

University of *Ljubljana*
Faculty of *Mathematics and Physics*



Department of Physics

Seminar: Symmetries in Physics

Icosahedral Symmetry of Viral Capsids

Anže LOŠDORFER BOŽIČ

Advisor: Doc. Dr. Primož ZIHERL

May 20, 2010

Abstract

In this seminar we discuss from a group theoretical viewpoint the icosahedral symmetry observed in capsids of many viral families. After giving the properties of the icosahedral point group, we present the first effort of explaining the icosahedral symmetry of viral capsids, the famous principle of quasi-equivalence proposed by Caspar and Klug. We then consider and compare two recent models improving on the Caspar-Klug principle, the approach of Lorman and Rochal based on the Landau theory of crystallisation, and the viral tiling theory. Lastly, we briefly comment on the possible origins of the icosahedral symmetry in viral capsids.

Contents

1	Introduction	3
2	Basic structure of viral capsids	3
3	Icosahedral symmetry of viruses and the Caspar-Klug theory	5
3.1	The icosahedral point group	5
3.2	Caspar-Klug principle of quasi-equivalence	8
4	Beyond Caspar and Klug	12
4.1	Landau theory of crystallisation applied to small icosahedral viruses . . .	12
4.2	Mathematical virology: the viral tiling theory	19
4.3	Comparison of the two approaches: L-A and Polyoma viruses	21
5	Origin of icosahedral symmetry in viruses	23
6	Conclusions	25

1 Introduction

When the early techniques of X-ray scattering to determine molecular structures were applied to viruses, the resulting diffraction patterns possessed striking symmetry. It came as an even greater surprise that the observed symmetry of many viruses was that of an icosahedron. This additionally implied that the symmetry of viruses, structures composed of many pieces of one or maybe few different proteins, is thus independent of protein contact details, and that a more general physical principle is hidden behind it.

In this seminar, we shall take a look at the icosahedral symmetry of viruses. First, we will outline the necessary theory of point groups and especially the properties of the icosahedral group. Then we will present the famous principle of Don Caspar and Aaron Klug, explaining the icosahedral symmetry from the viewpoint of quasi-equivalence of structural units. Afterwards, we will take a closer look at a recently proposed improved model for the observed symmetry which applies the Landau crystallisation theory to the structure of small viruses. We will compare this model with the so-called viral tiling theory, taking on the quasi-equivalence principle from another angle. Lastly, we will mention some possible explanations for the origin of icosahedral symmetry in viruses, which is not yet completely known.

2 Basic structure of viral capsids

But before we delve into the symmetries of viruses, we should spare a few words about their structure to help us understand the models and their implications. We will focus mainly on the structure of viral capsids; more details may be found in Ref. [1], wherefrom most of the material in this section is taken.

Even though viruses are entities that depend on other organisms in order to reproduce themselves, they can be found in almost every ecosystem, infecting both eukaryotes (animals and plants) and prokaryotes (bacteria). In spite of myriad different viruses adapted to their specific hosts, all of them have in common their basic structure, that of the *nucleocapsid*.

The nucleocapsid consists of the viral *genome*, enclosed inside the viral *capsid*. A major function of the viral capsid, which is formed from structural proteins, is the protection of the genome as well as recognition and attachment to the host cell, inside which the virus can be replicated. Furthermore, the capsid must also have the ability to alter its conformation in order to release its genome at the appropriate time, whilst being stable enough to survive in extracellular environment.

The nucleic acid of the viral genome can be either a DNA or an RNA molecule, and both may be either single- or double-stranded. The lengths of viral genomes range from very small (some 3 kbp) to lengths comparable with microbial genomes (e.g. 1200 kbp for Mimivirus). An important consequence of this is that smaller viruses with a correspondingly small genome encode only one to at most few structural proteins (*i.e.* different protein species) from which the capsid is constructed, whereas the number of different protein subunits increases with the increasing size of the genome.

2 Basic structure of viral capsids

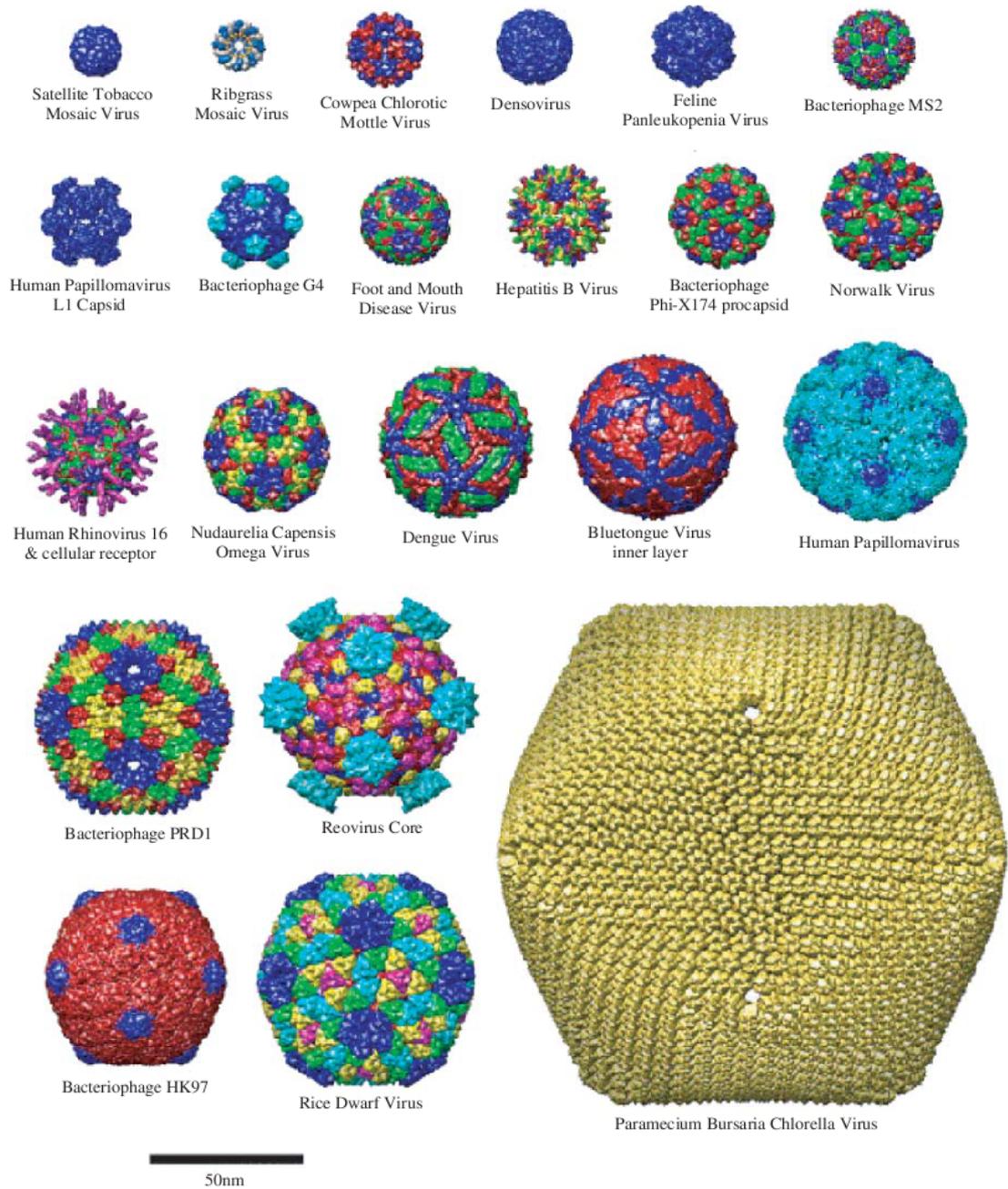


Figure 1: Examples of capsids with icosahedral symmetry [1]. The images were created with the molecular graphics program UCSF Chimera using data from cryo-electron microscopy and X-ray diffraction.

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

The individual protein subunits (*monomers*) are asymmetrical, yet they are organized into morphological subunits (*capsomeres*) which then form capsids possessing great symmetry. The vast majority of the viral capsids have either helical or icosahedral symmetry, though rods and cone-like capsids are also observed. Our focus will be on viruses having icosahedral symmetry, encountered in the capsids of many viral families (Fig. 1). Here, a distinction should be made between the symmetry and the shape of a capsid – icosahedral symmetry does not necessarily imply a shape of an icosahedron. Again, this can be clearly observed in Fig. 1, where we can see that the smaller viruses with icosahedral symmetry have nearly perfect spherical shape, the deviations from which grow larger with increasing capsid radius. Within the context of continuum elasticity theory, this can be explained by a single parameter, termed the Föppl-von Kármán number [2]:

$$\gamma = \langle R \rangle^2 \frac{Y}{\kappa} . \quad (1)$$

Here, $\langle R \rangle$ is the average capsid radius, Y is the two-dimensional Young modulus, and κ is the bending rigidity. This analysis shows that the larger the virus, the greater the influence of icosahedral symmetry on its shape, and such capsids thus possess structural deviations in forms of cones, ridges, ... [3]. On the other hand, the shapes of small viruses can be very well approximated as perfect spheres, with the constituent proteins still retaining icosahedral symmetry.

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

Since icosahedral symmetry is obviously a characteristic feature of numerous viral capsids, it indicates that some general physical principle and not protein-specific interactions is the cause of it. To be able to discuss the possible origins of the symmetry we will in this section first give the basic properties of the icosahedral symmetry from the viewpoint of point groups, and then present the most well-known explanation for the occurrence of such symmetry in viral capsids, the principle of quasi-equivalence put forth by Don Caspar and Aaron Klug.

3.1 The icosahedral point group

We will begin by giving the basic group structure pertaining to objects with icosahedral symmetry. For this purpose, we shall in this section try to follow the notation from Refs. [4] and [5], which is also where the reader should turn for the details of the group theory presented here.

Icosahedral point group \mathcal{Y} is the group of all proper coverings of the icosahedron, and consists of 12 (6 bilateral) five-fold rotation axes C_5 , 20 (10 bilateral) three-fold rotation axes C_3 , and 15 two-fold rotation axes C_2 (Fig. 2). The latter can be generated by consequently applying a C_5 and a C_3 rotation. We exclude inversion symmetry (group \mathcal{Y}_h), as the asymmetry of the constituent protein monomers means that the capsids are chiral in general. Nonetheless, the extension to the full icosahedral group is rather trivial, as we need only include the spatial inversion ($\mathcal{Y}_h = \mathcal{Y} \otimes S_2$).

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

The elements of the icosahedral group can be divided into five conjugacy classes: the identity (E), 12 rotations by $2\pi/5$ ($12C_5$), 12 rotations by $4\pi/5$ ($12C_5^2$), 20 rotations by $2\pi/3$ ($20C_3$), and 15 rotations by π ($15C_2$), respectively. Every other rotation can be generated from these elements; for instance, the rotation by $-2\pi/5$ around one five-fold axis is simply the rotation by $2\pi/5$ around its bilateral axis. Thus, the group order is obviously $g = |\mathcal{Y}| = 60$.

The number of conjugacy classes is equal to the number of irreducible representations of the group [4]. A well-known relation connects the dimensionalities of the irreducible representations s_α with the order of the group:

$$\sum_{\alpha=1}^5 s_\alpha^2 = g \quad . \quad (2)$$

The only possible solution in our case is $s_1 = 1$, $s_{2,3} = 3$, $s_4 = 4$, $s_5 = 5$. We thus have one one-dimensional representation (A), two three-dimensional representations ($T_{1,2}$), one four-dimensional representation (G), and one five-dimensional representation (H). From here the character table for the irreducible representations (Table 1) may be straightforwardly derived, using the orthogonality relations for group characters [4]:

$$\sum_p c_p |\chi_p|^2 = g \quad , \quad (3)$$

$$\sum_p c_p \chi_p^{(\alpha)} = 0 \quad \text{for } \alpha \neq A \quad , \quad (4)$$

$$\sum_\alpha s_\alpha \chi_p^{(\alpha)} = 0 \quad \text{for } p \neq \{E\} \quad , \quad (5)$$

where c_p is the number of elements in class p , and χ_p the character of the elements in the corresponding class.

Table 1: Character table for the icosahedral point group \mathcal{Y} . Here, $\eta_\pm = (1 \pm \sqrt{5})/2$.

	E	$12C_5$	$12C_5^2$	$20C_3$	$15C_2$
A	1	1	1	1	1
T_1	3	η_+	η_-	0	-1
T_2	3	η_-	η_+	0	-1
G	4	-1	-1	1	0
H	5	0	0	-1	1

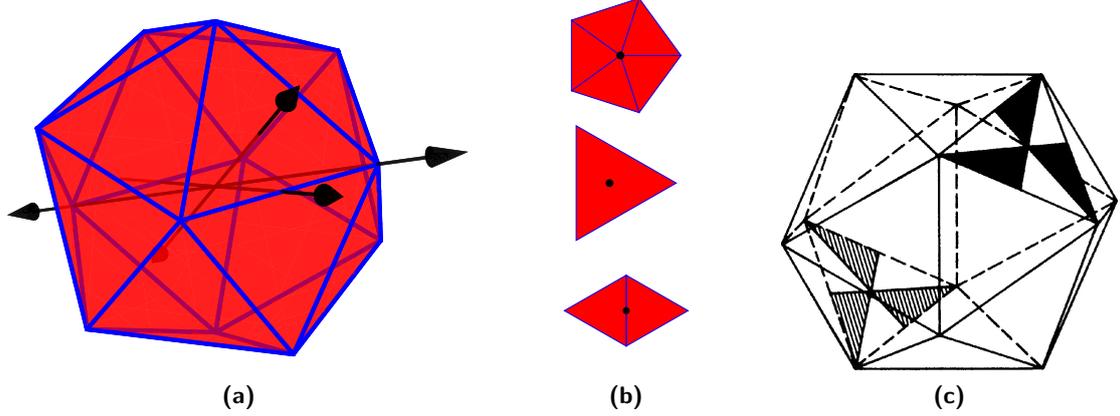


Figure 2: Proper symmetry elements of an icosahedron. The 12 (6 bilateral) five-fold rotation axes pass through the opposite vertices of the icosahedron, the 20 (10 bilateral) three-fold rotation axes pass through the centres of its opposite faces, whilst the 15 two-fold rotation axes pass through the centres of opposite edges. One of each rotation axes is shown in panel (a), and their nature clarified in panel (b). Panel (c) shows the absence of an inversion centre in the icosahedral group \mathcal{Y} [6].

Irreducible representations of $SO(3)$ and their subduction to point groups

In general, a point group \mathcal{G} consisting of proper rotations can be thought of as a subgroup of the special orthogonal group of all rotations in three dimensions, $\mathcal{G} \subset SO(3)$.¹ The irreducible representations $D^{(l)}$ of the special orthogonal group $SO(3)$ are $(2l + 1)$ -dimensional with $l = 0, 1/2, 1, 3/2, 2, \dots$, and are spanned by the spherical harmonics $Y_l^m(\vartheta, \varphi)$:

$$Y_l^m(\vartheta, \varphi) = (-1)^m \left[\frac{(2l + 1)(l - m)!}{4\pi(l + m)!} \right] e^{im\varphi} \sin^m \vartheta \left(\frac{d}{d(\cos \vartheta)} \right)^{l+m} \frac{(\cos^2 \vartheta - 1)^l}{2^l l!}, \quad (6)$$

with parity $PY_l^m = (-1)^l Y_l^m$. The characters of the irreducible representations $D^{(l)}$ are given by [7]

$$\chi^{(l)}(\varphi) = \sum_{m=-l}^l e^{im\varphi} = \frac{\sin((l + 1/2)\varphi)}{\sin(\varphi/2)}. \quad (7)$$

For the vector representation ($l = 1$) this reduces to the known expression $\chi^{(1)}(\varphi) = 2 \cos \varphi + 1$. It should also be mentioned that induced rotations by φ around a given axis \mathbf{k} , $R_{\mathbf{k}}(\varphi)$, do not mix different irreducible representations amongst themselves [7],

$$D^{(l)}(R_{\mathbf{k}}(\varphi))Y_l^m(\vartheta, \varphi) = Y_l^m(\vartheta', \varphi') = \sum_n D_{nm}^{(l)}(\mathbf{k}, \varphi)Y_l^n(\vartheta, \varphi). \quad (8)$$

¹Should we include improper rotations (roto-inversions), such a point group would be a subgroup of the orthogonal group $O(3)$.

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

As $D^{(l)}$ are representations of $SO(3)$, they will also be representations of a point subgroup, albeit in general not irreducible [4, 8]. Therefore, they can be written as a direct sum of the irreducible representations of the point subgroup; in our case, the icosahedral group \mathcal{Y} :

$$D^{(l)} = \bigoplus_{\alpha=1}^5 m_{\alpha} D_{\mathcal{Y}}^{(\alpha)} \quad , \quad (9)$$

where m_{α} are the multiplicities of each irreducible representation occurring in the reduction. Using again the orthogonality relations, the multiplicities can be obtained by [4]

$$m_{\alpha} = \frac{1}{g} \sum_p c_p \chi_p^{(\alpha)*} \chi_p^{(l)} \quad . \quad (10)$$

Little groups and orbits

Another important concept for the understanding of the geometrical approach to capsid construction is that of group orbits [9], so it shall be briefly outlined here. For a given group \mathcal{G} acting on a set of points Ω , its stabilizer subgroup (also called the little group) of an element $\omega \in \Omega$ is the set of all the group elements G that fix ω , $\mathcal{G}_{\omega} = \{G \in \mathcal{G} \mid G \circ \omega = \omega\}$. A closely related notion is that of a group orbit. The orbit of a point $\omega \in \Omega$ is the set of elements of Ω to which ω can be moved by the elements of \mathcal{G} , and is denoted by $\mathcal{G}\omega = \{G \in \mathcal{G} \mid G \circ \omega\}$. More details on this topic may be found in Ref. [8].

A *regular* (trivial) *orbit* is then such an orbit that for any group action on an element in orbit the resulting element is again in the orbit, and that the power of the stabilizer subgroup of each element is one, $\Omega_{reg} = \{\omega \in \Omega \mid G \circ \omega \in \Omega_{reg} \wedge |\mathcal{G}_{\omega}| = 1 \forall \omega \in \Omega_{reg}\}$. For a discrete symmetry group, the number of positions in a regular orbit is equal to the group order g [9]. In the case of an icosahedron, each point in one of the fundamental domains (black scalene triangles in Fig. 2c) can be used to generate its regular orbit. This notion plays an important role in the principle of quasi-equivalence discussed in the next sections, as asymmetric proteins placed in a regular orbit of an icosahedron have equivalent environments – consequently, no more than $|\mathcal{Y}| = 60$ proteins can be placed equivalently to construct a capsid with icosahedral symmetry.

3.2 Caspar-Klug principle of quasi-equivalence

One of the first who attempted to explain the highly symmetrical structure of spherical viruses (such as shown in Fig. 1) were Crick and Watson in the mid 1950s [10]. Based on experimental evidence they argued that these viruses possess the symmetry of a cubic point group – either tetrahedral, octahedral, or icosahedral. Not much later crystallographic studies of Tomato Bushy Stunt virus made by Caspar and Turnip Yellow Mosaic virus made by Klug showed spikes indicating five-fold symmetry (Fig. 3). At first, when each of them separately presented their results indicating that viral capsids are a sort of “surface crystals” with icosahedral symmetry, they were met with scepticism. Yet inspired by Buckminster Fuller’s geodesic domes and motivated by additional experimental

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

data, they then somewhat independently hit upon the idea of near- or *quasi-equivalence* of capsid subunits, and in 1962 they together published a now famous paper considering the geometrical aspects of virus capsid design [11].

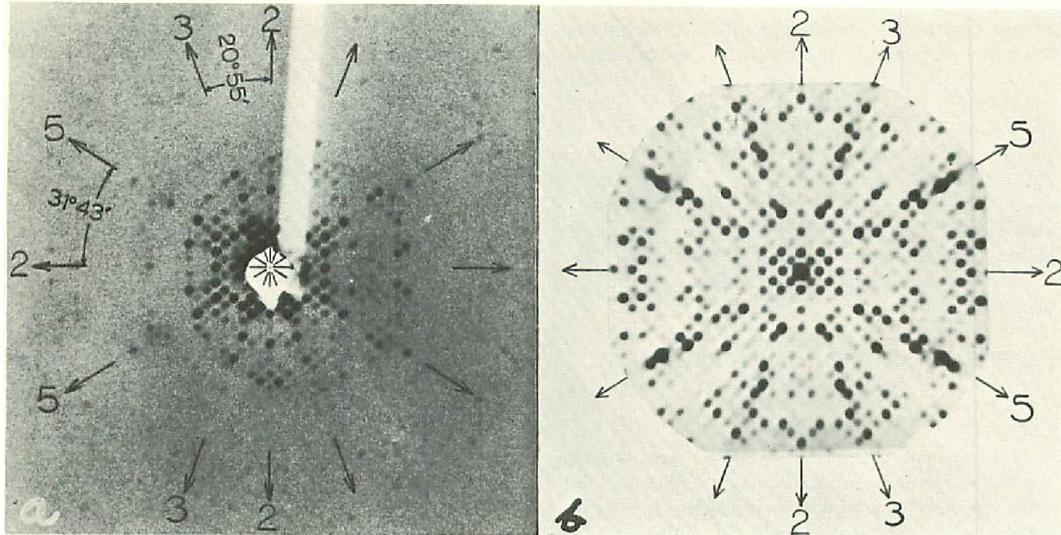


Figure 3: (a) An X-ray diffraction pattern of a poliovirus crystal, made by Finch and Klug (1959), (b) compared with an optical diffraction pattern of 60 points on the surface of a sphere with icosahedral symmetry [11]. Arrows show spikes of high intensity along certain directions, related to five-, three-, and two-fold axes of an icosahedron. Both patterns show the same symmetry relations.

In this paper they focused on simple spherical viruses with a regular structure, the main components being the nucleic acid and capsid proteins. They used three basic properties in the study of viral capsid formation: the lateral type of protein interaction (bonding), the intrinsic curvature of the bond deformation, and especially the asymmetry of capsid proteins [9]. Using the fact that the small viral genome can encode only one or a few types of proteins, Caspar and Klug (CK) formulated the main structural problem in physical virology: *how to construct a regular shell with the icosahedral symmetry formed by identical copies of identical asymmetric proteins* (considered as 2D units).

Such structures are formed by self-assembly, a process akin to crystallisation and governed by laws of statistical mechanics [11]. As in a crystal, molecules in a capsid should be in identical or at least physically indistinguishable environments, since interaction of identical proteins should lead to identical local environments. Arranging identical units in identical environments necessarily produces a symmetrical structure, and there is only a geometrically limited number of kinds of symmetry.

A spherical virus should thus have cubic symmetry, so that no direction in space can be preferred. Only three types of cubic symmetry exist: tetrahedral, octahedral, and icosahedral. The latter allows the greatest possible number of asymmetric units to be placed in equivalent positions (*i.e.* in one of its regular orbits), that is 60; tetrahedral and octahedral point group allow only 12 and 24 units, respectively. However, many

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

viruses contain more than 60 subunits in their shell: in general, their number is equal to $\mathcal{N} = 60\mathcal{T}$, where \mathcal{T} is a positive integer. Their positions belong to different regular orbits of the icosahedron and cannot be equivalent [9, 11]. Therefore, the insistence on strict mathematical equivalence must be dropped, whilst still retaining its physical essentials – a way has to be found to put identical proteins on a surface of a sphere so that they are quasi-equivalently related.

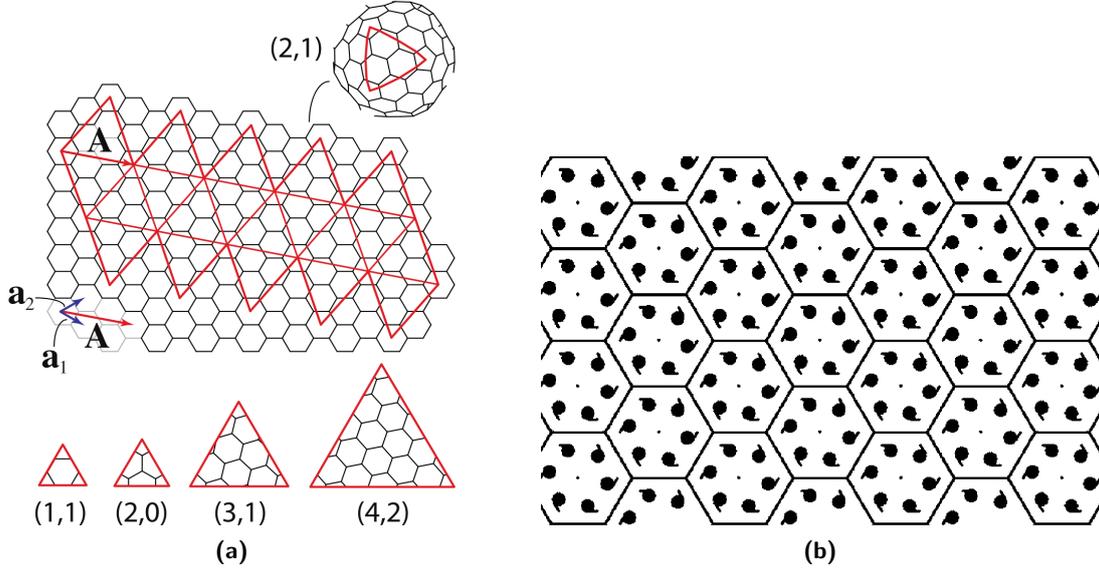


Figure 4: (a) Mapping of the plane hexagonal structure onto the icosahedron surface [12]. A net of an icosahedron containing 20 equilateral triangles can be embedded as a finite piece of a hexagonal tiling, with six asymmetric proteins located around each six-fold centre. Shown is a cut-and-fold construction of a $(h, k) = (2, 1)$ icosahedron, with triangulation number $\mathcal{T} = h^2 + hk + k^2 = 7$. Denoted are the basis vectors of the plane hexagonal lattice, \mathbf{a}_1 and \mathbf{a}_2 , and the vector $\mathbf{A} = h\mathbf{a}_1 + k\mathbf{a}_2$, directed along one of the sides of the icosahedron. The latter vector connects two pentagonal disclinations in the mapping of the hexagonal lattice onto the icosahedron surface. In the bottom of the figure are shown also the triangular faces of $(1, 1)$, $(2, 0)$, $(3, 1)$, and $(4, 2)$ icosahedra. (b) The plane hexagonal structure with one six-fold position filled with asymmetric structural units (proteins), presented as full circles with tails [9].

The task of CK was to find a way to put identical proteins in different but nearly equivalent positions, and to explain the origin of this quasi-equivalence. In contrast with the restriction of positions in regular orbits of 3D point groups, the translational symmetry of a 2D lattice makes infinite the number of positions belonging to the same regular orbit of a 2D crystal space group. CK were then looking for an *almost* regular mapping of the plane periodic structure onto the icosahedron surface. Mapping splits one regular orbit of the plane periodic structure into several different orbits of the icosahedron but maintains some “traces” of their former equivalence in the plane structure, thus making the positions in different orbits “quasi-equivalent” [9]. Symmetry analysis shows that the only type of crystalline order suitable for the almost regular mapping is the

3 Icosahedral symmetry of viruses and the Caspar-Klug theory

plane hexagonal structure with one six-fold regular position in the unit cell filled with asymmetric proteins.

A net of an icosahedron which consists of 20 equilateral triangles can be embedded as a finite piece of the considered plane hexagonal structure and then mapped (folded) onto the icosahedron surface (Fig. 4a). The mapping is chosen in such a way that the vertices of the regular triangular faces of the icosahedron coincide with the six-fold axes of the plane hexagonal lattice. By cutting out twelve 60° sectors, one can transform six-fold axes into five-fold ones, and then join the sector edges on the icosahedron surface. Joining of the edges represents a continuous matching operation as the edges of the cut-out sectors are equivalent by the symmetry of the hexagonal structure. This remarkable symmetry property allowed CK to state that their folding is done “without destroying the bonding pattern of the lattice” [11]. From a purely geometrical perspective, similar mapping can be achieved with a plane trigonal structure with three- and not six-fold symmetry axes. However, in such a case the joining of the edges would not be a matching but a “stitching” operation, leading to a seam formation.

As seen from Fig. 4a, the edge length of the icosahedron face for a given mapping is determined by the vector joining the vertices of the net triangles, situated in the nodes of the hexagonal lattice,

$$\mathbf{A} = h\mathbf{a}_1 + k\mathbf{a}_2 \quad , \quad (11)$$

where \mathbf{a}_i are the basis (Bravais) vectors of the hexagonal lattice, and h and k are some non-negative integers. The square of this vector length,

$$\mathcal{T} = h^2 + hk + k^2 \quad (12)$$

is equal to the number of lattice nodes contained in two net triangles – that is, the number of hexagons enclosed in them. As a consequence, the number of protein units in a CK structure with *triangulation number* \mathcal{T} is given by

$$\mathcal{N} = 60\mathcal{T} \quad . \quad (13)$$

Each of the CK structures can be thus characterised by the two non-negative integers h and k . These can be also thought of as the number of jumps through the vertices of a mapping that need to be performed in order to reach a center of a pentagon from its neighbouring pentagon. The structures with $h \neq k$, $h, k > 0$ are chiral (skew), meaning that their mirror image has different symmetry; a mirror image of a (h, k) mapping is a (k, h) mapping. To discriminate between the two, the triangulation number is sometimes used with words *laevo* (left) and *dextro* (right), and denoted by \mathcal{T} and \mathcal{T}_d , respectively [12].

Caspar and Klug also introduced the concept of a morphological unit (capsomere) [9, 11]: in the plane hexagonal structure, all protein positions belong to one six-fold regular position in the unit cell, and can be divided into groups of six positions situated around the same lattice node (Fig. 4b). The corresponding six proteins can be then considered as a morphological unit, which is a hexamer in the considered case. The CK geometrical model maps both the lattice nodes and the protein positions onto the icosahedron surface.

4 Beyond Caspar and Klug

After the mapping, the number of positions around an image of a node can be either six or five, and the corresponding morphological units can be considered as hexamers and pentamers, respectively. The number of pentamers situated in the five-fold axes of the icosahedron is 12, so that there are $10(\mathcal{T} - 1)$ hexamers situated around other nodes.

All in all, the Caspar-Klug principles may be summed up as follows [13]:

- Since the proteins are asymmetric, the capsid symmetry is lower than that of the regular icosahedron – we have no inversion symmetry or mirror planes, only rotational symmetry (group $\mathcal{Y} \subset \mathcal{Y}_h$).
- The asymmetric proteins can be located only in the regular 60-fold positions of the icosahedral point group \mathcal{Y} , and so the number of proteins in a capsid is always a multiple of 60,

$$\mathcal{N} = 60\mathcal{T}, \quad \text{where } \mathcal{T} \in \mathbb{N} \quad , \quad (14)$$

where \mathcal{T} is the triangulation number.

- The self-assembly of viral capsids is a process akin to crystallisation, and is governed by the laws of statistical mechanics.
- The proposed model of quasi-equivalent constructions is based on an almost regular mapping of a hexagonal lattice onto an icosahedron surface. Thus, only those values of \mathcal{T} and consequently \mathcal{N} are allowed which satisfy the relation

$$\mathcal{T} = h^2 + hk + k^2 \quad , \quad (15)$$

for some non-negative integers h and k .

4 Beyond Caspar and Klug

The Caspar-Klug principle of quasi-equivalence was deemed to be true of all spherical viruses, until it was discovered in the early 1980s that the Polyoma virus capsid has 360 subunits – a number forbidden by the selection rule [10]. More exceptions were found later on (see *e.g.* Ref. [13] and references therein), and whilst most spherical viruses still fall under the quasi-equivalent construction it became clear that a more general principle is needed for their description. Recently, two such descriptions emerged: one approaches the problem with the tools of statistical mechanics of crystallisation [9, 13, 14], whereas the second is still based on the geometrical coverings of the icosahedron [15, 16, 17]. In the following sections we shall take a more detailed look at the former, only briefly comparing it with the latter.

4.1 Landau theory of crystallisation applied to small icosahedral viruses

The assembly of viral capsids has similarities to micelle formation as well as crystallisation [18, 19]. The latter process is well described by Landau theory of crystallisation [20], which describes the change of the symmetry of a system during a phase transition. When

4 Beyond Caspar and Klug

considering crystals as well as capsids, this phase transition changes the symmetry from an isotropic to a lower one², and Lorman and Rochal have used this theory to explain the formation of capsids with icosahedral symmetry which do not necessarily fall within the CK classification, but have nevertheless been observed experimentally [9, 13, 14].

We characterise the system by introducing a density function $\rho(\mathbf{r})$ which gives the probability distribution of various positions of the proteins within the capsid [9, 20]. The symmetry of the capsid is thus defined as the group of all coordinate transformations under which this density is invariant. We may write

$$\rho = \rho_{iso} + \Delta\rho \quad , \quad (16)$$

where ρ_{iso} is the density distribution in the isotropic phase, invariant under $SO(3)$ symmetry, and $\Delta\rho$ is the symmetry-breaking term induced by the ordering. Similarly to the case of atomic positions in simple atomic crystals, the positions of the maxima of the corresponding density function are associated with protein centres in the capsid shell [13].

As we are interested only in *small* icosahedral viruses, their shape being almost perfectly spherical, we can take for basis functions the spherical harmonics,

$$\Delta\rho(\vartheta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm} Y_l^m(\vartheta, \varphi) \quad , \quad (17)$$

which also span the irreducible representations of the $SO(3)$ group. After the assembly of proteins into capsid, the density distribution acquires icosahedral symmetry $\mathcal{Y} \subset SO(3)$. This symmetry breaking is associated with one critical order parameter which spans an irreducible representation of the symmetry group of the disordered state [13, 20]. In the vicinity of the crystallisation point the structure of the ordered state defined by $\Delta\rho$ is determined by the critical order parameter only, the contributions of non-critical degrees of freedom being negligible in this region. As in the crystallisation process, the order parameter for the assembly process on a sphere represents a critical system of density waves (CDSW), the critical part of density being determined by a CDSW with the same wave number l [9, 20],

$$\Delta\rho_l(\vartheta, \varphi) = \sum_m A_{lm} Y_l^m(\vartheta, \varphi) \quad . \quad (18)$$

Additionally, since we do not wish for the capsid to have inversion symmetry, the parity of the harmonics has to be negative, thus limiting possible values of l to only odd ones, $l = \text{odd}$.

The free energy expansion of the assembly process is given in standard form as $\mathcal{F} =$

²For instance, one of the point or space groups.

4 Beyond Caspar and Klug

$\mathcal{F}_0 + \mathcal{F}_2 + \mathcal{F}_3 + \dots$, and contains only invariant terms [9]:³

$$\begin{aligned}\mathcal{F}_2 &= A(T, c) \sum_{m=-l}^l A_{lm} A_{l-m} \quad , \\ \mathcal{F}_3 &= B(T, c) \sum_{m_1, m_2, m_3} a_{m_1 m_2 m_3} A_{lm_1} A_{lm_2} A_{lm_3} \delta(m_1 + m_2 + m_3) \equiv 0 \quad , \\ \mathcal{F}_4 &= \sum_k C_k(T, c) \sum_{m_1, m_2, m_3, m_4} a_{m_1 m_2 m_3 m_4}^k A_{lm_1} A_{lm_2} A_{lm_3} A_{lm_4} \delta(m_1 + m_2 + m_3 + m_4) \quad .\end{aligned}$$

Here, a_i are the weight coefficients of the $SO(3)$ group, and $A(T, c)$, $B(T, c)$, and $C_k(T, c)$ are temperature- and composition-dependent coefficients of the Landau theory [20]. For odd wave numbers l the third-order term is identically zero, making the thermodynamics of asymmetric proteins assembly different with respect to the thermodynamics of atomic cluster formation, in spite of several common points in their formal description [13].

As seen in the previous sections, the irreducible representations $D^{(l)}$ of $SO(3)$ are also reducible representations of the icosahedral group \mathcal{Y} [Eq. (9)]. The formed capsid has to be invariant under all symmetry elements of the icosahedral group, and therefore the only possible representations $D^{(l)}$ are those that feature in their reduction at least one identical representation A . From Eq. (10) we have

$$m_A(l) = \frac{1}{60} \left[(2l + 1) + 12\chi^{(l)}(2\pi/5) + 12\chi^{(l)}(4\pi/5) + 20\chi^{(l)}(2\pi/3) - 15 \right] \neq 0 \quad , \quad (19)$$

where the characters $\chi^{(l)}(\varphi)$ are given by Eq. (7). In this way we obtain for the possible wave numbers

$$l = 15, 21, 25, 27, 31, 33, 35, \dots \quad . \quad (20)$$

The same wave numbers are obtained from analysis based on the theory of invariants,⁴ where it can be shown that any critical order parameter driving the icosahedral assembly of asymmetric proteins has the wave number l satisfying the equation [9]

$$l = 15 + 6i + 12j \quad , \quad (21)$$

where i and j are non-negative integers. This sequence of permitted wave numbers determines the possible capsid shell structures for small icosahedral viruses.

The explicit form of the critical density function $\Delta\rho_l$ is then given by the basis functions $f_l^k(\vartheta, \varphi)$, $k = 1, 2, \dots, m_A(l)$, of all m_A totally symmetric representations of the icosahedral group \mathcal{Y} in the restriction of the active irreducible representations of the

³No linear invariant can be formed from quantities which are transformed according to a non-unit irreducible representation of a group, for otherwise that representation would be reducible [20].

⁴Into which we shall not delve here, as it is not necessary for the understanding of the problem.

At the same time, it is heavily technical [21], so let us just mention that the relations for the allowed wave numbers are obtained by considering the full set of generators for the ring of invariant polynomials. The density function should then be a homogeneous polynomial of the l -th degree, and the homogeneity requirement gives the possible combinations.

4 Beyond Caspar and Klug

$SO(3)$ [Eq. (9)]. The CSDW is some linear combination of these functions invariant with respect to the icosahedral group,

$$\Delta\rho_l(\vartheta, \varphi) = \sum_{k=1}^{m_A} B_k f_l^k(\vartheta, \varphi) \quad , \quad (22)$$

where B_k are arbitrary coefficients. For small icosahedral capsids, the construction of the protein density distribution is simplified because the basis in Eq. (22) contains only one function. Specifically, Eq. (19) yields

$$m_A(l) \leq 1 \quad \text{for } l \leq 43 \quad , \quad (23)$$

so that

$$\Delta\rho_l(\vartheta, \varphi) = B f_l(\vartheta, \varphi) \quad \text{for } l \leq 43 \quad . \quad (24)$$

In such cases the positions of the maxima of the density distribution do not depend on the coefficient B , but are generated by a single function $f_l(\vartheta, \varphi)$ which has no fitting parameters. The explicit form of the symmetry adapted *irreducible icosahedral density function* $f_l(\vartheta, \varphi)$ for a given wave number l is obtained by averaging of spherical harmonics over the icosahedral symmetry group [9],⁵

$$f_l(\vartheta, \varphi) = \frac{1}{60} \sum_{G \in \mathcal{Y}} T(G) Y_l^m(\vartheta, \varphi) \quad . \quad (25)$$

For any fixed value of m this procedure gives either the same function $f_l(\vartheta, \varphi)$ we are looking for or zero. Since the functions are defined up to a constant complex factor, an appropriate choice makes the functions f_l real. The above equation may be rewritten as [7]

$$f_l(\vartheta, \varphi) = \frac{1}{60} \sum_{G \in \mathcal{Y}} \sum_{m'} D_{m'm}^{(l)}(G) \cdot Y_l^{m'}(\vartheta, \varphi) \quad , \quad (26)$$

and the actions of group elements are given in terms of the Euler angles (α, β, γ) [22, 23, 24]:

$$D_{m'm}^{(l)}(\alpha\beta\gamma) = e^{-im'\gamma} d_{m'm}^l(\beta) e^{-im\alpha} \quad , \quad (27)$$

$$d_{m'm}^l(\beta) = \sum_k \frac{(-1)^{k-m+m'} \sqrt{(l+m')!(l+m)!(l-m')!(l-m)!}}{(l-m'-k)!(l+m-k)!k!(k-m+m')!} \times \quad (28)$$

$$\times \cos^{2l+m-m'-2k}(\beta/2) \sin^{2k+m'-m}(\beta/2) \quad , \quad (29)$$

where k runs from $k = \max(0, m-m')$ to $k = \min(l-m', l+m)$. The definition of Wigner matrices $D_{m'm}^l$ follows the one given in Ref. [24] for the zyz Euler angles convention. The Euler angles for the proper icosahedral group \mathcal{Y} can be found in Ref. [23]. Thus we

⁵In fact, a projection of a spherical harmonic onto the identical representation.

4 Beyond Caspar and Klug

finally obtain in the expansion of the irreducible icosahedral density function over the spherical harmonics [Eq. (18)] the expression for the expansion coefficients A_{lm} :

$$A_{lm'} = \frac{1}{60} \sum_{G \in \mathcal{Y}} D_{m'm}^{(l)}(G(\alpha, \beta, \gamma)) \quad , \quad (30)$$

up to a normalisation factor. We may also easily verify whether such a function is truly invariant, as the set of coefficients must then satisfy

$$A_{lm'} = \sum_m D_{m'm}^{(l)}(G(\alpha, \beta, \gamma)) A_{lm} \quad \forall G \in \mathcal{Y} \quad . \quad (31)$$

With the solution for the irreducible icosahedral density function f_l , invariant under the icosahedral group, the positions of the asymmetric capsid proteins are entirely determined by the maxima of the density function. Due to 60 equivalent fundamental domains of the icosahedron we may search for the maxima only within one such domain, and obtain the positions within the other domains by application of the group elements. This number of different density maxima within the domain also has a clear connection with the CK triangulation number \mathcal{T} , and the total number of maxima of the density function equals $\mathcal{N} = 60\mathcal{T}$.

In this way we obtain the possible configurations of the capsid proteins. Several examples are given in Fig. 5, where, for the sake of clarity, only the positive part of the density function $f_l(\vartheta, \varphi) \geq 0$ is shown. On one hand, the distributions in Fig. 5a (e.g. Satellite Tobacco Mosaic virus), Fig. 5d (e.g. Cowpea Chlorotic Mottle virus), and Fig. 5e (e.g. Sindbis virus) give classical CK structures [13]. However, the distributions in Fig. 5b and Fig. 5f do not satisfy the CK selection rules for the triangulation number, and the distribution for $l = 25$ with $\mathcal{T} = 3$ (e.g. Dengue virus, Fig. 5c) shows no hexagonal arrangements of proteins and cannot be obtained by the CK model.

The protein environments as predicted by the Landau crystallisation theory violate the CK geometrical model and are not quasi-equivalent in the strict sense [9]. However, they do not violate the physical equivalence of the protein positions induced by the single irreducible density function. Each position in these constructions has five or six neighbours and the distances between these neighbours are approximately equal. In other words, if the asymmetric identical building blocks can be slightly deformed (as also assumed in the original CK theory), then there is no problem to put them together in the structure in slightly different environments [9].

Not only does the Landau theory of crystallisation predict capsid structures with all values of \mathcal{T} (in contrast with the CK theory), it also predicts several qualitatively different protein distributions for capsids with the same number of protein positions. An example for $\mathcal{T} = 3$ can be seen in Figs. 5c and 5d. This makes it possible to apply the theory to describe a reconstructive structural transformation of capsids, a possibility not present in the CK principle [14].

The reconstructive structural transformation is a process where the protein positions are rearranged during the maturation of the procapsid [1, 14]. This also involves different specific, often irreversible, biochemical features. In contrast to the procapsid formation,

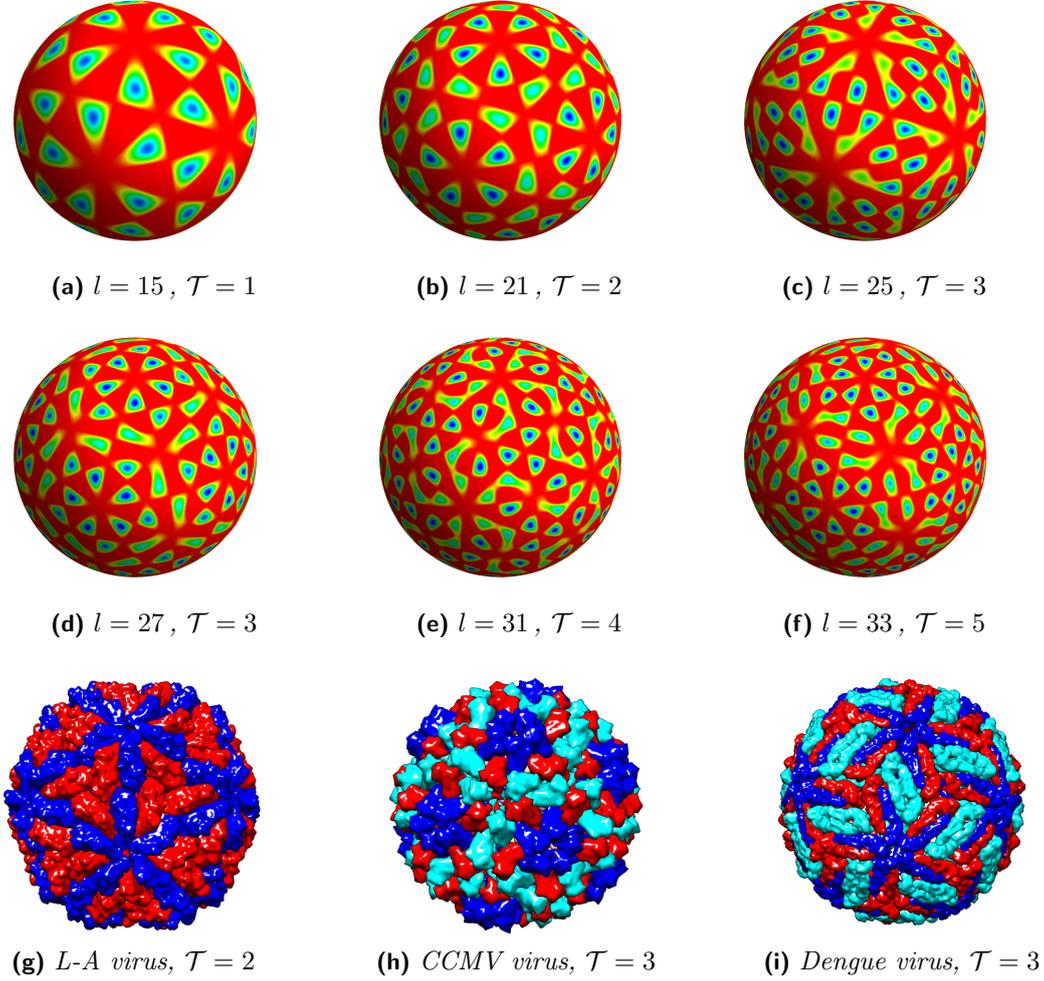


Figure 5: (a) – (f) The first six irreducible icosahedral density functions $f_l(\vartheta, \varphi)$ with wave numbers $l = 15, 21, 25, 27, 31, 33$. The corresponding numbers of different 60-fold positions of density maxima are $\mathcal{T} = 1, 2, 3, 3, 4, 5$. The distributions with $l = 15, 27, 31$ give classical CK structures, whilst the distributions with $l = 21, 25, 33$ cannot be obtained by the CK mapping. (g) – (i) Experimental capsid structures for several viruses (not to scale). The capsid of the CCMV can be classified as a CK structure with $\mathcal{T} = 3$ and is given by the density function with $l = 27$. On the other hand, the capsids of *L-A* and *Dengue* viruses cannot be obtained by the CK model, but are well described by density functions with $l = 21$ and $l = 25$, respectively. The experimental structures were rendered with the molecular graphics package UCSF Chimera [25].

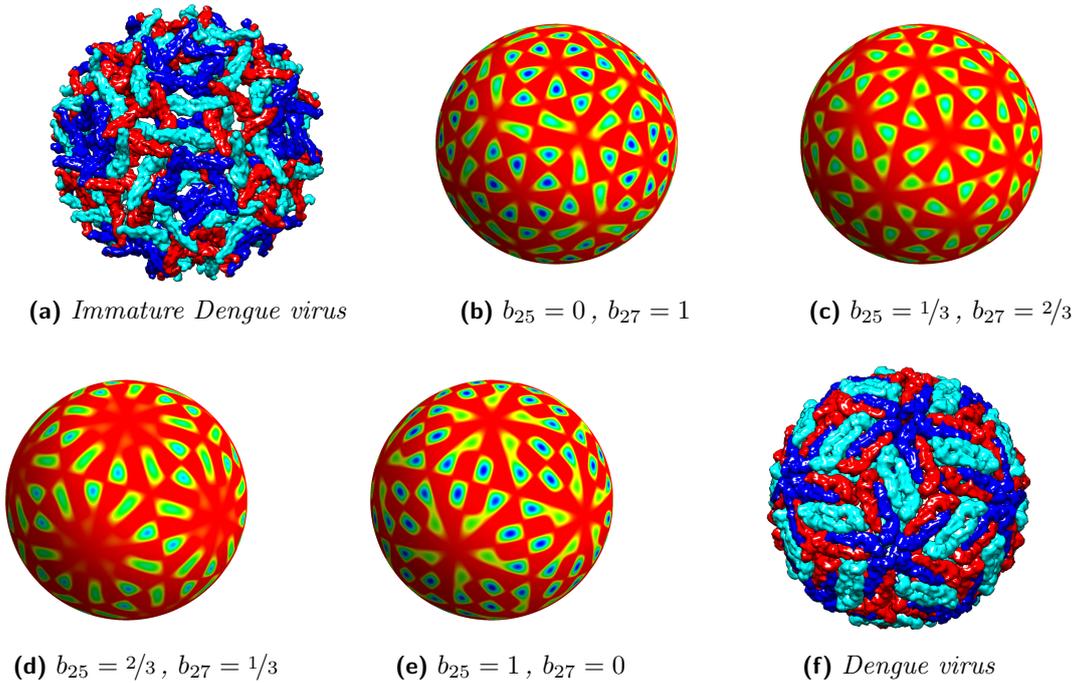


Figure 6: Variation of the protein density function in the intermediate states of the procapsid-capsid transformation in the Flavivirus family. The density function is chosen in the form $\Delta\rho = b_{27}f_{27} + b_{25}f_{25}$ with $b_{25} + b_{27} = 1$. The experimental structures of the Dengue virus were rendered with the molecular graphics package UCSF Chimera [25].

which displays universal properties, biochemical changes during maturation are not the same for different virus families. As it was shown that in the Flavivirus family (Dengue virus, West Nile virus, . . .) immature capsids first undergo reversible structural changes that then render them accessible to irreversible protein cleavage, this indicates that in this family the structural transformation from procapsid to a capsid is a reversible physical process [26, 27]. Thus, a continuous thermodynamical description of the reconstructive transformation in the capsids of the Flaviviridae may be given [14]. It is also important to note that the point symmetry of the protein shell remains icosahedral in all intermediate states of the reconstructive structural transformation in Flaviviridae, a characteristic observed in recent experiments [26, 27]. Conservation of the icosahedral symmetry during the transformation means that the protein density function in the intermediate states is a combination of

$$\Delta\rho = b_{27}f_{27} + b_{25}f_{25} \quad , \quad \text{where} \quad b_{25} + b_{27} = 1 \quad , \quad (32)$$

with the relative weight of the two functions being a thermodynamical variable depending on the external parameters (Fig. 6). A detailed discussion of this mechanism, along with the possibility of identifying the minima of the irreducible density function with the binding sites on the capsid surface, is given in Ref. [14].

Lastly, the approach of Lorman and Rochal to capsid assembly and structure has

another consequence: due to the absence of a cubic term in the free energy \mathcal{F} the icosahedral capsid assembly can be a second order phase transition [13]. This means that there is no latent heat involved in the transition, and the assembly takes place without a nucleation process. The latter feature implies that in equilibrium either intact virus shell or free proteins are dominant species while assembly intermediates are found only in trace concentrations. Nonetheless, no clear conclusion concerning the assembly kinetics can be drawn on the basis of the symmetry analysis alone [9].

4.2 Mathematical virology: the viral tiling theory

The other approach to improving the Caspar and Klug construction is based on re-considerations of its geometrical principles. As mentioned, the CK principle of quasi-equivalence may be thought of as a mapping of a hexagonal lattice onto a sphere. The viral tiling theory differs from the CK construction by introducing more general types of surface lattices, inspired by the theory of quasicrystals [16]. The hexagonal regular tiling in two dimension is extended to non-regular (Penrose) tilings that fill the space, permitting tessellations in shapes other than hexagons. An example of such a tiling is shown in Fig. 7.

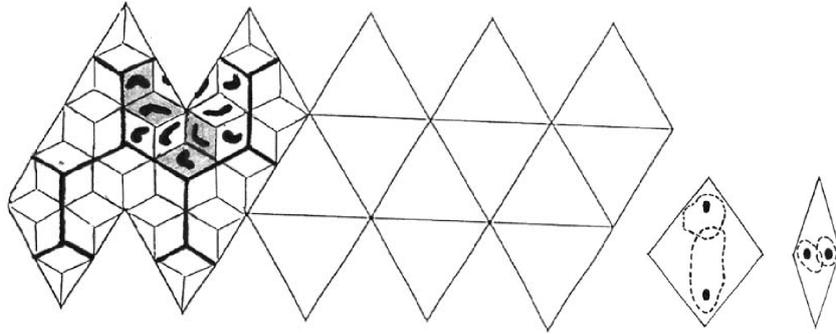


Figure 7: *An embedding of an icosahedral net into an irregular tiling, consisting of two types of rhombs [15]. The protein positions are shown as dotted decorations, and the dashed lines show how the proteins are positioned in the special case of the L-A virus.*

The viral tiling (VT) theory does not build upon the principle of quasi-equivalence in a strict sense as it allows for tessellations of spherical geometric objects with more than one building block (tile), as long as the tessellation preserves the overall icosahedral symmetry. Additionally, the tessellation must also abide by the generalised principle of quasi-equivalence: on any given tile protein subunits are located only at corners subtending the same angle (Fig. 8a) [15]. This assumption ensures that identical types of protein subunits can only occupy structurally or mathematically equivalent sites on the tiles. Mathematically, the generalised principle of quasi-equivalence manifests itself in the restriction on the tile decoration, that is, the prescription for the location of protein subunits on the tiles. The Caspar-Klug tessellation corresponds to triangle tilings, whereas another important class is given by the Penrose tilings, having the shapes of kites, darts, and rhombs [15].

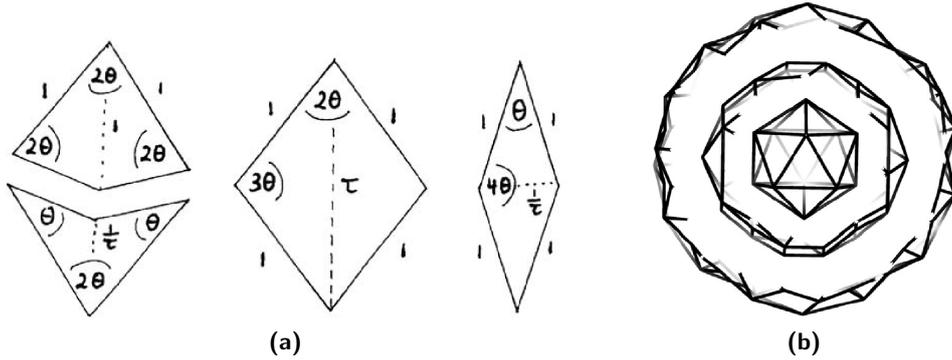


Figure 8: (a) Possible tilings in two dimensions which yield three-dimensional polyhedra compatible with the icosahedral symmetry [15]. Shown are the dart, kite, and rhomb tilings; a triangular tiling would again yield the CK constructions. The protein positions are restricted to decorations of the tiles along the edges subtending a same angle. (b) Shapes of the three polyhedra in the triacontahedral series, giving rise to structures not included within the CK quasi-equivalence principle [17].

In two dimensions, the Penrose tilings are obtained via projection from a regular lattice in five dimensions. Similarly, tilings relevant for the description of viruses can be obtained via projection from a suitable lattice in six dimensions, which is the smallest dimension in which a lattice invariant under the icosahedral group can occur [16, 28]. The polyhedra symmetric under the icosahedral group are obtained with a so-called affine extension of the non-crystallographic Coxeter group H_3 , given by a projection from the group D_6 [28, 29]. More details on this procedure along with examples are given in Refs. [17] and [28]. In this manner, finite dimensional point sets $\mathcal{S}(N)$ can be derived, defining nested point sets in three dimensions related to a six-dimensional lattice via the projection formalism mentioned above [16]. Here, N is termed the cut-off parameter because it limits the number of points in the set. By construction, these point sets contain the vertices of the desired polyhedra or tilings. For $N = 5$, a series of three polyhedra is obtained, containing a rhombic triacontahedron (or an icosahedron), a snub cube (which has octahedral symmetry), and a snub dodecahedron [28].⁶ From these polyhedra coordinates of the vertices (and the corresponding tilings) for the three cases can be obtained, and these structures can be matched with the capsids within the Polyomaviridae family [17]. The relations of the three tilings obtained are shown in Fig. 8b, and the relative radii of the structures match the measured values within the experimental accuracy [17].

In addition to specifying the locations of protein subunits in the capsids as in the CK principle, the VT theory is also used to predict the locations of inter-subunit bonds between proteins in different capsomeres [16]. The VT theory can thus model dimer and trimer interactions, these being different bonding environments of the protomers.

⁶It should be mentioned that in the latter two cases, the edge distances are not the same as one would expect for the true polyhedra; however, the deviations are rather small.

An example of this is provided in the following section. Another extension of the viral tiling theory to multi-level tilings was also made in order to include the possibility of crosslinking in some viral capsids, *i.e.* interconnecting of the protein subunits, resulting in a chainmail organisation covering the entire capsid [30].

4.3 Comparison of the two approaches: L-A and Polyoma viruses

Both the viral tiling theory and the Landau crystallisation approach provide an explanation for the non-CK structures of viral capsids, yet they consider the problem from different perspectives. An in-depth comparison would be beyond the scope of this seminar, but we will now sketch a comparison of the theories on the examples of two viruses, the L-A virus and the Murine Polyoma virus.

L-A virus

The first example of a virus which is not included in the CK principle of quasi-equivalence is the L-A virus, built from 120 GaG proteins organised into 60 dimers. According to the CK principle, such a virus has a forbidden triangulation number of $\mathcal{T} = 2$. The L-A virus symmetry is usually still classified as a $\mathcal{T} = 1$ virus, but with the basic morphological unit now being a dimer rather than a monomer [31, 32]. This means that the icosahedron net is no longer triangular, as can be seen from the grid in Fig. 9b. However, within the frameworks of the Landau crystallisation theory and the viral tiling theory, the resulting structure can be explained differently.

From the viewpoint of the VT theory, the structure of the L-A virus can be viewed as a rhomb tiling, with tiles and decorations as shown in an inset of Fig. 9b (the same structure is also shown in Fig. 7). On the other hand, the approach of Lorman and Rochal clearly allows all values of triangulation numbers, and the only icosahedral density function with $\mathcal{T} = 2$ shows nice correspondence with the experimental structure of the virus (Figs. 9c and 9a).

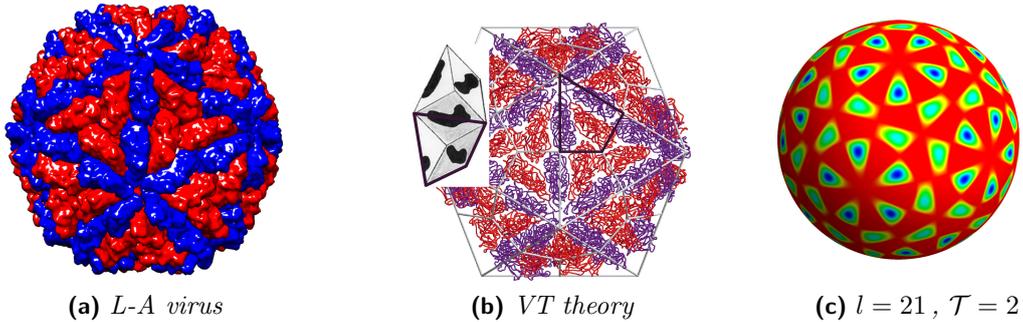


Figure 9: (a) *Experimental structure of the L-A virus capsid*, (b) *rhomb tiling of the capsid (inset) within the $\mathcal{T} = 1$ classification of the symmetry [32]*, (c) *and the irreducible icosahedral density function $f_{21}(\vartheta, \varphi)$ as obtained with the Landau crystallisation theory. The experimental structure of the virus was rendered with the molecular graphics package UCSF Chimera [25].*

Murine Polyoma virus

Another interesting case is the capsid of the Murine Polyoma virus which consists of 360 subunits, and is classified as a $\mathcal{T} = 7_d$ capsid [31] even though the number of subunits implies a triangulation number of $\mathcal{T} = 6$, again a forbidden number within the CK construction. Additionally, all the protein subunits are observed in clusters of pentamers (Fig. 10a), contrary to what the CK principle predicts.

The tessellation within the viral tiling theory is in this case based on rhombs and kites, and is shown in Fig. 10b superimposed on the 7_d CK hexagonal lattice. In this case the tiling approach also predicts the orientation and bonding of protein subunits (based on the three different tile decorations, the yellow-yellow rhomb, the red-blue rhomb, and the green-purple-white kite) [15]. In the context of the Landau crystallisation theory, the capsid structure is described by the $l = 37$ irreducible icosahedral density function with $\mathcal{T} = 6$ (Fig. 10c). The positions of the different maxima agree well with the experimentally observed structure.

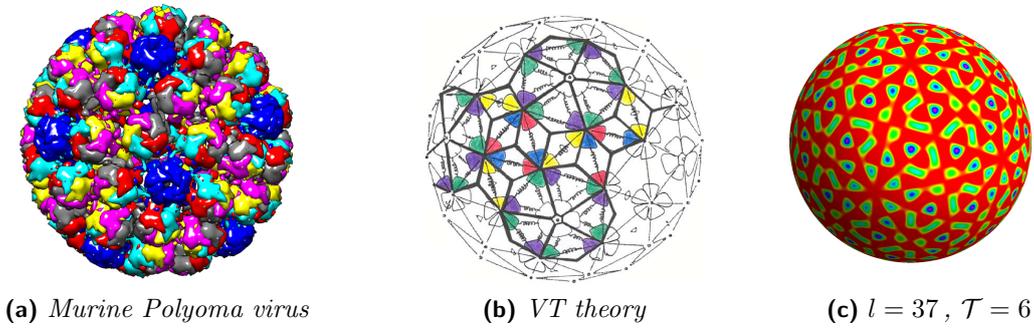


Figure 10: (a) *Experimental structure of the Murine Polyoma virus capsid, (b) kite and rhomb tiling of the capsid as predicted by the viral tiling theory [15], (c) and the irreducible icosahedral density function $f_{37}(\vartheta, \varphi)$ as obtained with the Landau crystallisation theory. The experimental structure of the virus was rendered with the molecular graphics package UCSF Chimera [25].*

Discussion

As we have seen, both the VT theory and the Landau crystallisation approach produce several capsid structures with icosahedral symmetry that are not included within the CK principle. However, there are certain differences between the two approaches. It is interesting to note that the VT theory claims predicting the correct bonding patterns in the case of Murine Polyoma virus, whereas nothing is said of the matter in the case of the L-A virus. Consequently, protein bonding prediction does not appear to be a general property of the VT theory; on the other hand, the theory of Lorman and Rochal makes no such claims altogether.

It should also be mentioned that the concept of the basic morphological unit being a monomer or a dimer must principally be based on the energetics of the inter-protein

bonds, and not on pure geometrical considerations and convenience [12]. The dimer and trimer interactions in the VT theory do not seem to follow from the generalised quasi-equivalence principle, which restricts only the possible decorations of the tiles. The Landau crystallisation theory predicts well the positions of the proteins in the capsid, but perhaps also a connection should be made, or at least considered, between the properties of the irreducible icosahedral density functions and the relative orientations and bonds between the protein subunits.

Each of the two theories gives some new predictions of their own. The VT theory for instance predicts the possibility of cross-linking protein subunits in some of the phage capsids, and the approach of Lorman and Rochal predicts a structural transformation of a capsid that some viruses can undergo. And lastly, there are also some capsid structures (for example that of the Dengue virus) which are obtained by the Landau crystallisation theory, but have not as of yet been considered by the viral tiling theory.

The main objection to the viral tiling theory would therefore be that, in contrast to the tiling decorations, protein positions and inter-protein bonds do not follow from a general principle. The theory thus predicts some new structures, but so far not all of the ones that the theory of Lorman and Rochal has covered. The latter theory appears to have a sounder foundation, but it would be necessary to explore in greater detail what exactly are the implications of the theory besides the new icosahedral capsid structures. Even so, the expansion of the CK classification by the two theories is an achievement in itself.

5 Origin of icosahedral symmetry in viruses

Whereas the new approaches to the construction of icosahedral viruses improve the model proposed by Caspar and Klug, which in itself describes well a number of different capsids, the origin of icosahedral symmetry in viruses and the physical principles underlying the quasi-equivalence principle have yet to be fully elucidated [33].

The icosahedral point group is the one that generates the maximum enclosed volume for shells composed of a given number of subunits [11], but the fact that many viruses self-assemble under *in vitro* conditions indicates that the icosahedral symmetry should be a generic feature of the free energy minima of aggregates of viral capsid proteins [33]. An interesting fact is that the models of capsid assembly which dealt with identical capsomeres regularly produced capsid symmetries lower than icosahedral.⁷ These models range from the closest packing of N disks (capsomeres) on the surface of a sphere, which is in turn related to the Thomson problem of repelling charges, to the self-assembly of adhering hard disks [33, 34, 35, 36].

Several models considered optimal equilibrium structures of the assembled capsids, not describing individual subunits but focusing on the morphological units (capsomeres) [33, 36]. In contrast with individual protein subunits whose interactions can be asymmetric and species-specific, capsomeres interact through a more isotropic and generic interaction

⁷As we have noted before, the tetrahedral and octahedral group are also sufficient, and the latter is, in fact, sometimes observed as the energetic minimum [33].

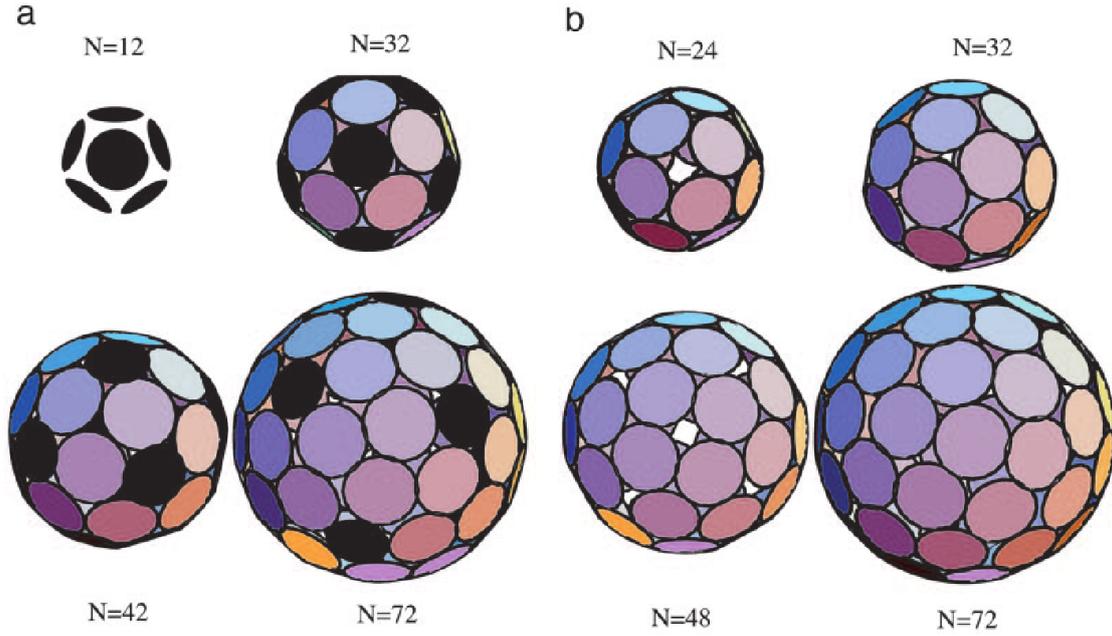


Figure 11: *Minimum energy structures of capsid assembly from N capsomeres (pentamers, shown in black, and hexamers), produced by Monte Carlo simulations [33]. (a) The pentamer and hexamer states have the same energies, and differ only in their size. The resulting structures correspond to CK icosahedra. (b) Minimum energy structures for large energy difference between different capsomeres, i.e., only one size (type) of capsomere. The $N = 24$ and $N = 48$ now show octahedral symmetry, whereas $N = 32$ is icosahedral. Structure with $N = 72$ is highly degenerate, fluctuating over structures with different symmetry.*

potential. Within the class of viral capsids satisfying the CK principle an essential feature is the existence of two different types of capsomeres, pentamers and hexamers.

Zandi *et al.* carried out Monte Carlo simulations of the capsid assembly from these two types of capsomeres interacting via Lennard-Jones potential⁸ [33]. The only difference in the interaction potential of different capsomeres was that the equilibrium spacing (minimum of the potential) included the geometrical size difference between pentamers and hexamers of the same edge length. The energy difference between the pentamers and hexamers, reflecting differences between individual contact interactions and folding conformations of capsomere proteins, entered as a Boltzmann factor that provided the relative thermal probability for a non-interacting unit to be in a pentamer state. For a fixed number of capsomeres N the number of pentamers was thus permitted to vary and was not fixed to 12 as in the CK construction.

For the case where the energy difference between different capsomeres was zero and the only difference between them was their size, the pronounced energy minima were found for values of N where the capsid structures coincided with icosahedral CK struc-

⁸However, different potentials were tested, and the results were found to be robust.

tures (Fig. 11a). The appearance of icosahedral symmetry thus seems to be a direct consequence of the free energy minimisation of a generic interaction, containing attraction and excluded volume repulsion, and the existence of two different morphological units. However, when there was only one type of capsomere present (*i.e.* the energy difference between the two types was significantly increased), the observed minima differed greatly (Fig. 11b). Some minima now had octahedral symmetry, whereas others fluctuated over structures with different symmetries. It was also found that the presence of a small compression of a capsid systematically facilitated the appearance of icosahedral symmetry [33].

Altogether, the study concluded that the existence of two different types of morphological units is not absolutely necessary for obtaining capsids with icosahedral symmetry, even though their presence strongly favours it [33]. It would be thus interesting to compare low- \mathcal{T} structures obtained with the Landau crystallisation theory for the cases of icosahedral and octahedral symmetry. Given an appropriate description of interactions between positions of protein subunits, one would be able to see which symmetry yields an energetic minimum.

6 Conclusions

That icosahedral symmetry should feature so prominently in the capsids of viruses is not at all obvious, especially considering the many different types of viruses and their constituents. It is observed in various families of viruses, and is not dependent on protein-specific interactions but rooted in physical origins. These are not yet fully understood, and the first theory explaining the existence of icosahedral symmetry, the Caspar-Klug principle of quasi-equivalence, focused mainly on the possible structures composed of identical units, whatever the mechanism of their assembly.

Its shortcomings were recently addressed by two models, the viral tiling theory and the Landau crystallisation on a spherical surface. Both include new types of structures not included in the CK construction, yet differ between themselves. The VT theory is still based on geometrical principles, and consequently seems to possess limited prediction power. On the other hand, the approach of the Landau theory considers the capsid self-assembly as a crystallisation on a sphere, and from this yields icosahedral structures both within the CK principle and those outside it. None of the models tries to explain the origin of the symmetry, but the crystallisation approach might give some insight to results of various self-assembly models.

Understanding the origin of the icosahedral symmetry in the self-assembly of identical units would be of much value, and the two models represent a significant step in this direction. After all, the symmetry of the capsid has important implications for the symmetry of the entire virus, including the shape of the packaged genome, the stress distribution over the capsid, and more. But we have seen that the simplicity of the symmetry clearly does not imply a simple explanation, and much still remains to be done.

References

- [1] J. Carter and V. Saunders, *Virology: Principles and Applications*. Wiley, Chichester, 2007.
- [2] J. Lidmar, L. Mirny, and D. R. Nelson, “Virus shapes and buckling transitions in spherical shells,” *Phys. Rev.* **E 68** (2003) 051910.
- [3] A. Šiber, “Buckling transition in icosahedral shells subjected to volume conservation constraint and pressure: Relations to virus maturation,” *Phys. Rev.* **E 73** (2006) 061915.
- [4] J. P. Elliott and P. G. Dawber, *Symmetry in Physics: Principles and Simple Applications*, vol. 1. Macmillan, Houndmills, 1986.
- [5] J. P. Elliott and P. G. Dawber, *Symmetry in Physics: Further Applications*, vol. 2. Oxford University Press, New York, 1979.
- [6] L. S. Levitov and J. Rhyner, “Crystallography of quasicrystals; application to icosahedral symmetry,” *J. Phys. France* **49** (1988) 1835–1849.
- [7] W. Ludwig and C. Falter, *Symmetries in Physics*. Springer, Berlin, 1996.
- [8] S. K. Kim, *Group Theoretical Methods and Applications to Molecules and Crystals*. CUP, Cambridge, 1999.
- [9] V. L. Lorman and S. B. Rochal, “Landau theory of crystallization and the capsid structures of small icosahedral viruses,” *Phys. Rev.* **B 77** (2008) 224109.
- [10] G. J. Morgan, “Historical review: Viruses, crystals and geodesic domes,” *Trends Biochem. Sci.* **28** (2003) 86–90.
- [11] D. L. D. Caspar and A. Klug, “Physical principles in the construction of regular viruses,” *Cold Spring Harbor Symposia on Quantitative Biology* **27** (1962) 1.
- [12] A. Šiber, “Icosadeltahedral geometry of fullerenes, viruses and geodesic domes.” arXiv:0711.3527v1 [physics.pop-ph], 2007.
- [13] V. L. Lorman and S. B. Rochal, “Density-wave theory of the capsid structure of small icosahedral viruses,” *Phys. Rev. Lett.* **98** (2007) 185502.
- [14] S. B. Rochal and V. L. Lorman, “Theory of a reconstructive structural transformation in capsids of icosahedral viruses,” *Phys. Rev.* **E 80** (2009) 051905.
- [15] R. Twarock, “A tiling approach to virus capsid assembly explaining a structural puzzle in virology,” *J. Theor. Biol.* **226** (2004) 477–482.
- [16] R. Twarock, “Mathematical virology: a novel approach to the structure and assembly of viruses,” *Phil. Trans. R. Soc.* **A 364** (2006) 3357–3373.

References

- [17] T. Keef, R. Twarock, and K. M. Elsayy, “Blueprints for viral capsids in the family of Polyomaviridae,” *J. Theor. Biol.* **253** (2008) 808–816.
- [18] D. G. Angelescu and P. Linse, “Modelling of icosahedral viruses,” *Curr. Opin. Colloid Interface Sci.* **13** (2007) 389–394.
- [19] D. G. Angelescu and P. Linse, “Viruses as supramolecular self-assemblies: modelling of capsid formation and genome packaging,” *Soft Matter* **4** (2008) 1981–1990.
- [20] L. D. Landau and E. M. Lifšic, *Statistical Physics, Part 1*, vol. 5 of *Course of theoretical physics*. Butterworth-Heinemann, 3rd ed., 2006.
- [21] T. A. Springer, “Invariant theory,” in *Lecture Notes in Mathematics*, A. Dold and B. Eckmann, eds. Springer, Heidelberg, 1977.
- [22] J. Navaza, “On the three-dimensional reconstruction of icosahedral particles,” *J. Struct. Biol.* **144** (2003) 13–23.
- [23] P.-D. Fan, J.-Q. Chen, and J. P. Draayer, “Reduced projection operators and algebraic expressions for the symmetry-adapted functions of the icosahedral group,” *Acta Cryst. A* **55** (1999) 871–883.
- [24] M. E. Rose, *Elementary Theory of Angular Momentum*. Wiley, New York, 1957.
- [25] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF chimera – a visualization system for exploratory research and analysis,” *J. Comput. Chem* **25** (2004) 1605–1612.
- [26] L. Li, S.-M. Lok, I.-M. Yu, Y. Zhang, R. J. Kuhn, J. Chen, and M. G. Rossmann, “The flavivirus precursor membrane-envelope protein complex: structure and maturation,” *Science* **319** (2008) 1830–1834.
- [27] I.-M. Yu, W. Zhang, H. A. Holdaway, L. Li, V. A. Kostyuchenko, P. R. Chipman, R. J. Kuhn, M. G. Rossmann, and J. Chen, “Structure of the immature Dengue virus at low pH primes proteolytic maturation,” *Science* **319** (2008) 1834–1837.
- [28] T. Keef and R. Twarock, “A new series of polyhedra as blueprints for viral capsids in the family of Papovaviridae.” [arXiv:q-bio/0512047v1 \[q-bio.BM\]](https://arxiv.org/abs/q-bio/0512047v1), 2005.
- [29] M. Baake, D. Joseph, P. Kramer, and M. Schlottmann, “Root lattices and quasicrystals,” *J. Phys. A: Math. Gen.* **23** (1990) 1037–1041.
- [30] R. Twarock and R. Hendrix, “Crosslinking in viral capsids via tiling theory,” *J. Theor. Biol.* **240** (2006) 419–424.
- [31] M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks III, and V. S. Reddy, “VIPERdb2: an enhanced and web API enabled relational database for structural virology,” *Nucl. Acids Res.* **37** (2009) D436–D442. <http://vipperdb.scripps.edu/>.

References

- [32] H. Naitow, J. Tang, M. Canady, R. B. Wickner, and J. E. Johnson, “L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism,” *Nat. Struct. Biol.* **9** (2002) 725–728.
- [33] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, “Origin of icosahedral symmetry in viruses,” *Proc. Natl. Acad. Sci. USA* **101** (2004) 15556–15560.
- [34] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, “Viral self-assembly as a thermodynamic process,” *Phys. Rev. Lett.* **90** (2003) 248101.
- [35] S. Čopar, “Thomsonov problem.” Faculty of Mathematics and Physics, Seminar II, 2008.
- [36] H. D. Nguyen and C. L. Brooks III, “Generalized structural polymorphism in self-assembled viral particles,” *Nano Lett.* **8** (2008) 4574–4581.