Taylor & Francis
Taylor & Francis Group

Exploring the interplay between virology and molecular crystallography

Aloysio Janner

Theoretical Physics, Faculty of Science, Radboud University, Toernooiveld, NL-6525 ED Nijmegen, The Netherlands

Polyhedra with icosahedral symmetry and vertices labelled by rational indices of points of a six-dimensional lattice left invariant by the icosahedral group allow a morphological characterization of icosahedral viruses which includes the Caspar–Klug classification as a special case. Scaling transformations relating the indexed polyhedron enclosing the surface with the one delimiting the central cavity lead to models of viral capsids observed in nature. Similar scaling relations can be obtained from projected images in three dimensions of higher-dimensional crystallographic point groups having the icosahedral group as a subgroup. This crystallographic approach can be extended to axial-symmetric clusters of coat proteins around icosahedral axes of the capsid. One then gets enclosing forms with vertices at points of lattices left invariant by the corresponding point group and having additional crystallographic properties also observed in natural crystals, but not explained by the known crystallographic laws.

1. Introduction

Molecular crystallography is a concept based on empirical morphological properties of biomacromolecules, which appears at first as a self-contradiction. Indeed, crystallography is based on lattice periodicity, a property not occurring in molecules, which are finite even in the ideal case. The first puzzling property consists of polyhedral forms enclosing the molecule and having vertices at points of a lattice as observed in icosahedral viruses and in a number of proteins and nucleic acids with axial symmetry [3–6,8].

Lattices with axial point group symmetry have lattice parameters a and c with axial ratio $\gamma = c/a$. The parameters a and c correspond to the length of the basis vectors in the plane of rotation and the length of the one along the axis, respectively. This implies that the value of the axial ratio γ does not depend on the point group symmetry. The second puzzling property is that for lattices enclosing molecular forms, the square of this ratio is a rational number with small integers p and q. Triggered by these observations, distributions of lattices of crystals belonging to a same crystallographic system have been plotted as a function of the axial ratio. One then finds sharp peaks at rational values for γ^2. In particular, the value $\gamma = 1$ dominates, both in the molecular and in the crystal cases [1,2,7], a property which is not explained by the known crystallographic laws. Molecules and crystals seem, therefore, to have at a deeper level similar properties which can be used for the geometrical characterization of viruses as well.

The viral morphology is considered first, followed by that of clusters of coat proteins with a given axial symmetry. Finally, an attempt is mentioned which could possibly solve the present puzzles.

---

*Email: a.janner@science.ru.nl.

2. Morphology of icosahedral viruses

Since the pioneering work of Caspar and Klug in 1962, the icosahedral viruses are classified according to a triangulation number $T = h^2 + hk + k^2$, where h, k are the integral coordinates of a point $P(h, k)$ in the hexagonal lattice of a planar honeycomb net. The folding of this net into the triangular faces of an icosahedron is obtained by cutting and sticking the planar net and replacing with pentagons the honeycomb hexagons centred at the vertices of a triangular icosahedral face. This procedure is repeated for the other faces as well.

There are, however, icosahedral viruses, like the Polyoma virus and the Simian Virus 40 which do not obey the Caspar–Klug rules underlying their model. The puzzle has been solved by Twarock [12], who replaced the triangular tessellation of Caspar and Klug with a Penrose-like tiling with kites and darts. The approach has been generalized by Keef and Twarock [11], making use of the mathematical structure of affine extensions of Lie algebras.

The alternative viral classification scheme adopted here is based on the property that the icosahedral point group 235 leaves a six-dimensional lattice invariant, and in three dimensions the projected image of this lattice generated by all integral linear combinations of the six vectors $\{a_1, a_2, \ldots, a_6\}$ pointing from the centre to the non-aligned vertices of an icosahedron. These vectors form a basis $\{a\}$ of the projected lattice which, mathematically speaking, is a $\mathbb{Z}$-module M_{ico} of dimension 3 and range 6. Accordingly, a point P of the projected lattice M_{ico} can be labelled (like the icosahedral lattice points) by a set of six integer-valued indices denoted as $[n_1 n_2, \ldots, n_6]$, which are the components of P with respect to the basis $\{a\}$.

The icosahedral group applied to one indexed point generates a polyhedron with icosahedral symmetry and vertices with integral indices. In particular, the icosahedron is generated by 235 from [100000]. The Euler symbol (*VEF*) of the icosahedron is (12 30 20), where *V*, *E* and *F* indicate the number of vertices, of edges and of faces, respectively, satisfying the relation $V + F = E + 2$. The dodecahedron (20 30 12) is generated in a similar way from [111000]. Additional icosahedral forms are obtained starting from one, two or more points of M_{ico}. For example, the triacontahedron, with 32 vertices and 30 rhombic faces discovered by Kepler in 1611, which is the projection in space of the six-dimensional hypercube, has icosahedral vertices generated from [200000] and dodecahedral ones generated from [111$\bar{1}$1$\bar{1}$].

Indexed polyhedra enclosing the capsid can be used for a classification of viruses. For example, the Satellite Panicum Virus has an icosahedron as enclosing form, whereas the Canine Parvovirus is enclosed by a dodecahedron, both being $T = 1$ viruses according to Caspar–Klug [9]. The $T = 3$ Bacteriophage MS2, together with the various serotypes of the Human Rhinovirus, are enclosed by triacontahedra [8,9].

In the same way, one can classify the surface of the central hole where the viral genome is located. The relevance of the crystallographic approach considered is supported by the property, observed in the few cases investigated so far, that the internal polyhedron delimiting the central hole and the external polyhedron enclosing the capsid have vertices belonging to the same $\mathbb{Z}$-module M_{ico}. Moreover, the two polyhedra are related by a scaling, which transforms rational indices into rational indices. Scaling transformations with this property are denoted as crystallographic. For example, the capsid of the Rhinovirus is delimited by two triacontahedra radially scaled by a factor $\tau = (1 + \sqrt{5})/2$, the golden mean [8]. Expressed in the icosahedral basis $\{a\}$ indicated above, this scaling is represented by an invertible six-dimensional matrix with half-integer entries and is, therefore, crystallographic. There are also additional crystallographic planar and linear scalings which allow the decoration of faces, of edges and even of vertices of the polyhedral forms, in a similar way as triangular facets are obtained from the Caspar–Klug construction when h and k are not

relatively prime [9]. The crystallographic approach is very rich and includes the tessellations in Caspar–Klug Theory as special cases. This does not mean that it covers adequately the morphology of all possible viruses.

The application of these ideas to viruses was induced by properties observed in molecular systems (proteins and nucleic acids) with axial symmetry. For being in a similar situation, clusters of protein subunits obtained from axial subgroups of 235 are considered in the next section.

3. Axial-symmetric clusters of coat proteins

As already mentioned, the investigation of axial-symmetric proteins has revealed two generic crystallographic properties of their enclosing forms:

1. Vertices at points of a given lattice, the *form lattice*, with same or larger point group symmetry than the molecular system.
2. The form lattice is integral.

The concept of *molecular crystallography* has been introduced because of these phenomenological properties, and of similar additional ones not discussed here. A theoretical foundation is still missing.

Generic means that these properties are observed in a number of proteins with different rotational symmetry independently from their biological activities. A lattice is said to be integral if the matrix of the scalar products of the lattice basis vectors (the Gram matrix) has integral entries. This may require the choice of an appropriate unit of length. In the axial case, lattices whose axial ratio squared is a rational number are integral, as shown further on for the hexagonal case. Typical situations where one finds the first property is that of multimers with a central hole. In practically all axial-symmetric cases the external and the internal enclosing forms are related by a scaling transformation leaving the form lattice invariant [3–5].

An icosahedral viral capsid is not uniaxial, but one can consider clusters of coat proteins having twofold, threefold or fivefold rotational symmetry. The question then arises whether these clusters share similar crystallographic properties as the proteins already investigated. In the case of the Rhinovirus of various serotype (HRV1A, HRV2, HRV3, HRV14 and HRV16) and for each of the coat proteins involved (VP1, VP2, VP3 and VP4) the answer is yes. It is convenient to consider tetramers (instead of dimers), trimers together with hexamers and pentamers together with decamers. The corresponding form lattices are orthorhombic (instead of monoclinic), hexagonal and decagonal, respectively. In the last case, the lattice is five-dimensional giving rise by projection to a decagonal $\mathbb{Z}$-module M_{10} of dimension 3 and range 5. Each of the coat proteins occurs in the capsid as one of the 60 copies obtained by the icosahedral group from a given 'starting' monomer. These identical copies appear differently when viewed along one of the icosahedral axes. It is, therefore, appropriate to label the monomers by the corresponding element of the icosahedral group ordered in succession $R_0, R_1, \ldots, R_{59}$, with R_0 the identity element. For example, if one has $2_x = R_{37}$ and A_0 is the chain whose coordinates are given in the Brookhaven Data Bank for the VP1 coat protein, then the monomer $2_x A_0$ is noted as $VP1_{37}$ and [0,32,37,42] indicates the tetrameric cluster obtained by applying the subgroup 222 of 235 to A_0.

For each of the four coat proteins, one gets 15 different 222-clusters and in all cases the corresponding enclosing forms appear to be independent of the serotype and to have vertices at an orthorhombic lattice with lattice parameters $a = a_c/z_1, b = a_c/z_2, c = a_c/z_3$ with z_1, z_2, z_3 integers and a_c the half edge of the cube enclosing the capsid. All these lattices are integral. In the example of the tetrameric cluster [10,15,29,43] of the coat protein VP2 of the Human Rhinovirus

HRV3 shown in Figure 1a, one has $z_1 = 7, z_2 = 9, z_3 = 11$. In a similar way, one gets trimers from a subgroup 3 and hexamers from a subgroup 32 of 235. Again in all the possible cases (20 trimeric and 10 hexameric ones) the enclosing forms appear to have vertices at points of a hexagonal lattice with lattice parameters $a = r_h/z$ and $c = \gamma a$, where r_h is the hexagonal radius of the enclosing triacontahedron projected along a threefold axis and $\gamma^2 = p/q$ is rational. Accordingly, the Gram matrix for the corresponding hexagonal lattice with basis vectors chosen in the standard way is integral:

$$g_h(z, \gamma) = \frac{r_h^2}{2qz^2} \begin{pmatrix} 2q & -q & 0 \\ -q & 2q & 0 \\ 0 & 0 & 2p \end{pmatrix}. \tag{1}$$
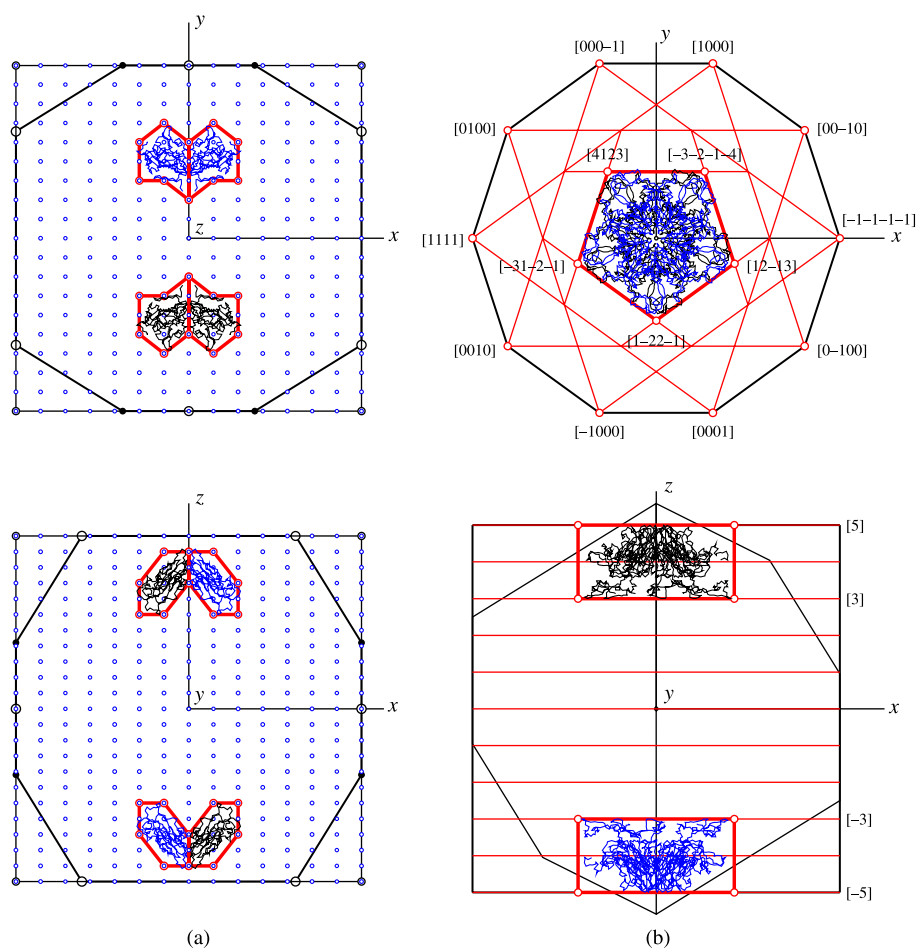

Figure 1.   (a) The tetramer [10,15,29,43] with 222 symmetry of the VP2 coat protein of HRV3 has an orthorhombic form lattice satisfying the relation $7a = 9b = 11c$. (b) The decamer [0,1,2,3,4;30,31, 32,33,34] of VP1 has enclosing forms with vertices indexed by a decagonal lattice with axial ratio $c/a = 1/5$, rationally equivalent to the isometric case $c = a$. The projected boundary of the capsid is also indicated. The results are independent of the serotype of the Rhinovirus. Available in colour online.

Figure 2a shows the HRV3 hexameric cluster [11,27,46;16,37,41] of VP3, where $z = 6$, $\gamma = 1/2$ and in Figure 2b one finds the enclosing forms of the VP4 cluster [11,27,46;16,37,41;12,26,47; 17,38,41] with $z = 5$ and $\gamma = 1/3$. As in both cases, γ already is rational (and not only γ^2), these form lattices are equivalent to the isometric hexagonal lattice, where $c = a$ [8]. Finally, there are six decameric clusters with symmetry 52 and 10 pentamers with symmetry five. The dyadically related pentamers of a given decamer are indicated by two sets of five monomers, like in [5,17,22,41,49;12,27,35,51,59]. In a similar way as in the previous clusters the enclosing forms of these pentamers and decamers are at points of a decagonal $\mathbb{Z}$-module M_{10}, obtained as a projection in space of a five-dimensional decagonal lattice. Four of the five basis vectors of M_{10} can be chosen pointing from the centre to four vertices of a regular pentagon with radius a_p and the fifth one perpendicular to the pentagonal vectors and of length $c_p = \gamma a_p$. One finds again for the enclosing polyhedra $\gamma^2 = p/q$ rational, implying that the corresponding decagonal lattice is integral. In the particular case of the HRV3 decameric cluster [0,1,2,3,4;30,31,32,33,34] of VP1, shown in Figure 1b, γ is 1/5 leading again to the isometric $\gamma = 1$ case, as observed in most of the pentagonal or decagonal clusters of the Rhinovirus [8].
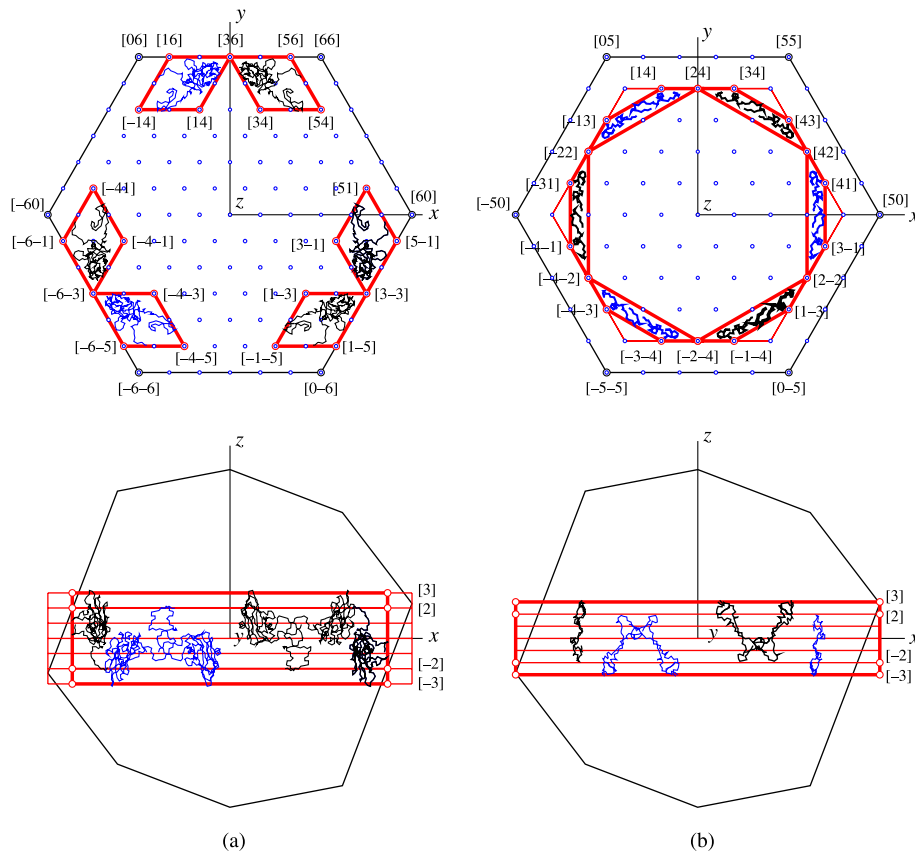

Figure 2. (a) The hexamer [11,27,46;16,37,41] of the coat protein VP3 of the same Rhinovirus as in Figure 1, has an hexagonal form lattice with $a = r_h/6$ and $c = a/2$, where r_h is the hexagonal radius of the projected capsid, whereas one has $a = r_h/5$ and $c = a/3$ for the dodecameric cluster [11,27,46;16,37,41;12,26,47;17,38,42] of VP4 shown in (b). Both form lattices are equivalent to the isomeric hexagonal lattice. Available in colour online.

4. Remarks and perspectives

There is one aspect which deserves special attention. It is the role played in a finite molecular system like a virus by transformations of infinite order as the crystallographic scalings mentioned above, or the translations generated by reflections in the Keef–Twarock approach [11]. In both cases, one derives non-trivial structural informations, for example on the nested shells observed in a number of icosahedral viruses, by restricting to a finite number the infinite-order transformations considered. From this point of view, even the triacontahedron appears to be composed by an icosahedron and a dodecahedron mutually related by a crystallographic scaling. Doing so is meaningful and relevant, but these infinite order transformations cannot be included in the symmetry group of the system. The question remains whether the same results can be obtained by a finite group of transformations leaving the finite set of points invariant and containing the known Euclidean symmetry point group as a subgroup. The larger point group would then in fact represent a hidden symmetry group of the molecular system.

This is possible in principle. Consider as an example the group 2354 of order 240, semidirect product of the icosahedral group 235 by a fourfold rotation, which leaves the icosahedral lattice invariant. When applied to the point $[11111\bar{1}]$ it generates nested icosahedra scaled by τ^2. When applied to [111000], $[11111\bar{1}]$ and [110000] of the module M_{ico}, one gets pairs of dodecahedra scaled by τ^3 and by τ and icosidodecahedra scaled by τ, respectively [10]. Both the icosahedral and dodecahedral vertices of the triacontahedron can be obtained from [111111] by the six-dimensional hypercubic group. In this case, however, one also gets nested an additional icosahedron and a dodecahedron, so that the hypercubic group cannot be considered to be the symmetry group of the triacontahedron. In a similar way, from projection of a higher-dimensional lattice (like the hypercubic lattice) one can get three-dimensional integral lattices. Possibly, therefore, the puzzles indicated in the Introduction can be solved by considering three-dimensional images of a higher-dimensional crystallography.

Thinking at the evolutionary time required by nature to realize the present viral structures, it is not surprising that much more work is needed for an adequate understanding of their complex order. The perspectives opened by the approaches mentioned justify further exploration.

References

[1] R. de Gelder and A. Janner, *Remarkable features in lattice-parameter ratios of crystals. I. Orthorhombic, tetragonal and hexagonal crystals*, Acta Cryst. B61 (2005), pp. 287–295.

[2] ———, *Remarkable features in lattice-parameter ratios of crystals. II. Monoclinic and triclinic crystals*, Acta Cryst. B61 (2005), pp. 296–303.

[3] A. Janner, *Strongly correlated structure of axial-symmetric proteins. I. Orthorhombic, tetragonal, trigonal and hexagonal symmetries*, Acta Cryst. D61 (2005), pp. 247–255.

[4] ———, *Strongly correlated structure of axial-symmetric proteins. II. Pentagonal, heptagonal, octagonal, nonagonal and ondecagonal symmetries*, Acta Cryst. D61 (2005), pp. 256–268.

[5] ———, *Strongly correlated structure of axial-symmetric proteins. III. Complexes with DNA/RNA*, Acta Cryst. D61 (2005), pp. 269–277.

[6] ———, *Polygrammal symmetries in biomacromolecules: Heptagonal poly d(As^4T).poly d(As^4T) and heptameric α-hemolysin*, Struct. Chem. 13 (2002), pp. 277–287.

[7] ———, *Integral lattices*, Acta Cryst. A60 (2004), pp. 198–200.

[8] ———, *Crystallographic structural organization of human rhinovirus serotype 16, 14, 3, 2 and 1A*, Acta Cryst. A62 (2006), pp. 270–286.

[9] ———, *Towards a classification of icosahedral viruses in terms of indexed polyhedra*, Acta Cryst. A62 (2006), pp. 319–330.

[10] ———, *Higher-dimensional point groups in superspace crystallography*, Acta Cryst. A64 (2008), pp. 280–283.

[11] T. Keef and R. Twarock, *A new series of polyhedra as blueprints for viral capsids in the family of Papovaviridae*, arXiv: q-bio/0512047v1.

[12] R. Twarock, *A tiling approach to virus capsid assembly explaining a structural puzzle in virology*, J. Theor. Biol. 226 (2004), pp. 477–482.