matrices"* giving the links between various proteins belong the the set of the so-called *circulant matrices*, characterized by the fact that there is only one unit in each row and in each column, the rest of the entries being equal to 0

The square of any such matrix is the unit matrix, and its eigenvalues are 1 or $-1$. Due to this circumstance, each protein appears exactly 60 times in the fully built capsid.

## 4.2 General scheme with three hexamer types

In this section we consider only three hexamer types, (*ababab*), (*abcabc*) and (*abcdef*). We can organize all the capsids obtained with these hexamers in a single table below. To each value of triangular number $T$ corresponds a unique partition into $1 + (T - 1)$ where the "1" represents the unique pentamer type and $(T - 1)$ is partitioned into a sum of certain number of

different hexamer types, according to the formula

$$(T-1) = 6\,\alpha + 3\,\beta + 2\,\gamma$$

with non-negative integers $\alpha, \beta$ and $\gamma$, of which $\beta$ and $\gamma$ can take on only the values 0 or 1.

| Type (p,q) | $T = p^2 + pq + q^2$ | $N_6$ | $T$ decomposition |
|---|---|---|---|
| (1,1) | 3 | 20 | $1+2$ |
| (2,0) | 4 | 30 | $1+3$ |
| (2,1) | 7 | 60 | $1+6$ |
| (3,0) | 9 | 80 | $1+6+2$ |
| (2,2) | 12 | 110 | $1+6+2+3$ |
| (3,1) | 13 | 120 | $1+6+6$ |
| (4,0) | 16 | 150 | $1+6+6+3$ |
| (3,2) | 19 | 180 | $1+6+6+6$ |
| (4,1) | 21 | 200 | $1+6+6+6+2$ |
| (5,0) | 25 | 240 | $1+(4\times 6)$ |

Table II: Classification of icosahedral capsids. The last column gives the number and type of hexamers needed for the construction

To each value of the triangular number $T$ corresponds a unique partition into $1 + (T-1)$, where the " 1" represents the unique pentamer type and $(T-1)$ is partitioned into a sum of certain number of different hexamer types according to the formula:

$$(T-1) = 6\,\alpha\, + 3\,\beta\, + 2\,\gamma$$

with non-negative integers $\alpha, \beta$ and $\gamma$, the numbers $\beta$ and $\gamma$ taking on exclusively the values 0 or 1. This leads to four different classes, according to the choices:

$A:\ \ \beta = 0,\ \gamma = 0; \quad B:\ \ \beta = 1,\ \gamma = 0;$
$C:\ \ \beta = 0,\ \gamma = 1; \quad D:\ \ \beta = 1,\ \gamma = 1.$

This classification is based on the fact that the corresponding hexamers are centered on a *three-fold* or a *two-fold symmetry axis*, so that the first type must be found at the center of icosahedron's triangular face, whereas the second type must be found in the center of an edge between elementary

triangles. The number $\alpha$ of maximally differentiated hexamers follows then from the corresponding partition of a given triangular number, as shown in the following table (Fig.26).

| T and (p,q) | 1+(k×6) | 1+(k×6)+2 | 1+(k×6)+3 | 1+(k×6)+2+3 |
|---|---|---|---|---|
| 1 (1,0) | 1+(0×6) | | | |
| 3 (1,1) | | 1+(0×6)+2 | | |
| 4 (2,0) | | | 1+(0×6)+3 | |
| 7 (2,1) | 1+(1×6) | | | |
| 9 (3,0) | | 1+(1×6)+2 | | |
| 12 (2,2) | | | | 1+(1×6)+2+3 |
| 13 (3,1) | 1+(2×6) | | | |
| 16 (4,0) | | | 1+(2×6)+3 | |
| 19 (3,2) | 1+(3×6) | | | |
| 21 (4,1) | | 1+(3×6)+2 | | |
| 25 (5,0) | 1+(4×6) | | | |
| 27 (3,3) | | 1+(4×6)+2 | | |
| 28 (4,2) | | | 1+(4×6)+3 | |
| 31 (5,1) | 1+(5×6) | | | |
| 36 (6,0) | | | | 1+(5×6)+2+3 |
| 37 (4,3) | 1+(6×6) | | | |
| 39 (5,2) | | 1+(6×6)+2 | | |
| 43 (6,1) | 1+(7×6) | | | |
| 48 (4,4) | | | | 1+(7×6)+2+3 |
| 49 (7,0) | 1+(8×6) | | | |
| 49 (5,3) | 1+(8×6) | | | |
| 52 (6,2) | | | 1+(8×6)+3 | |
| 57 (7,1) | | 1+(9×6)+2 | | |
| 61 (5,4) | 1+(10×6) | | | |
| 63 (6,3) | | 1+(10×6)+2 | | |
| 64 (8,0) | | | 1+(10×6)+3 | |
| 67 (7,2) | 1+(11×6) | | | |
| 73 (8,1) | 1+(12×6) | | | |
| 75 (5,5) | | 1+(12×6)+2 | | |
| 76 (6,4) | | | 1+(12×6)+3 | |
| 79 (7,3) | 1+(13×6) | | | |
| 81 (9,0) | | 1+(13×6)+2 | | |

| T and (p,q) | 1+(k×6) | 1+(k×6)+2 | 1+(k×6)+3 | 1+(k×6)+2+3 |
|---|---|---|---|---|
| 84 (8,2) | | | | 1+(13×6)+2+3 |
| 91 (6,5) | 1+(15×6) | | | |
| 91 (9,1) | 1+(15×6) | | | |
| 93 (7,4) | | 1+(15×6)+2 | | |
| 96 (8,3) | | | | 1+(15×6)+2+3 |
| 100 (10,0) | | | 1+(16×6)+3 | |
| 103 (9,2) | 1+(17×6) | | | |
| 108 (6,6) | | | | 1+(17×6)+2+3 |
| 109 (7,5) | 1+(18×6) | | | |
| 111 (10,1) | | 1+(18×6)+2 | | |
| 112 (8,4) | | | 1+(18×6)+3 | |
| 117 (9,3) | | 1+(18×6)+2 | | |
| 121 (11,0) | 1+(20×6) | | | |
| 124 (10,2) | | | 1+(20×6)+3 | |
| 127 (7,6) | 1+(21×6) | | | |
| 129 (8,5) | | 1+(21×6)+2 | | |
| 133 (11,1) | 1+(22×6) | | | |
| 133 (9,4) | 1+(22×6) | | | |
| 139 (10,3) | 1+(23×6) | | | |
| 144 (12,0) | | | | 1+(23×6)+2+3 |
| 147 (7,7) | | 1+(24×6)+2 | | |
| 147 (11,2) | | 1+(24×6)+2 | | |
| 148 (8,6) | | | 1+(24×6)+3 | |
| 151 (9,5) | 1+(25×6) | | | |
| 156 (10,4) | | | | 1+(25×6)+2+3 |
| 157 (12,1) | 1+(26×6) | | | |
| 163 (11,3) | 1+(27×6) | | | |
| 169 (13,0) | 1+(28×6) | | | |
| 169 ((8,7) | 1+(28×6) | | | |
| 171 (9,6) | | 1+(28×6)+2 | | |
| 172 (12,2) | | | 1+(12×6)+3 | |
| 181 (11,4) | 1+(30×6) | | | |
| 183 (13,1) | | 1+(30×6)+2 | | |
| 189 (12,3) | | 1+(31×6)+2 | | |
| 192 (8,8) | | | | 1+(31×6)+2+3 |

Figure 26: Left: Classification of icosahedral capsids ) up to $T = 81$.

It is easy to see what is the arithmetic nature of triangular numbers of each group.

- The first column (type $A$) the triangular number $T$ must be either a prime number, or a square of a prime number.

- The $T$-numbers in the second column (type $B$) are divisible by 3;

- The $T$-numbers in the third column (type $C$) are multiples of 4;

- The fourth column (type $D$) contains triangular numbers divisible both by 3 and by 4, i.e. their $T$-numbers are multiples of 12.

| **T, (p,q)** | **A(6)** | **B(6 + 2)** | **C(6 + 3)** | **D(6 + 2 + 3)** |
|---|---|---|---|---|
| **1**, (**1**, **0**) | (**1**, **0**, **0**, **0**) | | | |
| **3**, (**1**, **1**) | | (**1**, **0**, **1**, **0**) | | |
| **4**, (**2**, **0**) | | | (**1**, **0**, **0**, **1**) | |
| 6 | – | – | – | (1, 0, 1, 1) |
| **7**, (**2**, **1**) | (**1**, **1**, **0**, **0**) | | | |
| **9**, (**3**, **0**) | | (**1**, **1**, **1**, **0**) | | |
| 10 | – | – | (1, 1, 0, 1) | – |
| **12**, (**2**, **2**) | | | | (**1**, **1**, **1**, **1**) |
| **13**, (**3**, **1**) | (**1**, **2**, **0**, **0**) | | | |
| 15 | – | (1, 2, 1, 0) | – | – |
| **16**, (**4**, **0**) | | | (**1**, **2**, **0**, **1**) | |
| 18 | – | – | – | (1, 2, 1, 1) |
| **19**, (**3**, **2**) | (**1**, **3**, **0**, **0**) | | | |
| **21**, (**4**, **1**) | | (**1**, **3**, **1**, **0**) | | |
| 22 | – | – | (1, 3, 0, 1) | – |
| 24 | – | – | – | (1, 3, 1, 1) |
| **25**, (**5**, **0**) | (**1**, **4**, **0**, **0**) | | | |
| **27**, (**3**, **3**) | | (**1**, **4**, **1**, **0**) | | |
| **28**, (**4**, **2**) | | | (**1**, **4**, **0**, **1**) | |
| 30 | – | – | – | (1, 4, 1, 1) |
| **31**, (**5**, **1**) | (**1**, **5**, **0**, **0**) | | | |
| 33 | – | (1, 5, 1, 0) | – | – |
| 34 | – | – | (1, 5, 0, 1) | – |
| **36**, (**6**, **0**) | | | | (**1**, **5**, **1**, **1**) |

Table III: Periodic Table of icosahedral capsids.

The four types, $A, B, C$ and $D$ are put in separate columns.

A good example of predictive ability of our scheme is the analysis of the *Herpes virus* $T = 16$ capsid.

The total number of major capsid proteins (capsomer forming blocks) can

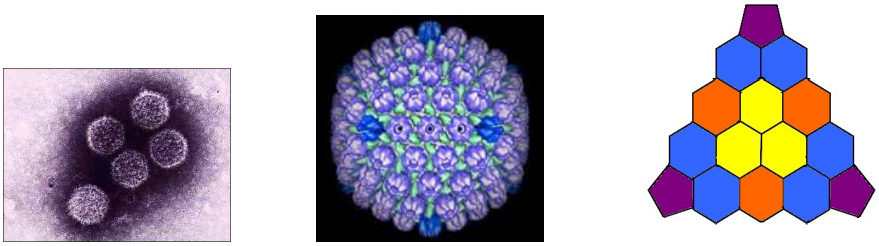


Figure 27: The *Herpesvirus* and its basic triangle with different hexamer types

be easily found via the construction algorithm. It is confirmed by observations, as has been found by B.L. Trus and al., ([15])

The $T = 16$ capsid's basic triangle contains two totally differentiated hexamers (blue and yellow), and one "axially symmetric" hexamer (red). It is quite easy now to proceed to the following protein counting:

Indeed, inside the basic triangle there are three copies of " $6A$", three copies of " $6B$" hexamers, which amounts to $3 \times (6 + 6) = 36$ proteins, and three halves of type "3" hexamers, i.e. $3 \times 3 = 9$ proteins, giving altogether 45 proteins; there are also three "p" proteins coming from pentamers (3/5) fragments, giving the total of 48. The capsid contains 20 such triangles, i.e. $20 \times 48 = 960$ major proteins.

# 5 Symmetric reduction

## 5.1 Higher symmetries

The classification scheme presented above is based on the exclusive use of three types of hexamer symmetries, $(ababab)$, $(abcabc)$ and $(abcdef)$. However, among the capsids with all possible values of triangular number $T$ there are two classes that display an extra internal symmetry. These are the ones corresponding to the particularly symmetric choice of two integers $(p, q)$. Capsids with $T$-numbers generated by combinations $(p, 0)$ *and* $(p, p)$ display an additional symmetry of the edges, which may contain hexamers of other types than the three ones used up to now: $(abbabb)$, $(abccba)$, etc.

New internal symmetries are displayed in the figure (28) below. Let us consider the edge symmetry. The first case when such symmetry occurs is in $T = 4$ capsids. The second case with additional symmetry on the edge can be realized under the condition that the $b$-sides are polarized so that

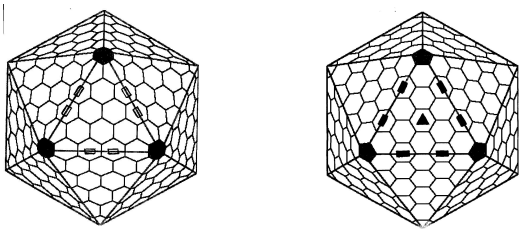


Figure 28: Two symmetric capsids: $(p,q) = (5,0)$, $T = 25$ and $(p,q) = (3,3)$, $T = 27$. One can observe the position of two-fold symmetry (on the edge) and the three-fold symmetry (in the center of basic triangle).

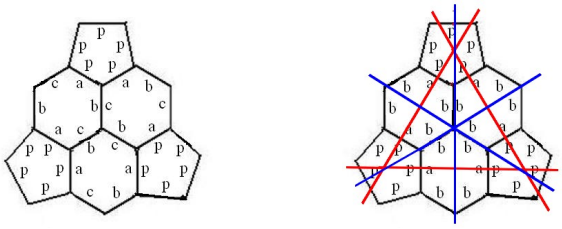


Figure 29: Two possible symmetries of the $T = 4$ capsid: with the $Z_3$ symmetry, and with the $S_3 \times Z_2^I$ symmetry. 3 different proteins instead of 4.

they can stick together only with one of the two possible polarizations: In

Figure 30: allowed forbidden

most icosahedral virus capsids the polarized character of major building coat proteins is indeed the case.

## 5.2 Reduction of coat proteins number

Almost all non-chiral capsids, especially the "perfect" ones corresponding to triangular numbers givenb by pairs $(p,0)$ or $(p,p)$ admit higher degrees of symmetry, and therefore, the reduction of number of different coat proteins needed for their construction. It can be seen on the following examples of

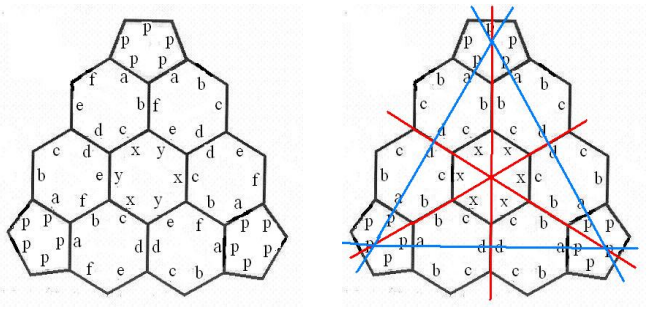


Figure 31: Two versions of the $T = 9$ capsid: the first one with $Z_3$ symmetry and 9 different coat proteins, the second one with the $S_3 \times Z_2$ symmetry and with 6 different coat proteins.

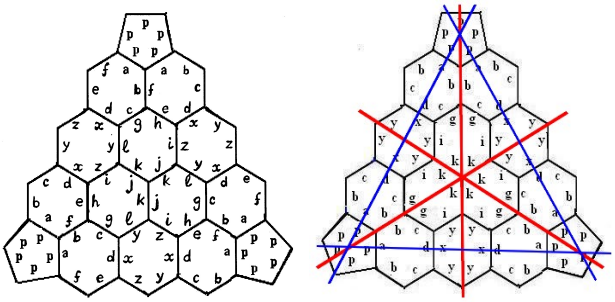


Figure 32: The T16 capsid symmetries: with $Z_3$ only and with $S_3 \times Z_2$ group

$T = 9$ and $T = 16$ capsids. The most symmetric capsids display a full $S_3 \times Z_2$ symmetry:

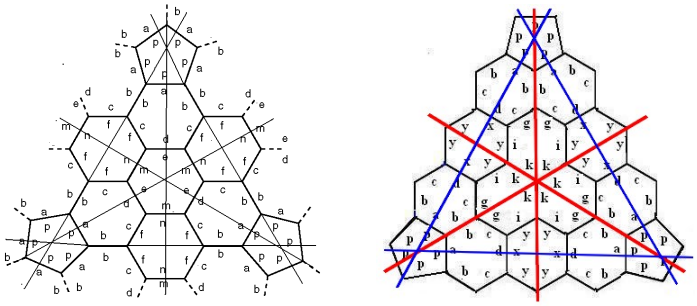


Figure 33: The $T = 12$ capsid with six symmetry axes; 9 different proteins instead of 12. same for the $T = 16$ capsid 10 different proteins instead of 16.

## 5.3 The $T = 25$ example

The best way to see how the reduction due to the additional symmetry does work is to consider a gradual construction respecting the full symmery group, which in this case will be $Z_3 \times Z_2 \times Z_2$.

Let us do it on the example of the $T = 25$ viruses: the *Adenovirus* and the *PRD1* virus, both having the same $T$-number equal to 25, but not the same number of different coat proteins, as can be seen in the next two figures, (34) and (35): The adenovirus structure is well known, and it is based on

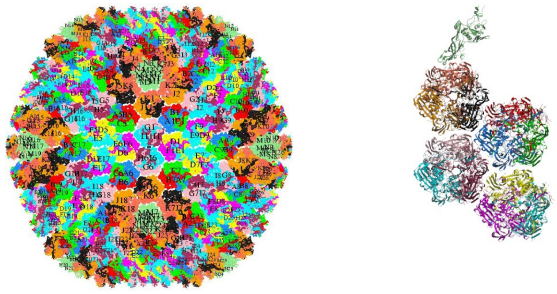


Figure 34: The $T = 25$ *Adenovirus* capsid and its coat proteins.

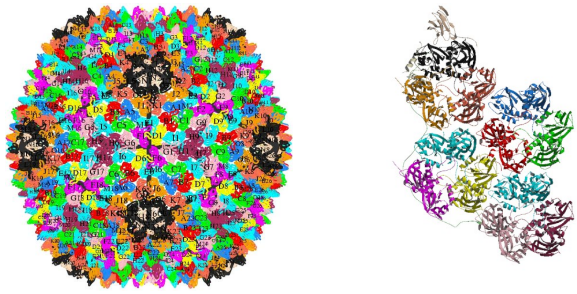


Figure 35: The $T = 25$ *PRD1* virus capsid and its coat proteins.

four species of totally differentiated hexamers. Its symmetry group is $Z_3$.

The complete analysis of the $PRD1$-capsid symmetry shows how the number of different coat proteins can be reduced due to the consequent use of symmetry.

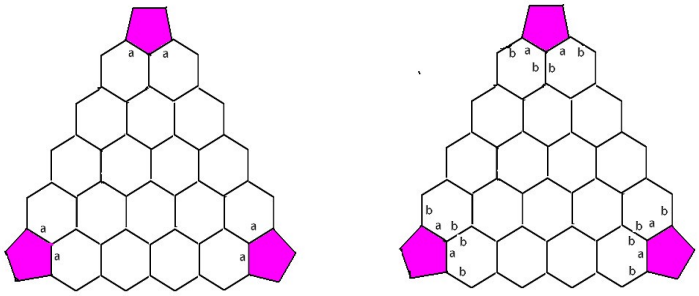


Figure 36: The $T = 25$ PRD1 capsid. If protein "b" is placed next to "a", it must be repeated in all other positions obtained via symmetry transformation.

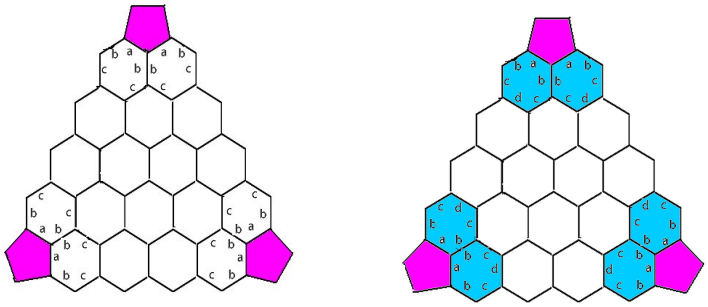


Figure 37: The $T = 25$ PRD1 capsid. With proteins "c" and "d", hexamers adjacent to pentamers are complete. fre.

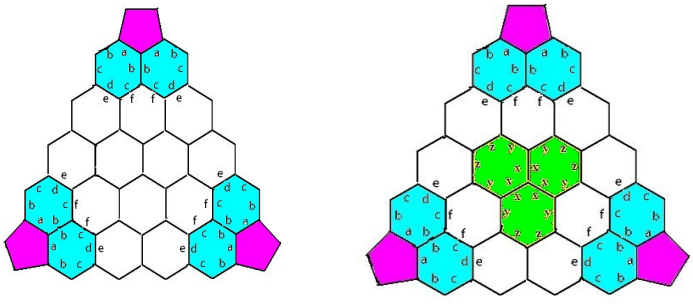


Figure 38: The $T = 25$ capsid. With new proteins "x", "y" and "z", the central hexamers are complete.

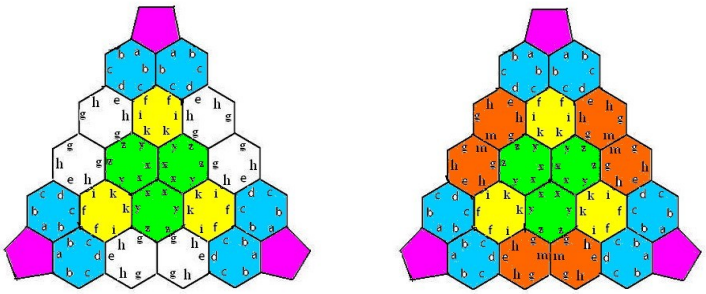


Figure 39: The $T = 25$ capsid. With all proteins placed, the capsid is complete.

# 6 Conclusions

From the point of view of the symmetrical reduction icosahedral capsids can be divided in three categories:

*The irreducible ones, with $T$ = a prime number $6n + 1$.*

$$T = 3, 7, 13, 19, 31, 37, 43, 61, 67, 73, ...$$

*The reducible ones, with $T$ = a square*

$$T = 4, 9, 16, 25, 36, 49, 64, 81...$$

*The reducible ones, with $T = 3 \times n^2$*

$$T = 12, 27, 48, 75, 108, 147...$$

and the rest:

$$T = 21, 28, 39, 52, 63, 76...$$

The total number of basic coat proteins remains the same, and in all cases is equal to

$$N_p = T.$$

As a consequence, in the case of the reduced number of different proteins due to symmetry reduction, the total number of those which appear not three, but six times in the elementary triangle, is equal to 120. It seems

plausible that the three types are genetically related, because they display a similar hexamer structure:

*all hexamers maximally differentiated in the first case, and symmetrically reduced in the second and third cases.*

*The families with additional symmetries are liable to possess some evolutionary kinship.*

*The viruses with capsids having a prime number $T$ should be also genetically close:* $T = 7, 13, 19, 31, 37, ...$

The two schemes, the less symmetric and the maximally symmetric one, may have their own advantages and shortcomings. The first one displays the same number of copies of each particular coat protein, i.e. 60, which supposedly simplifies their chain production. In the second case the number of different coat proteins is reduced, but they must be produced in different amounts, some in 60, some in 120 copies. It would be interesting to know why certain types of icosahedral capsid viruses choose the particular scheme between these two possibilities.


## Acknowledgements

The Author is greatly indebted to Prof. Reidun Twarock of the University of York for her invaluable help, numerous suggestions, enlightening remarks and constructive criticism. Discussions with Dr. Tom Keefe and his help with documentation are gratefully acknowledged.

## References


[1] R. Kerner *Model of viral capsid growth.* Journal Computational and Mathematical Methods in Medicine, **6**, Issue 2, (2007), 95–97

[2] R. Kerner *Classification and evolutionary trends of icosahedral viral capsids.* Journal Computational and Mathematical Methods in Medicine, **9**, Issue 3 & 4, (2008), 175–181.

[3] R. Kerner, *Models of Agglomeration and Glass Transition.* Imperial College Press, (2007).

[4] R. Kerner, *Self-Assembly of Icosahedral Viral Capsids: the Combinatorial Analysis Approach*, "Mathematical Modelling of Natural Phenomena", Vol. 6, No. 6, (2011), 136-158

[5] D. D. Richman, R. J. Whitley and F. G. Hayden, *Clinical Virology.* (second edition); ASM Press, Washington DC (2009)

[6] M.C.M. Coxeter, *"Regular polytopes"*, Methuen and C,, London, (1948)

[7] M. Eigen, 1971, *Selforganization of matter and the evolution of biological molecules*, Springer-Verlag, *D*ie Natutwissenschaften **58** heft 10,

[8] H. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley. $C_{60}$*: Buckminsterfullerene.* Nature, 318 (1995), 162–163.

[9] D. L. D. Caspar and A. Klug, *Physical Principles in the Construction of Regular Viruses.* Cold Spring Harbor Symp. Quant. Biology, **27** (1962), No 1, 1–24.

[10] A. Zlotnick, *To Build a Virus Capsid : An Equilibrium Model of the Self Assembly of Polyhedral Protein Complexes.* J. Mol. Biology **241**, (1994), 59–67.

[11] S. B. Larson *e*t al., 1998, *Refined structure of satellite tobacco mosaic virus at 1.8 A resolution.* Journal of Molecular Biology **277** (1998) 37–59.

[12] D. J. McGeogh and A. J. Davison, , *The descent of human herpesvirus.* 8.Semin. Cancer Biology, **9** (1999), 201–209.

[13] D. J. McGeogh and A. J. Davison, *The molecular evolutionary history of the herpesviruses: origins and evolution of viruses*, Academic Press Ltd., London (1999).

[14] P. L. Stewart, R. M. Burnett, M. Cyrklaff, S. D. Fuller, *Image reconstruction reveals the complex molecular organization of adenovirus.* Cell, **67** (1991) October 4, 145–154

[15] B. L. Trus *e*t al., *Capsid structure of Kaposi's sarcoma-associated herpesvirus*, Journal of Virology, **75** (2001) No 6, 2879-2890.

[16] Q. Wang, T. Lin, L. Tang, J. E. Johnson and M. G. Finn, *Icosahedral Virus Particles as Addressable Nanoscale Building Blocks.* Angewandte Chemie, **114** (2002), o 3, 477–480. (2003) 167–236

[17] H. R. Hill, N. J. Stonehouse, S. A. Fonseca and P. . Stockley, *Analysis of phage MS2 coat protein mutants expressed from a reconstituted phagemid reveals that proline 78 is essential for viral infectivity. J*ournal of Molecular Biology, **266**, (1997), 1–7.

[18] P. E. Prevelige, D. Thomas and J. King, *Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells.* Biophys. Journal **64** (1993), 824–835;

[19] B. Buckley, S. Silva and S. Singh, S. 1993, *Nucleotide sequence and in vitro expression of the capsid protein gene of tobacco ringspot virus.* Virus Research, **30** (1993), 335–349

[20] R. Twarock, *A tiling approach to virus capsid assembly explaining a structural puzzle in virology.* Journal of Theoretical Biology, **226** (2004), No 4, 477–482

[21] R. Kerner, *The principle of self-similarity* , in " Current Problems in Condensed Matter", ed. J. Moran-Lopez, (1998), 323–341