Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 31 May 2012
Accepted 16 November 2012

From an affine extended icosahedral group towards a toolkit for viral architecture

A. Janner

Theory of Condensed Matter, IMM, Radboud University, Heyendaalseweg 135, NL-6525 AJ Nijmegen, The Netherlands. Correspondence e-mail: a.janner@science.ru.nl

The affine extensions (there are 55 different ones) of the icosahedral group developed by T. Keef and R. Twarock of the York Centre for Complex Systems Analysis of the University of York [see in particular Keef *et al.* (2013). *Acta Cryst.* A**69**, 140–150], and applied to the investigation of the architecture of a number of icosahedral viruses, are here considered in the framework of molecular crystallography. The basic ideas of such molecular description involve positions with rational indices which approximate backbone positions in viral polypeptide and RNA chains. The test case of the Pariacoto virus suggests that the best-fit algorithm used in the York group's approach should be adapted to a more specific toolkit suited for the investigation of the architecture of icosahedral viruses. Typical problems which could be solved by means of such a toolkit are exemplified and put in the perspective of viral properties.

1. Introduction

An affine extension of the icosahedral group has been formulated by T. Keef and R. Twarock in terms of point arrays with icosahedral symmetry and six-dimensional rational indices (Keef & Twarock, 2008, 2009*a*). These discrete finite sets, suitably fitted to the backbone structure of icosahedral viruses, have been shown to approximate in a meaningful way the architecture of both the viral capsid and the genome enclosed in it (Keef & Twarock, 2009*b*; Keef *et al.*, 2013). Combined with the theory of structural phase transitions in crystals, the Keef–Twarock approach allowed analysis of viral maturation (Indelicato *et al.*, 2012).

The assignment of a set of n rational indices to a position in three dimensions leads to a corresponding nD description (where the indices are the coordinates of lattice points in n dimensions) and, therefore, to a crystallographic approach. Such an approach, which is the natural one for incommensurate crystals and for quasicrystals, has been shown to lead to non-trivial structural relations in axial-symmetric biomacromolecules, like proteins and nucleic acids (see *e.g.* Janner, 2008*a*, and references therein), and is denoted as *molecular crystallography*.

The application of molecular crystallography to icosahedral viruses (they are biomacromolecules with axial symmetry) allows a natural classification going beyond the standard one of Caspar and Klug (Janner, 2006; Caspar & Klug, 1962). Moreover, the geometrical results obtained for the architecture of the capsid could be extended to the genome as well (Janner, 2011*b*; Keef *et al.*, 2013) in a way consistent with the phenomenological results obtained for proteins and nucleic acids. Typical for such an approach are enclosing forms with vertices having rational indices and crystallographic scalings relating these forms.

Molecular crystallography deals with the morphological properties of single molecules. By considering lattice-periodic packing of these molecules in crystals, one obtains relations going beyond both the classical crystallography [involving *e.g.* integral lattices (Janner, 2004)] and the molecular one [*e.g.* through encoding of the space-group symmetry in a viral capsid (Janner, 2010, 2011*a*)].

The relation of these various approaches to the alternative one based on an affine extension of the icosahedral group is far from trivial and requires further work. The present paper, where the Keef–Twarock approach is analysed in the spirit of the molecular crystallography mentioned above, represents a first step in this direction.

Considered as an example are icosahedral viruses already analysed according to these alternative approaches and whose structural data are reported in the Protein Data Bank (PDB): the Pariacoto virus (PaV) (PDB key 1f8v; Tang *et al.*, 2001), the cowpea chlorotic mottle virus (CCMV) (PDB key 1cwp; Speir *et al.*, 1995), the satellite tobacco mosaic virus (STMV) (PDB key 1a34; Larson *et al.*, 1998), the RNA MS2 bacteriophage (MS2) (PDB key 1zdh; Valegård *et al.*, 1997), the tomato bushy stunt virus (TBSV) (PDB key 2tbv; Hopper *et al.*, 1984), the cucumber mosaic virus (CMV) (PDB key 1f15; Smith *et al.*, 2000), the Seneca Valley virus (SVV-001) (PDB key 3cji; Venkataraman *et al.*, 2008), the simian virus 40 (SV40) (PDB key 1sva; Stehle *et al.*, 1996) and the murine polyoma virus (murine py) (PDB key 1sid; Stehle & Harrison, 1996).

Discussed are: the Pariacoto virus (Keef & Twarock, 2008, 2009*a*,*b*; Keef *et al.*, 2013; Janner, 2011*b*,*c*); the cowpea

chlorotic mottle virus (Keef & Twarock, 2009*b*; Indelicato *et al.*, 2012; Janner, 2011*b*,*c*); the satellite tobacco mosaic virus (Janner, 2011*b*,*c*); the MS2 bacteriophage (Keef *et al.*, 2013; Janner, 2011*b*,*c*); the tomato bushy stunt virus, the Seneca Valley virus and the cucumber mosaic virus (Keef & Twarock, 2009*b*); the simian virus 40 (Keef *et al.*, 2013); and the polyoma virus (Keef *et al.*, 2008).

2. A crystallographic approach for icosahedral viruses

A crystallographic characterization of icosahedral viruses is here outlined, because it is shared by all the various approaches mentioned in §1.

The icosahedral group 235 = $\{R_5, R_3 | R_5^5 = R_3^3 = (R_5R_3)^2 = 1\}$ leaves invariant a lattice in six dimensions, but not a three-dimensional one. Indeed the lowest integral faithful representation of 235 is six dimensional.

The action of 235 mentioned is on the six non-aligned vectors pointing to vertices of an icosahedron that are the projections of the six-dimensional simple-cubic lattice basis into one of the three-dimensional subspaces invariant under the icosahedral group. This action defines a $\mathbb{Z}$-module $\Lambda_{\rm ico}$ of rank 6 and dimension 3. A possible basis for $\Lambda_{\rm ico}$, with components with respect to the orthonormal basis $e = \{e_1, e_2, e_3\}$, is given by

$$\begin{array}{lll} a_1 = a_0(e_1 + \tau e_3), & a_2 = a_0(\tau e_1 + e_2), & a_3 = a_0(\tau e_2 + e_3), \\ a_4 = a_0(-e_1 + \tau e_3), & a_5 = a_0(-\tau e_2 + e_3), & a_6 = a_0(\tau e_1 - e_2), \end{array} \quad (1)$$

where a_0 is the icosahedral lattice parameter and $\tau = (1 + 5^{1/2})/2$, so that

$$\Lambda_{\rm ico} = \left\{ \sum_{i=1}^{i=6} n_i a_i \, | n_i \in \mathbb{Z} \right\}. \quad (2)$$

As the six vectors a_i are linearly independent on the rational numbers $\mathbb{Q}$, a six-dimensional indexed position I, with components (x, y, z) with respect to the orthonormal basis $e = \{e_1, e_2, e_3\}$, is uniquely defined by the *rational indices* $[q_1, q_2, \ldots, q_6]$ where the $q_i \in \mathbb{Q}$ are the components of I with respect to the $\mathbb{Z}$-module basis $a = \{a_1, \ldots, a_6\}$.

The basic idea of a molecular crystallographic approach to a viral structure is to approximate backbone positions, or vertices of an enclosing form, by corresponding positions with rational indices.

Applying to an initial indexed position I_0 the icosahedral group 235, one gets the orbit $O(I_0)$ of 235-equivalent positions, which all have rational indices, and a polyhedron with icosahedral symmetry generated from I_0 with vertices at those positions.

In the general case, the order of an orbit is 60, which is the order of the group 235. Disregarding the trivial case of an orbit of order 1, obtained from the position [000000] at the origin, the other possible orbits have order 12, 20 and 30, and correspond to an icosahedron, a dodecahedron and an icosidodecahedron, generated (for example) from the indexed positions [100000], [111000] and [110000], respectively.

Table 1
Standard polyhedra with icosahedral symmetry and $a_0 = 1$.

Standard polyhedron	(V E F)	Generator	Sym.	Radius
Icosahedron (ICO)	(12 30 20)	[010000] = $(\tau, 1, 0)$	5	$r_{\rm ICO} = (2 + \tau)^{1/2}$
Dodecahedron (DOD)	(20 30 12)	$\frac{1}{2}[11\bar{1}1\bar{1}1] = (1, 1, 1)$	3	$r_{\rm DOD} = 3^{1/2}$
Icosidodecahedron (IDD)	(30 60 30)	$\frac{1}{2}[110000] = \frac{1}{2}(\tau, \tau^2, 1)$	2	$r_{\rm IDD} = 2$

It is convenient to define the following standard icosahedral polyhedra, according to the choice one finds in Indelicato *et al.* (2012), with characteristics indicated in Table 1, where (V E F) denotes the number of vertices (V), edges (E) and faces (F) of the polyhedron, and Sym. is the site symmetry of the vertices. Here and further on, the indices -1 are noted as $\bar{1}$.

3. Affine extended pentagonal group

Before considering the icosahedral group, it is convenient to illustrate the concepts and procedure of the affine extension of a symmetry group by the two-dimensional pentagonal case.

One starts from the pentagonal $\mathbb{Z}$-module $\Lambda_{\rm penta}$ of dimension 2 and rank 4, with as basis the vectors pointing from the centre to four vertices of a regular pentagon. One can choose:

$$b_k = (\cos k\varphi, \sin k\varphi), \quad k = 1, 2, 3, 4, \quad \varphi = 2\pi/5, \quad (3)$$

so that

$$\Lambda_{\rm penta} = \left\{ \sum_{i=1}^{4} n_i b_i | n_i \in \mathbb{Z} \right\}. \quad (4)$$

As reported in Keef & Twarock (2008, 2009*a*), Keef *et al.* (2013) and illustrated here in Fig. 1, an affine extension involves the following steps:

(i) One considers a *starting pentagon* with vertices at

$$[1000], \quad [0100], \quad [0010], \quad [0001], \quad [\overline{1111}]$$

(Fig. 1*a*).

(ii) These positions are translated by the *pentagonal vector* $T_{\rm penta} = -(b_1 + b_2 + b_3 + b_4)$ yielding the five additional indexed points:

$$[0\overline{111}], \quad [\bar{1}0\overline{11}], \quad [\overline{11}0\bar{1}], \quad [\overline{111}0], \quad [\overline{2222}]$$

(Fig. 1*b*). Note that this pentagonal vector is only one of a finite number of possible translations in this classificato.

(iii) One applies to all these ten points the fivefold rotation R_5,

$$R_5 = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \bar{1} \\ 1 & 0 & 0 & \bar{1} \\ 0 & 1 & 0 & \bar{1} \\ 0 & 0 & 1 & \bar{1} \end{pmatrix}, \quad \varphi = 2\pi/5, \quad (5)$$

and one gets a first-order *affine system* $P1$ of cardinality 20, decomposed into four orbits A, B, C, D of order 5, as indicated in Table 2 and Fig. 1(*c*).

(iv) The higher-order affine systems (not considered in this article) are obtained by repeating steps (ii) and (iii).

Note that the orbits A, D and C are related by the crystallographic scaling S_τ with scaling factor τ, which is invertible and integral, like its inverse $S_{1/\tau}$, possibly combined with the two-dimensional total inversion.

$$S_\tau = \begin{pmatrix} 0 & 1 & 0 & \bar{1} \\ 0 & 1 & 1 & \bar{1} \\ \bar{1} & 1 & 1 & 0 \\ \bar{1} & 0 & 1 & 0 \end{pmatrix}, \quad S_{1/\tau} = \begin{pmatrix} \bar{1} & 1 & 0 & \bar{1} \\ 0 & 0 & 1 & \bar{1} \\ \bar{1} & 1 & 0 & 0 \\ \bar{1} & 0 & 1 & \bar{1} \end{pmatrix}. \quad (6)$$

4. Affine extended icosahedral group

In a similar way as in the pentagonal case, one derives the systems of the affine extension of the icosahedral group and the decomposition of these affine systems into icosahedral affine orbits. The starting polyhedra are the standard ones indicated in Table 1. The admitted translations are discussed in detail in Keef & Twarock (2008, 2009*a*) and in Indelicato *et al.* (2012), so repetition is avoided here.

The 55 affine systems (up to first order) An, $n = 1, \ldots, 55$, derived by these authors have been numbered in the same successive order as in their publications. In particular, $A1$ to $A41$ appear in Table 5 of Keef & Twarock (2009*a*), the additional 13 ones $A42$ to $A54$ in Table 1 of Keef & Twarock (2008) and the last one is indicated as shell26 in Table 2 of Indelicato *et al.* (2012).

These systems have been computed again. The only deviation found is in the cardinality of $A42$, which is 420 and not 360, as indicated in Table 1 of Keef & Twarock (2008). These orbits have been numbered for each affine system in decreasing order of the radius of the corresponding standard polyhedron.

Denoting by An_m the mth orbit, the one in the affine system An with maximal radius is indicated as An_1.

Table 2
Orbits of the pentagonal affine system $P1$ (ordered according to decreasing radius).

Orbit	Name	Generator	Order	Radius
B	$P1_1$	[2000]	5	2
D	$P1_2$	[1100]	5	τ
A	$P1_3$	[1000]	5	1
C	$P1_4$	[$\bar{1}$011]	5	$1/\tau$

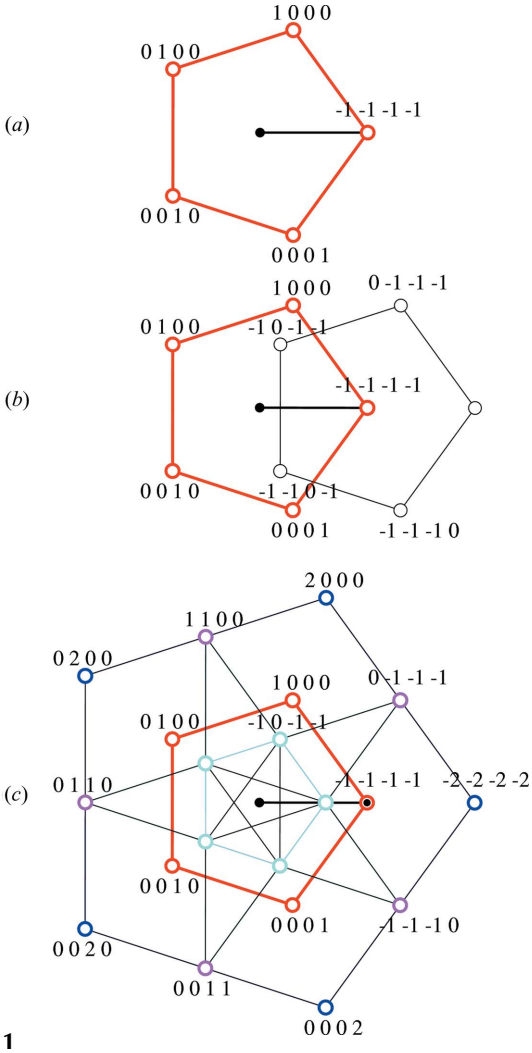

Figure 1
Affine extension of the pentagonal group (up to first order). (*a*) Start configuration. (*b*) Pentagonal translation. (*c*) Fivefold rotation.

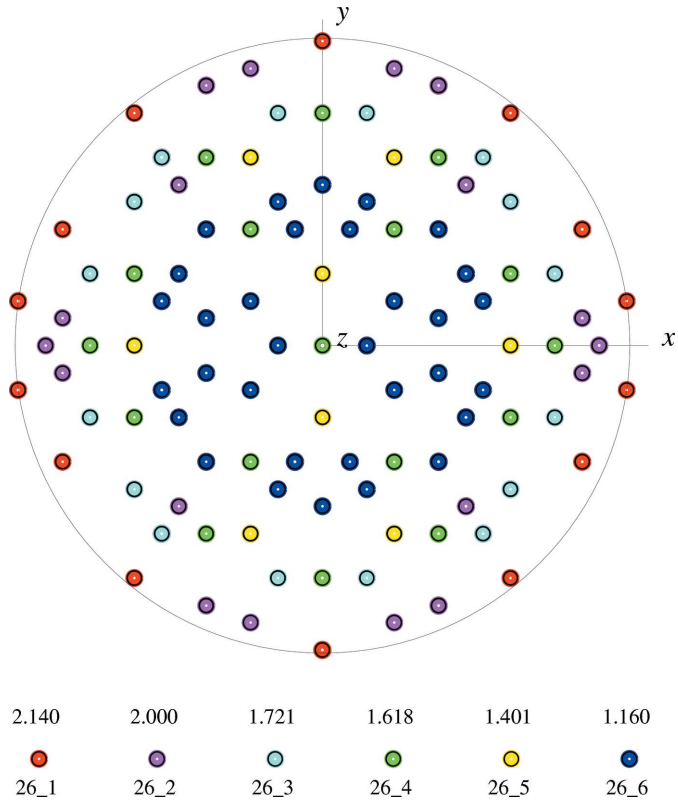

Figure 2
Points of the six orbits of the icosahedral affine system $A26$ and corresponding standard radii.

Table 3
Orbits of the affine systems $A1$, $A26$ and $A55$.

System	Start polyhedron	Translation	Cardinality	Orbit	Order	Radius (standard, as in Table 1)	Generator
$A1$	ICO	$-(1/\tau)T_5$	116	$A1_1$	12	3.0776	$\frac{1}{2}[\bar{1}11111]$
		$T_5 = (\tau, 1, 0)$		$A1_2$	60	2.6458	$\frac{1}{2}[1\bar{3}1111]$
		(ICO, $-1/\tau$)		$A1_3$	12	1.9021	[100000]
				$A1_4$	20	1.7321	$\frac{1}{2}[11\bar{1}111]$
				$A1_5$	12	0.7265	$\frac{1}{2}[31\bar{1}111]$
...	...	...	...	...	...	...	...
$A26$	IDD	$-1/(2\tau)T_5$	290	$A26_1$	60	2.1404	$\frac{1}{4}[\bar{1}3\bar{1}111]$
		[ICO, $-1/(2\tau)$]		$A26_2$	60	2.0000	$\frac{1}{4}[11\bar{1}133]$
				$A26_3$	60	1.7214	$\frac{1}{4}[11\bar{1}3\bar{1}1]$
				$A26_4$	30	1.6180	$\frac{1}{2}[101000]$
				$A26_5$	20	1.4013	$\frac{1}{4}[1\bar{1}11\bar{1}1]$
				$A26_6$	60	1.1600	$\frac{1}{4}[3\bar{1}111]$
...	...	...	...	...	...	...	...
$A55$	DOD	$\tau^2 T_3$	360	$A55_1$	20	6.2665	$\frac{1}{2}[3\bar{1}33\bar{1}1]$
		$T_3 = (0, 1/\tau, \tau)$		$A55_2$	60	5.9388	$\frac{1}{2}[3113\bar{1}1]$
		(DOD, τ^2)		$A55_3$	60	5.3662	$\frac{1}{2}[113\bar{3}\bar{1}1]$
				$A55_4$	60	5.3662	$\frac{1}{2}[3\bar{1}31\bar{1}\bar{1}]$
				$A55_5$	60	4.2808	$\frac{1}{2}[113111]$
				$A55_6$	60	3.4430	$\frac{1}{2}[3\bar{1}11\bar{1}1]$
				$A55_7$	20	2.8024	$\frac{1}{2}[11111\bar{1}]$
				$A55_8$	20	1.7321	$\frac{1}{2}[11\bar{1}1\bar{1}1]$

As illustration the orbits of the affine systems $A1$, $A26$ (which plays a role in the structure of the Pariacoto virus, as discussed further on) and $A55$ are listed in Table 3. In particular, the orbit points of $A26$ are plotted in Fig. 2 in a view along the icosahedral twofold axis.

In the 55 affine systems one finds 470 orbits of the point group 235: 36 are of order 12 (icosahedral orbits), 45 of order 20 (dodecahedral orbits), 70 of order 30 (icosidodecahedral orbits), 316 of order 60 (general orbits) and 3 of order 1 (trivial orbits).

All icosahedral orbits are mutually related by crystallographic scaling transformations (invertible, with rational entries), and this is also the case for the dodecahedral and the icosidodecahedral orbits. In Table 4 the orbits of order 12 are indicated together with their standard radius, scaling factor and set of the rational indices of a chosen generator. In a similar way, some few illustrative examples are reported for the dodecahedral, the icosidodecahedral and the general case. Scaling transformations mutually relate only some of the 60-order orbits; therefore the corresponding scaling factors have been omitted.

5. Fitting an affine system to the viral capsid

The central idea of the structural relevance of the affine extended symmetry group is that the fitting of one, or more, affine systems to a given virus yields a one-parameter characterization of its architecture.

To begin with, one affine orbit is identified as an *outer orbit*, among those having the largest standard radius in each affine system (conventionally indicated by An_1), which optimizes the fitting of the residue of a given chain at the greatest radial distance from the centre (the *outer residue*). Such a residue $C_\alpha^{\rm out}(j)$ is the jth one of what can be denoted as the kth outer chain $C^{\rm out}[k]$, as representative of the icosahedral equivalent ones:

$$C_\alpha^{\rm out}(j) \in C^{\rm out}[k], \quad r[C_\alpha^{\rm out}(j)] = r_{\rm max}. \tag{7}$$

The orbits An_1 of all possible affine systems An are then rescaled with respect to $C_\alpha^{\rm out}(j)$ in such a way that their radius is equal to $r_{\rm max}$. Accordingly, the C_α-rescaled generator I_{C_α} is related to the the standard one I of An_1, with radius $r_{\rm st}$, by

$$I_{C_\alpha} = \frac{r_{\rm max}}{r_{\rm st}} I = kI, \tag{8}$$

with k the rescaling factor.

Among all outer affine orbits, the fitted An_1 is the one generated by the $I_{C_\alpha}^{\rm out}(n)$ at the minimal distance $D_{\rm min}$ from the outer residue $C_\alpha^{\rm out}$. Thus

$$\text{Distance}\,[C_\alpha^{\rm out}(j), I_{C_\alpha}^{\rm out}(n)] = D_{\rm min}, \quad I_{C_\alpha}^{\rm out}(n) \in An_1. \tag{9}$$

Table 4
Scaling properties of 235-orbits.

Orbits	Radius (standard, as in Table 3)	Scaling	Generator
36 orbits of order 12 (icosahedral)			
$A33_9$	0.5878	$1/2\tau$	$\frac{1}{4}[\bar{1}1\bar{1}11\bar{1}]$
$A1_5, A22_7, A24_8$	0.7265	$1/\tau^2$	$\frac{1}{2}[3\bar{1}11111]$
$A27_6, A35_8$	0.9511	$1/2$	$\frac{1}{2}[100000]$
$A3_5, A5_5, A16_5, A18_6$	1.1756	$1/\tau$	$\frac{1}{2}[\bar{1}11111]$
$A28_5, A32_6$	1.5388	$\tau/2$	$\frac{1}{4}[111\bar{1}\bar{1}1]$
$A1_3, A2_4, A3_4, A4_4, A5_4, A6_4, A7_5, A8_5, A9_4, A10_5, A11_4, A12_5, A13_6, A15_3, A20_5, A38_7$	1.9021	1	[100000]
$A33_4$	2.4898	$\tau^2/2$	$\frac{1}{4}[311111]$
$A1_1, A6_3, A24_4, A25_5, A40_7$	3.0766	τ	$\frac{1}{2}[111\bar{1}\bar{1}1]$
$A2_1$	3.8042	2	[200000]
$A3_1$	4.9795	τ^2	$\frac{1}{2}[311111]$
45 orbits of order 20 (dodecahedral)			
$A18_7$	0.6616	$1/\tau^2$	$\frac{1}{4}[111\bar{3}3\bar{3}]$
...	...	...	...
$A55_1$	6.2665	$\tau + 2$	$\frac{1}{2}[333\bar{1}1\bar{1}]$
70 orbits of order 30 (icosidodecahedral)			
$A51_{16}$	0.3090	$1/4\tau$	$\frac{1}{4}[11\bar{1}00\bar{1}]$
...	...	...	...
$A54_1$	6.8540	$\tau^4/2$	$\frac{1}{2}[332002]$
316 orbits of order 60 (general)			
$A44_{10}$	0.8740	-	$\frac{1}{2}[01\bar{1}2\bar{1}1]$
...	...	...	...
$A13_1$	6.9266	-	$[110\bar{1}02]$

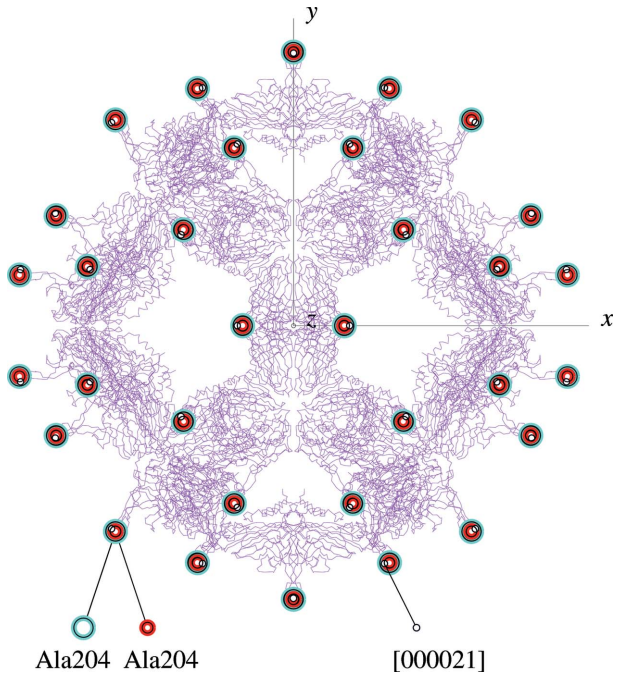

Figure 3
Outer fitting of the chain C with outer C_α residue Ala204, outer affine orbit $A26_1$, generator [000021] at $D_{\min} = 3.4$ Å and (accidentally) same nearest residue Ala204 at $d_{\min} = 3.4$ Å.

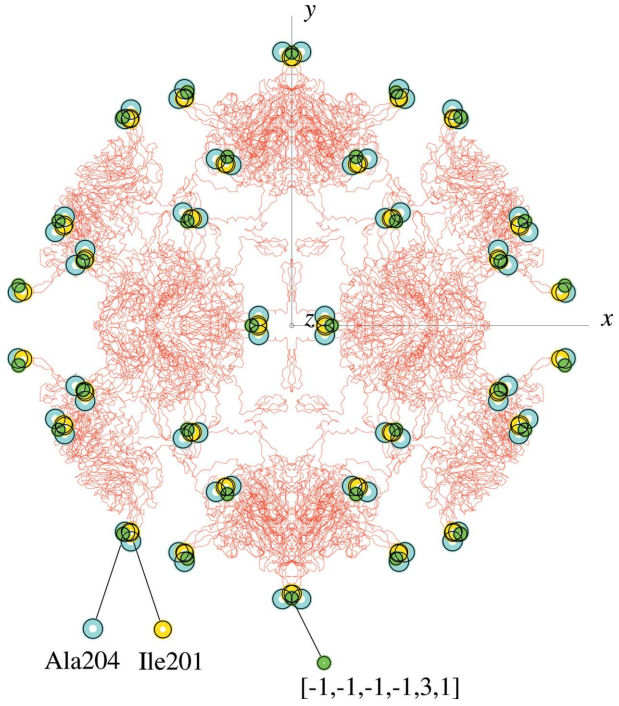

Figure 4
Outer fitting of the chain A (in red) with outer C_α residue Ala204, outer affine orbit $A13_1$, generator $[\bar{1}11\bar{1}31]$ at $D_{\min} = 7.1$ Å and nearest residue Ile201 at $d_{\min} = 4.7$ Å.

Note that the residue $C_\alpha^{\rm out}(j)$ is not necessarily the nearest one to the indexed position $I_{C_\alpha}^{\rm out}(n)$, so that another minimal distance $d_{\min}$ occurs, that between the indexed (rescaled) position and a residue $C_\alpha(j')$,

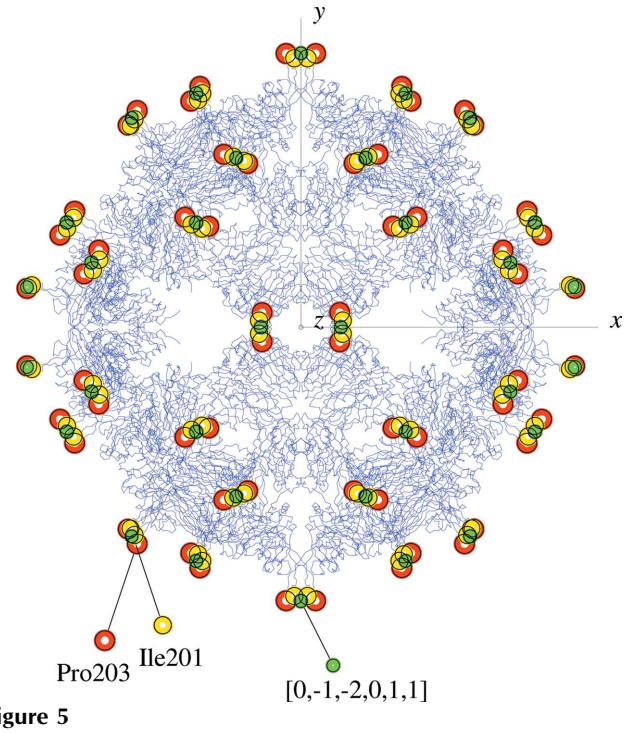

Figure 5
Outer fitting of the chain B (in blue) with outer C_α residue Pro203, outer affine orbit $A12_1$, generator $[0\bar{1}2011]$ at $D_{\min} = 9.1$ Å and nearest residue Ile201 at $d_{\min} = 5.2$ Å.

$$\text{Distance}\,[I_{C_\alpha}^{\rm out}(n),\, C_\alpha(j')] = d_{\min}. \qquad (10)$$

Applying these ideas to each of the polypeptide chains A, B, C of the Pariacoto virus, one finds the positions shown in Figs. 3, 4 and 5. The corresponding fitting parameters are reported in Table 5.

The icosahedral basis, as one finds in equation (1), relates the (rational) indices with the corresponding orthogonal coordinates of an indexed position. This basis (denoted as ico1) is the one adopted in Figs. 3, 4 and 5 of the outer fitting of the chains A, B, C of the Pariacoto virus, with the value of a_0 depending on the appropriate C_α rescaling. This basis, however, is not compatible with the orientation adopted in the plot of the tomato bushy stunt virus according to its PDB file. This incompatibility is demonstrated in Fig. 6, where the outer fitting of the chain A is plotted in a similar way as for the Pariacoto virus and using the same icosahedral basis ico1. One sees indeed that the minimal distances between orbit points of the outer residues Tyr244 and the corresponding rescaled indexed ones (in green) are not orbit invariant as they should be. The alternative basis ico3 adopted for the same fitting in Fig. 7 appears to be a compatible one. The basis elements compatible with the orientation used for various viruses are indicated in Table 6.

The icosahedral bases compatible with the cowpea chlorotic mottle virus and the Seneca Valley virus have not yet been identified.

In the York group's approach, the identification of the affine system(s) characterizing the architecture of a given virus does not follow the way presented here for the fitting between

Table 5
Fitting parameters of the (rescaled) outer orbits with the chains A, B and C of the Pariacoto virus.

Icosahedral basis ico1 = {(1, 0, τ)}], distances in Å.

Chain	Outer residue	r_{max}	Orbit	Generator	D_{min}	Nearest residue	d_{min}
A	C_α(197) = Ala204	171	$A13_1$	$[\overline{1}\overline{1}\overline{1}31]$	7.1	C_α(194) = Ile201	4.7
B	C_α(154) = Pro203	172	$A12_1$	$[0\overline{1}\overline{2}011]$	9.1	C_α(152) = Ile201	5.2
C	C_α(153) = Ala204	174	$A26_1$	[000021]	3.4	C_α(153) = Ala204	3.4

a given chain and an associated affine orbit. It makes use instead of a *best-fit algorithm*, as described in Keef *et al.* (2013), which involves several steps. In particular, a *sample preparation* (*a*), based on the viral structure one finds in the PDB, is followed by an *alignment and scaling* (*b*), a *sifting* (*c*) and finally by an evaluation of *goodness of fit* (*d*).

Here we do not try to reconstruct and discuss all these steps, but the whole is analysed from another point of view and restricted to the single representative test case of the Pariacoto virus, making use of the data presented in Fig. 4 of the same paper (Keef *et al.*, 2013).

From the orbit points indicated, whose radii range from 174 to 43 Å, one recognizes two affine systems: the first one denoted here as $A26$ of cardinality 290 and with six orbits (listed in Table 3) and a second one, the system $A15$ of cardinality 172 and with five orbits.

Table 6
Compatibility between virus orientation and corresponding icosahedral bases.

Viruses	Compatible icosahedral basis
PaV, MS2	ico1: $a_1 = (1, 0, \tau)$, $a_2 = (\tau, 1, 0)$, $a_3 = (0, \tau, 1)$, ..., $a_6 = (\tau, -1, 0)$
STMV	ico2: $a_1 = (1, \tau, 0)$, $a_2 = (-1, \tau, 0)$, $a_3 = (0, 1, \tau)$, ..., $a_6 = (0, 1, -\tau)$
TBSV, murine py, SV40, CMV	ico3: $a_1 = (0, 1, \tau)$, $a_2 = (-1, \tau, 0)$, $a_3 = (-\tau, 0, 1)$, ..., $a_6 = (1, \tau, 0)$

The $A26$ system is rescaled with respect to the outer residue Ala204 of the chain C, which in this virus has the maximal radial distance $r_{max} = 174$ Å from the centre. The rescaled orbits have then the radii given by: 174 (26_1), 162 ($A26_2$), 140 ($A26_3$), 131 ($A26_4$), 114 ($A26_5$) and 94 ($A26_6$). The fitting of this affine system to the Pariacoto virus is shown in Fig. 8.

The affine system $A15$ belongs to the associated *skeletal configuration* discussed in Indelicato *et al.* (2012), given here as

$$A_{\text{skeletal}} = A26 \cup \frac{1}{2} A15 \tag{11}$$

and rescaled accordingly, with the rescaling factors k_{26} = 81.3 Å and $k_{15} = 1/2k_{26}$, so that the whole only depends on one fitting parameter, the r_{max}. The five rescaled orbits of $A15$ then have the radii: 112 ($A15_1$), 92 ($A15_2$), 77 ($A15_3$), 70 ($A15_4$) and 43 ($A15_5$). The corresponding fitted orbit points are shown

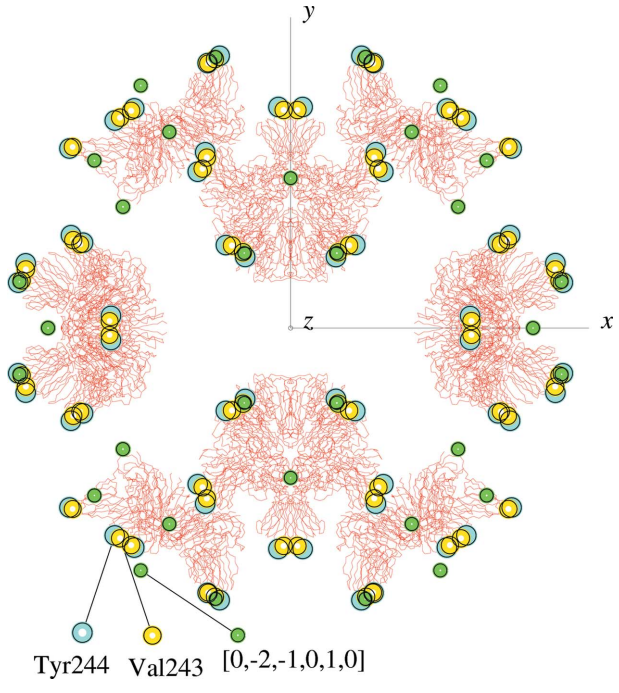

Figure 6
Incompatibility of the icosahedral basis ico$_1$: $a_1 = (1, 0, \tau)$, ... used for the fitting of the outer C_α residue Tyr244 belonging to the chain A of the tomato bushy stunt virus with the affine system $A7$, the outer affine orbit $A7_1$, the generator $[0\overline{2}1010]$ and nearest residue Val243. Compare the affine orbit points (in green) with those in Fig. 7.

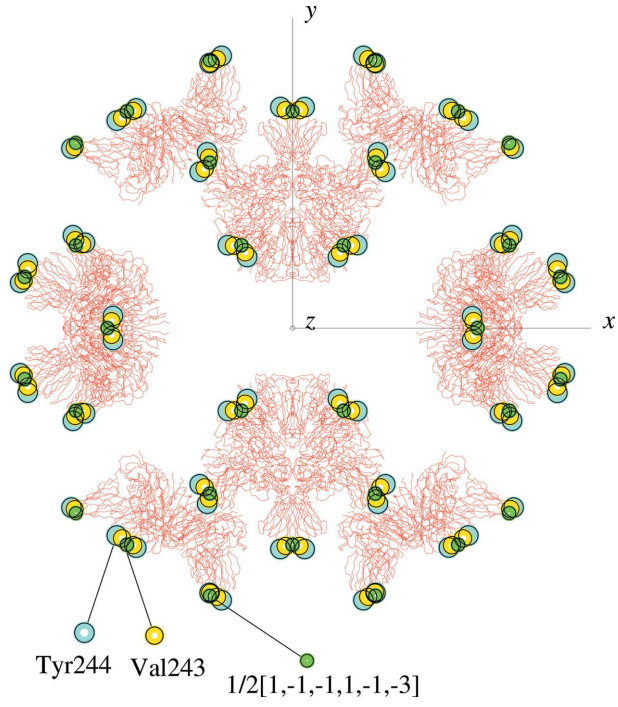

Figure 7
Compatibility of the alternative icosahedral basis ico$_3$: $a_1 = (0, 1, \tau)$, ... used for the fitting of the outer C_α residue Tyr244 belonging to the chain A of the tomato bushy stunt virus with the affine system $A5$, the outer affine orbit $A5_1$, the generator $\frac{1}{2}[1\overline{1}1\overline{1}\overline{3}]$ and nearest residue Val243. Compare the affine orbit points (in green) with those in Fig. 6.

Table 7
Minimal distances (in Å) of the Pariacoto chains with respect to the rescaled orbits of $A26$ and $A15$.

Orbit	Order	A	B	C	D	E	F	R
26_1	60	6.8	6.4	5.9	*59.3*	*63.6*	*63.9*	*68.1*
26_2	60	*33.9*	*15.5*	*17.2*	*66.2*	*48.2*	*46.9*	*49.3*
26_3	60	2.6	5.6	*12.7*	*27.2*	*27.1*	*29.7*	*37.5*
26_4	30	*14.6*	*14.9*	4.0	*35.1*	*34.3*	*30.4*	*24.5*
26_5	20	*10.9*	*20.0*	*17.1*	*50.2*	*19.7*	*18.9*	*11.4*
26_6	60	8.6	*12.3*	*14.7*	*18.5*	*12.0*	*12.7*	*15.5*
15_1	60	*16.6*	9.7	9.1	*23.6*	5.6	4.9	*11.6*
15_2	60	*13.4*	*27.0*	*23.1*	*39.2*	*21.7*	*22.4*	8.0
15_3	12	*36.6*	*36.4*	*43.0*	*35.5*	*32.5*	*37.1*	*27.4*
15_4	20	*37.3*	*48.8*	*44.5*	*58.0*	*42.9*	*43.8*	*24.8*
15_5	20	*60.1*	*66.8*	*64.5*	*74.4*	*65.1*	*66.4*	*50.5*

in Fig. 9. The radial values indicated are exactly those one finds in Keef *et al.* (2013).

The goodness of fit of these indexed positions is here evaluated in terms of the residues of all the polypeptide chains A, B, C, D, E, F and the P-backbone positions of the RNA chain R of the Pariacoto virus having minimal distances $d_{\min}$ within a 10 Å range ($d_{\min} < 10$ Å) from one of the indexed positions of the skeletal configuration.

The result is shown graphically in Figs. 10, 11 and 12. All minimal distances of the various chains with respect to the rescaled orbits of $A26$ and $A15$ are indicated in Table 7. Given in italics are the distances larger than the 10 Å limit.

As can be seen, the chain D and the orbits of order 12 (15_3), 20 ($A26_5$, $A15_4$, $A15_5$) and 60 (26_2) are outside the 10 Å range and can be considered as not containing valuable structural information for the architecture of the Pariacoto virus. The 10 Å limit is not fully assumed *ad hoc*. It reflects the expected spreading between viral positions and related indexed ones and is indicative only. In any case, the values outside the 10 Å range are also indicated and all the relations within this limit are useful and not trivial. We recall that they depend on one fitting parameter only. Moreover, please note that there are points located more towards the interior of the capsid, which is occupied by genomic RNA but for which no data are available in the PDB file.

Similar situations are expected to occur in all the other virus cases considered so far.

6. Towards a toolkit for icosahedral viruses

The discussion of the relations between the skeletal configuration and the structural properties of the Pariacoto virus presented in the previous section is an indication that the goal of characterizing the viral architecture in terms of affine systems derived from an extension of the icosahedral group is not fully realised.

This situation supports the idea of specializing the best-fit algorithm mentioned above to a more specific toolkit for the architecture of icosahedral viruses.

The presence in our fitting approach of two conceptually different minimal distances ($D_{\min}$ and $d_{\min}$) suggests consideration of two complementary problems: the fitting of an indexed position nearest to a given residue (at distance $D_{\min}$) and the finding of the residue (of more generally of a back-

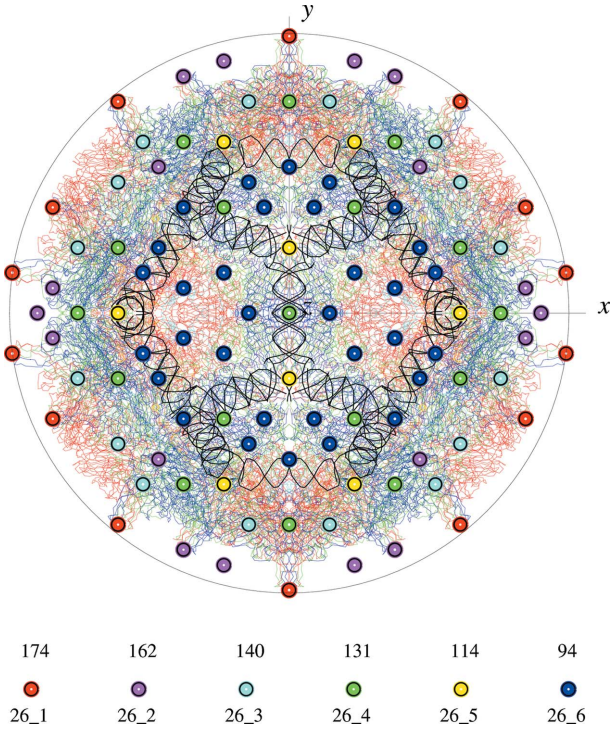

Figure 8
Pariacoto virus: affine system $A26$ rescaled by $r_{\max} = 174$, the radial distance of the outer residue Ala204 of chain C [compare with Fig. 2 of this article and Fig. 4 of Keef *et al.* (2013)].

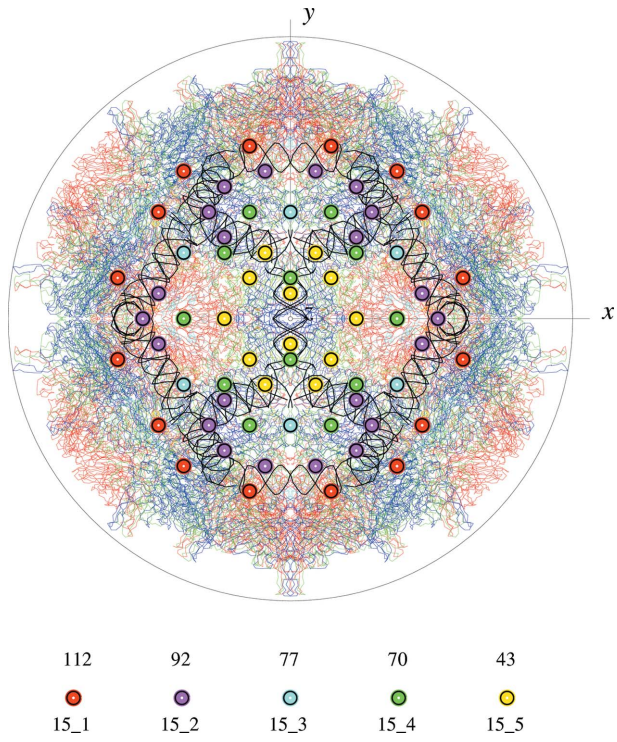

Figure 9
Pariacoto virus: affine system $A15$ of the skeletal configuration $A26 \cup \frac{1}{2}A15$, rescaled by the factor $k_{15} = 40.65$ Å $= \frac{1}{2}k_{26}$, where k_{26} (Å) is the rescaling factor of $A26$, as in the previous figure.

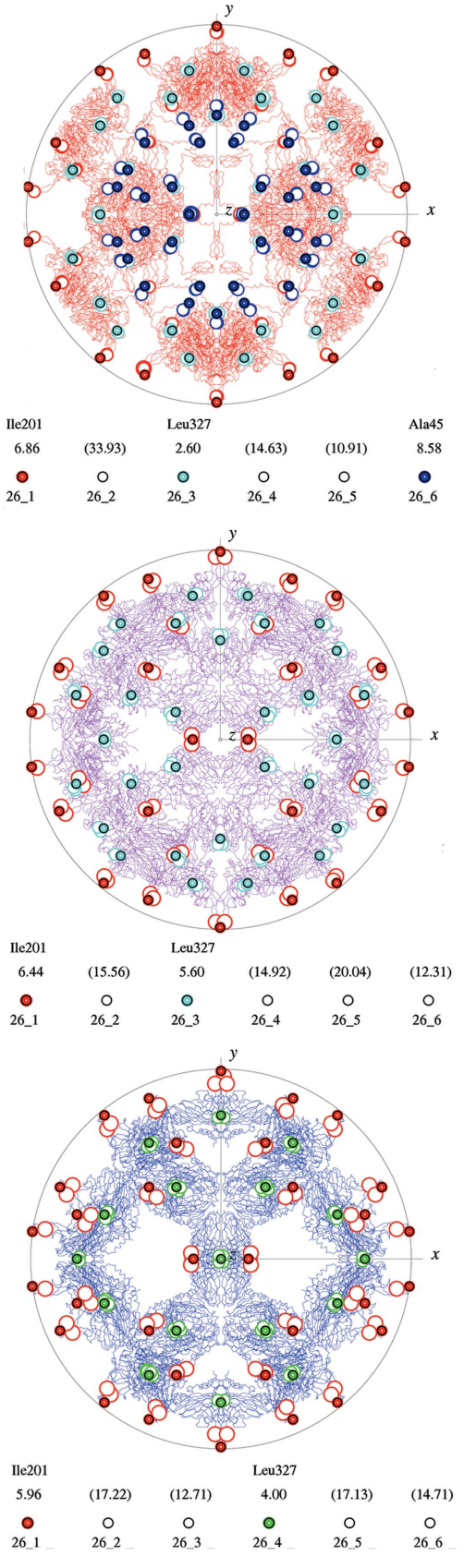

Figure 10
Pariacoto virus: residues of the chains *A* (top view), *B* (middle view) and *C* (bottom view) within a 10 Å range from the various orbits of the affine system *A*26 (compare with Fig. 8).

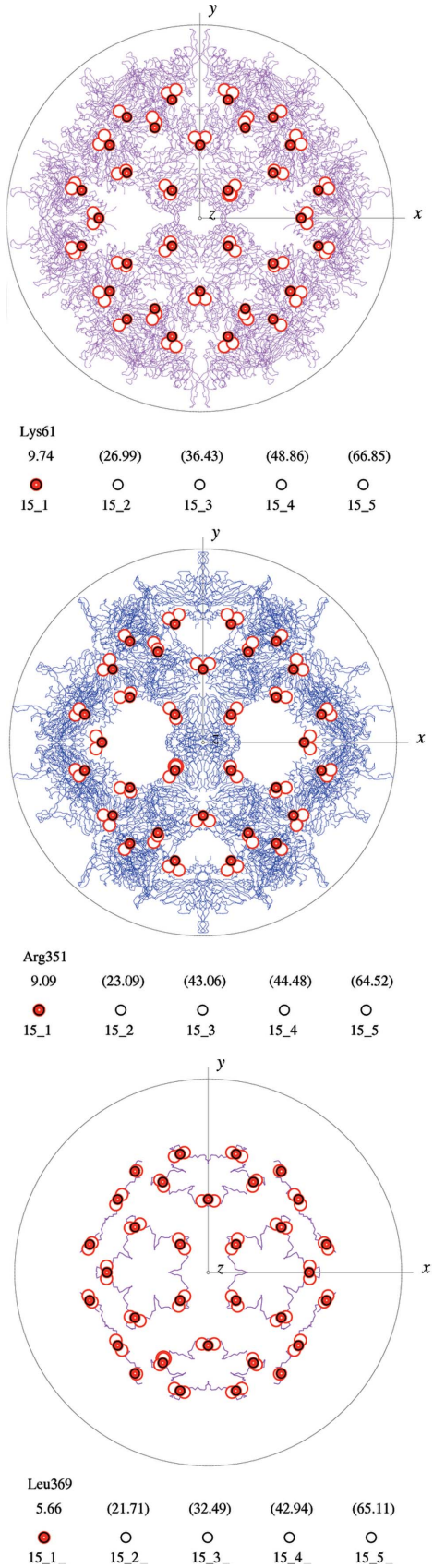

Figure 11
Pariacoto virus: residues of the chains *B* (top view), *C* (middle view) and *E* (bottom view) within a 10 Å range from the various orbits of the affine system *A*15 (compare with Fig. 10).

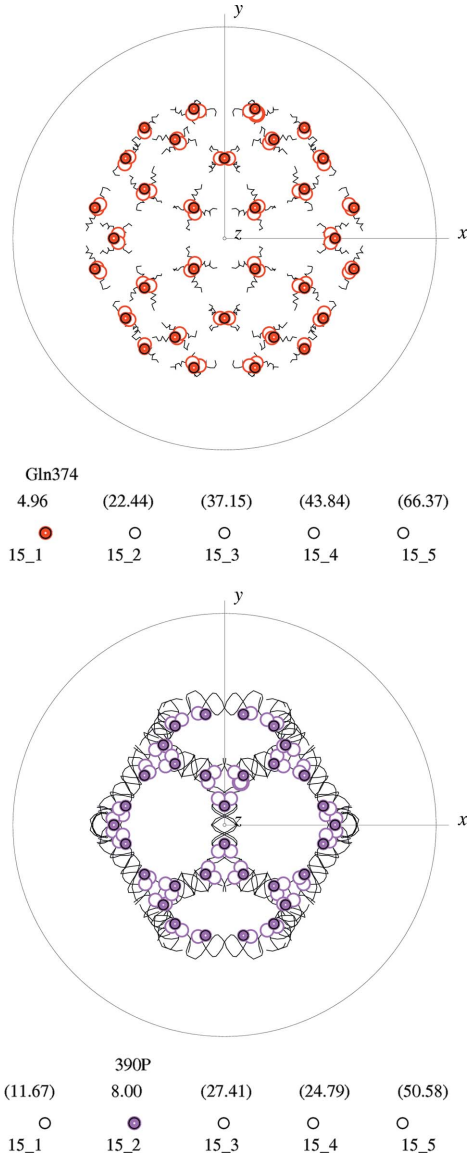

Figure 12
Pariacoto virus: residues of the chains *F* (top view) and backbones *P* of the RNA chain *R* (bottom view) within a 10 Å range from the various orbits of the affine system *A*15 (compare with Figs. 10 and 11).

bone position) nearest to a given indexed position (at distance $d_{\min}$).

Only positions with small integral indices are structurally relevant, as one knows from the point-group symmetry of crystal growth forms, and this requirement can be generalized to rational indices as well. The affine extension of the icosahedral group, to low order of the translations considered (here to the first order only), can be seen as a way to generate finite sets of *structure-adapted low-indices positions* which possibly are at different radial distances and are, therefore, related by transformations not limited to the icosahedral symmetry ones.

6.1. From a given residue to an indexed position

As an example of how to find an indexed position fitted to a starting residue one looks for the generator of a rescaled orbit

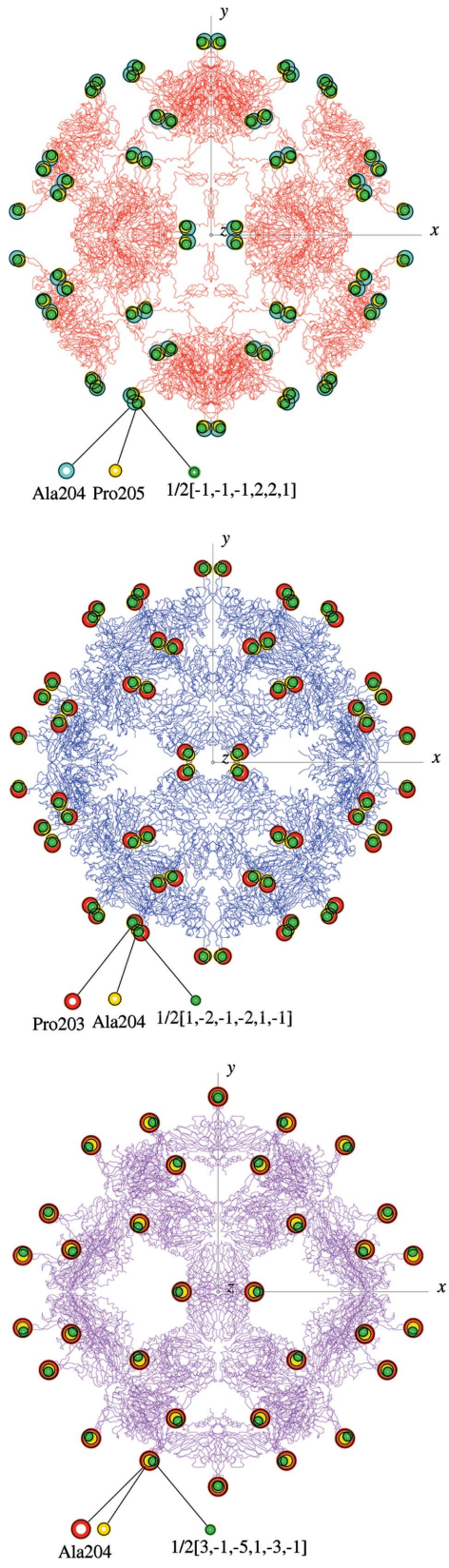

Figure 13
Pariacoto virus. Chain *A* (top view): from residue Ala204 to indexed position $\frac{1}{2}[\bar{1}\bar{1}\bar{1}221]$ of $A47_3$ with nearest Pro205. Chain *B* (middle view): from residue Pro203 to indexed position $\frac{1}{2}[1\bar{2}\bar{1}2\bar{1}\bar{1}]$ of $A44_3$ with nearest Ala204. Chain *C* (bottom view): from residue Ala204 to indexed position $\frac{1}{2}[3\bar{1}5\bar{1}3\bar{1}]$ of $A43_5$ with nearest Ala204.

Table 8
Fitting parameters of the (rescaled) affine orbits of order 60 with the outer residues C_α^0 of the chains A, B and C of the Pariacoto virus.

Compare with Table 5. Distances are in Å.

Chain	Given residue	Radius	Indexed position	Orbit	$D_{\min}$	Nearest residue	$d_{\min}$
A	C_α^0 = Ala204	170.95	$I^0 = \frac{1}{2}[\overline{1}\overline{1}\overline{1}221]$	47_3	2.66	C_α = Pro205	2.03
B	C_α^0 = Pro203	171.67	$I^0 = \frac{1}{2}[1\overline{2}\overline{1}21\overline{1}]$	44_3	2.59	C_α = Ala204	1.76
C	C_α^0 = Ala204	173.60	$I^0 = \frac{1}{2}[3\overline{1}\overline{5}\overline{1}31]$	43_5	3.49	C_α = Ala204	3.49

Table 9
Fitting parameters of the residues belonging to the chains A, B and C of the Pariacoto virus, nearest to a (rescaled) indexed affine position, with site symmetry 5, 3 and 2, respectively.

Distances are given in Å.

Generator I^0	Orbit	Site symmetry	Chain	C_α	Radius	$d_{\min}$
[000010]	$A1_3$	5	A	Met175	157.4	4.7
			B	Asn49	104.3	28.2
			C	Leu340	111.0	38.6
[100011]	$A12_2$	3	A	Ala21	124.8	3.9
			B	Met175	137.0	5.3
			C	Asn173	136.5	5.1
[0$\overline{1}$0010]	$A2_2$	2	A	Gly33	118.0	8.0
			B	Arg220	139.4	13.8
			C	Thr140	103.4	3.0

of order 60 at the nearest distance $D_{\min}$ from the given residue. In the present case we choose an outer residue (at maximal radial distance $r_{\max}$) of a given polypeptide chain.

The choice of a general affine orbit is based on the idea that, generically, an outer residue also gives rise by the action of the icosahedral group to an orbit of order 60. This problem has already been solved for outer chains in §4, the difference being that there the restriction was imposed to the affine orbits with maximal standard radius instead of orbits of order 60.

The result for the chains A, B and C of the Pariacoto virus is presented in Fig. 13 and in Table 8, to be compared with Figs. 3, 4 and 5, and Table 5, respectively. So, for example, the starting residue C_α^0 of the chain A is Ala204 having the maximal radius $r_{\max}(C_\alpha^0) = 170.95$ Å. The fitted (rescaled) orbit of order 60 is $A47_3$. The nearest indexed position of this affine orbit is $I^0(C_\alpha) = 1/2[\overline{1}\overline{1}\overline{1}221]$. The minimal distance between C_α^0 and I^0 is $D_{\min} = 2.66$ Å, whereas the residue nearest to I^0 is C_α = Pro205 at a distance $d_{\min} = 2.03$ Å.

6.2. From indexed position to nearest residue(s)

The problem is to determine the nearest residue to a given indexed position. Actually, this problem has already been solved in §4 for outer affine orbits and in §6.1 for orbits of order 60.

Considered here are positions belonging to affine orbits of order 12, order 20 and order 30, which have positions with site symmetry 5, 3 and 2, respectively. Determined are the residues at minimal distance $d_{\min}$ from the given indexed positions and this for the polypeptide chains A, B and C of the Pariacoto virus.

The results are shown in Figs. 14, 15 and 16, and the corresponding fitting parameters are given in Table 9. Note that the results for each type of orbit (icosahedral, dodecahedral or icosidodecahedral) are independent of the chosen representative because of mutual scaling equivalence (as pointed out in Table 4) so that, after rescaling, the images are correspondingly the same. The rescaling factor and thus the images of the indexed position, however, do depend on the chain of the fitted residue.

7. Conclusions and perspectives

The physics of viruses (which in this paper are assumed to have icosahedral symmetry) can be approached directly or through a preliminary geometrical characterization of their structure.

One can distinguish between three different geometrical approaches based either on tiling models, or on affine extensions of the icosahedral group (developed by the York group), or on molecular crystallography (a concept introduced by the present author).

The seminal classification scheme of Caspar & Klug (1962) is of the tiling type and leads to the so-called T-numbers. A tiling description is also possible for viruses not fitting into the Caspar–Klug scheme (Twarock, 2004). The relations between T-numbers, affine extensions and the alternative molecular crystallographic classification of indexed polyhedra have been discussed elsewhere (Keef & Twarock, 2009*b*; Janner, 2006).

The need to go beyond the icosahedral symmetry, typical for the last two approaches, follows from the existence of structural relations between viral backbone positions at different radial distance from the centre, which are, therefore, non-equivalent with respect to the icosahedral symmetry group. This allows one, in particular, to extend the structural characterization of the capsid to the genomic positions (not considered by Caspar & Klug).

Common to the York group's approach and my own approach is the fact that the additional structural relations are obtained from geometrical transformations of infinite order, like translations and scaling. The number of atomic positions involved in these relations is finite, like the different atomic positions in a virus. This requires, therefore, a truncation. The price is that the truncated set of transformations admitted does not form a group and one can only speak of *structural relations* and no longer of symmetry. An attempt to recover a finite group for the truncated set has been formulated, for special cases only, in terms of higher-dimensional crystallographic point groups projected in three dimensions in a similar way as for the icosahedral group (Janner, 2008*d*).

The present article tries to arrive at an understanding of the approach mainly developed at the York Centre for Complex Analysis of the University of York by T. Keef and R. Twarock of an affine extended icosahedral group, with the aim of characterizing the architecture of icosahedral viruses. It

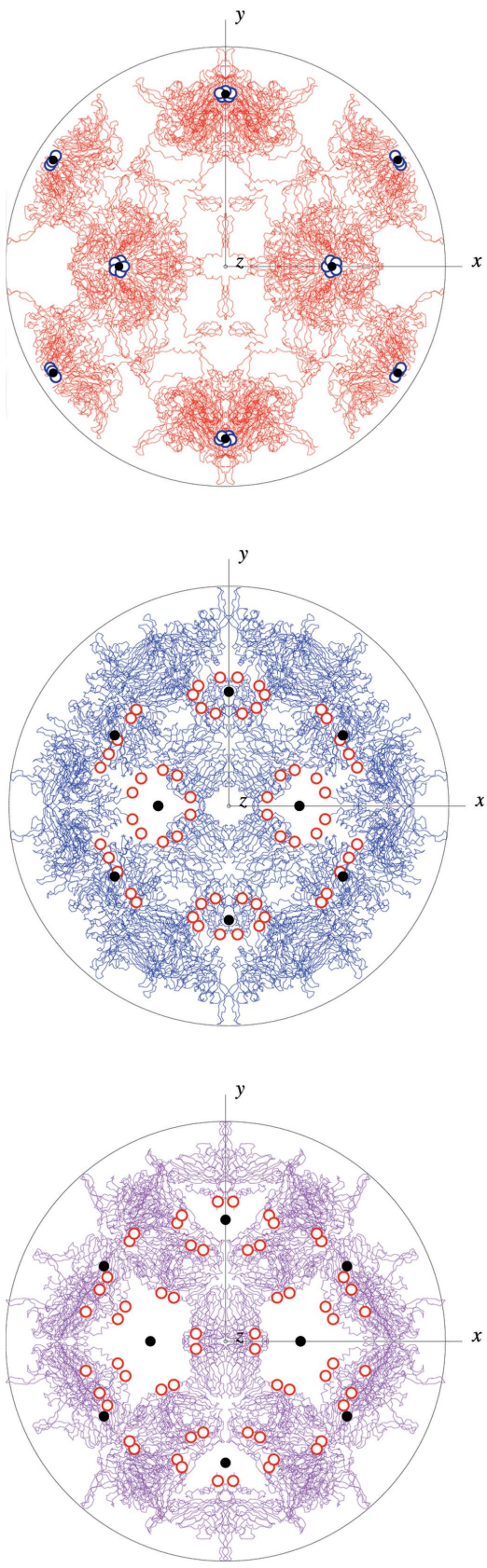

Figure 14
Pariacoto virus, icosahedral orbit $A1_3$ (black points), generator [000010]. Nearest residue: Met175 of chain *A* (top view), Asn49 of chain *B* (middle view) and Leu340 of chain *C* (bottom view) (see Table 9).

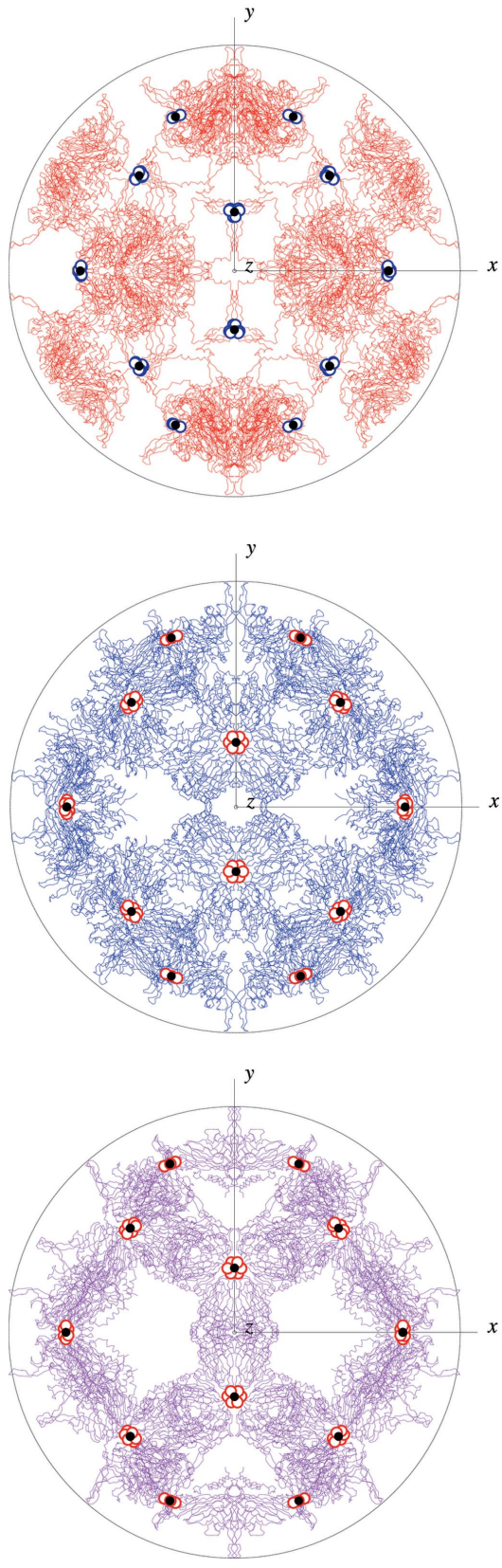

Figure 15
Pariacoto virus, dodecahedral orbit $A12_2$ (black points), generator [100011]. Nearest residue: Ala21 of chain *A* (top view), Met175 of chain *B* (middle view) and Asn173 of chain *C* (bottom view) (see Table 9).

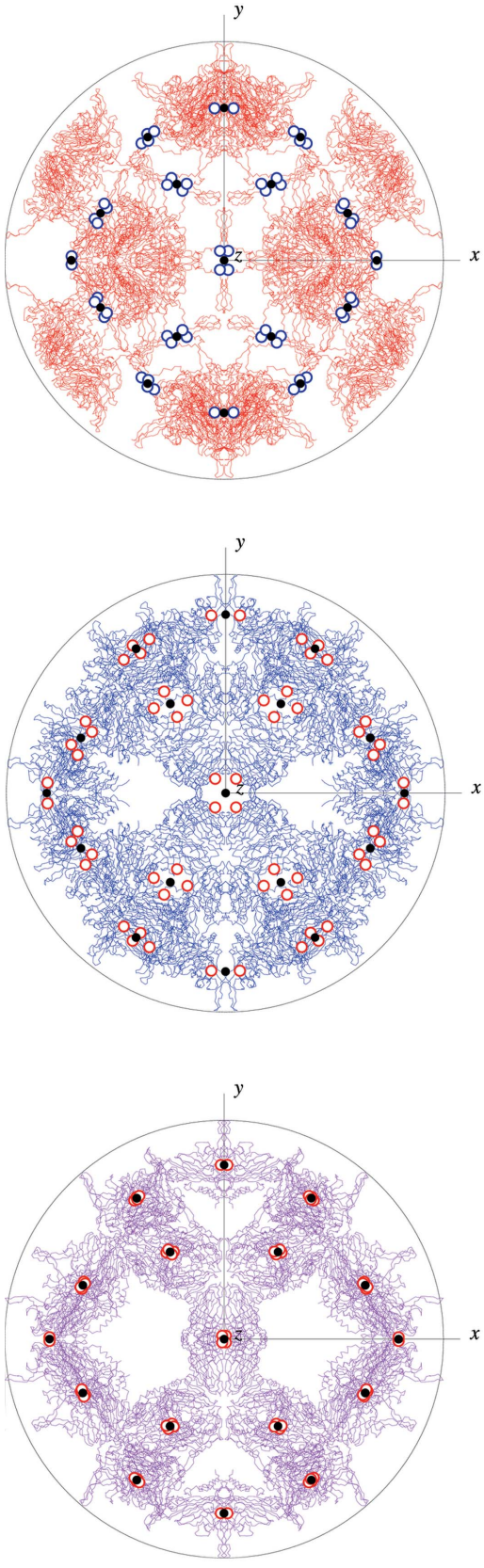

Figure 16
Pariacoto virus, icosidodecahedral orbit $A2_2$ (black points), generator [$0\bar{1}0010$]. Nearest residue: Gly33 of chain *A* (top view), Arg220 of chain *B* (middle view) and Thr140 of chain *C* (bottom view) (see Table 9).

represents a first step only, as the connections with alternative approaches like the one developed by the author in a number of publications have not yet been worked out in detail. Only some basic fundamental ideas dealing with what is denoted as molecular crystallography have been taken into account here.

To complete the present discussion, the affine systems and the affine orbits involved should be specified, as has been done here for the Pariacoto virus, also in all other viruses whose architecture has already been characterized according to the Keef–Twarock–Wardman approach.

Moreover, the problem of the compatible icosahedral basis for viruses like the Seneca Valley virus and the cowpea chlorotic mottle virus, mentioned in §5, should be considered in general and not only *ad hoc*; at least so far it has not yet been solved by means of the best-fit algorithm (Keef *et al.*, 2013).

Despite the preliminary character of the present work, some general conclusions can be drawn leading to perspectives open for further investigations.

(i) The systems of an *affine extended icosahedral group* allow a one-parameter characterization of structural properties of both the viral capsid and the genome.

(ii) Not every viral chain, nor every orbit of the fitted affine systems is involved in this characterization.

(iii) By the affine extension method, one generates a fairly large number of positions with integral (and rational) low indices, which appear to be relevant for the structural characterization of viruses.

(iv) From the York group's *best-fit algorithm*, a more specific toolkit should be developed for the analysis of the architecture of icosahedral viruses involving structural relations based, for example, on: indexed backbone positions; regions with a given site point symmetry; crystallographic scalings; polyhedral enclosing forms having vertices at positions with rational indices; and symmetry properties of lattice-periodic packed structures.

In particular, similar relations can be expected between the affine orbit points fitted sequentially to the C_α's of the primary structure of the various chains and turning points of their secondary structure (α-helices, β-strands and loops) as has been shown to be the case for octahedral holoenzymes (Janner, 2008*b*,*c*).

In this purely geometrical approach, one should keep in mind that the final goal is a better understanding of the physical, chemical and biological properties of viruses involving typical phenomena like mutation, conservation, maturation and expansion. This allows possible comparisons between viruses of the same family but with different serotype, and viruses of different families.

As already known, the approach can be generalized to other biological systems, like proteins with a given axial point-group symmetry.

A better insight into the research activity of the York group has been obtained as a result of a visit to the York Centre for Complex Systems Analysis, the University of York, at the

beginning of May 2012. The valuable comments made by these colleagues during the visit are gratefully acknowledged. Thanks are expressed to one of the referees whose suggestions helped to improve the manuscript.

References

Caspar, D. L. D. & Klug, A. (1962). *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24.
Hopper, P., Harrison, S. C. & Sauer, R. T. (1984). *J. Mol. Biol.* **177**, 701–713.
Indelicato, G., Cermelli, P., Salthouse, D. G., Racca, S., Zanzotto, G. & Twarock, R. (2012). *J. Math. Biol.* **64**, 745–773.
Janner, A. (2004). *Acta Cryst.* A**60**, 198–200.
Janner, A. (2006). *Acta Cryst.* A**62**, 319–330.
Janner, A. (2008*a*). *Models, Mysteries and Magic of Molecules*, ch. 11, edited by J. C. A. Boeyens & J. F. Ogilvie, pp. 233–254. Dordrecht: Springer.
Janner, A. (2008*b*). *Acta Cryst.* A**64**, 494–502.
Janner, A. (2008*c*). *Acta Cryst.* A**64**, 503–512.
Janner, A. (2008*d*). *Acta Cryst.* A**64**, 280–283.
Janner, A. (2010). *Acta Cryst.* A**66**, 312–326.
Janner, A. (2011*a*). *Acta Cryst.* A**67**, 174–189.
Janner, A. (2011*b*). *Acta Cryst.* A**67**, 517–520.
Janner, A. (2011*c*). *Acta Cryst.* A**67**, 521–532.
Keef, T. & Twarock, R. (2008). *Comput. Math. Methods Med.* **9**, 221–229.
Keef, T. & Twarock, R. (2009*a*). *J. Math. Biol.* **59**, 287–313.
Keef, T. & Twarock, R. (2009*b*). *Emerging Topics in Physical Virology*, edited by P. G. Stockley & R. Twarock, p. 332. London: Imperial College Press.
Keef, T., Twarock, R. & Elsawy, K. M. (2008). *J. Theor. Biol.* **253**, 808–816.
Keef, T., Wardman, J. P., Ranson, N. A., Stockley, P. G. & Twarock, R. (2013). *Acta Cryst*. A**69**, 140–150.
Larson, S. B., Day, J., Greenwood, A. & McPherson, A. (1998). *J. Mol. Biol.* **277**, 37–59.
Smith, T. J., Chase, E., Schmidt, T. & Perry, K. L. (2000). *J. Virol.* **74**, 7578–7586.
Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. (1995). *Structure*, **3**, 63–77.
Stehle, T., Gamblin, S. J., Yan, Y. & Harrison, S. C. (1996). *Structure*, **4**, 165–182.
Stehle, T. & Harrison, S. C. (1996). *Structure*, **4**, 183–194.
Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M. & Johnson, J. E. (2001). *Nat. Struct. Biol.* **8**, 77–83.
Twarock, R. (2004). *J. Theor. Biol.* **226**, 477–482.
Valegård, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. & Liljas, L. (1997). *J. Mol. Biol.* **270**, 724–738.
Venkataraman, S., Reddy, S. P., Loo, J., Idamakanti, N., Hallenbeck, P. L. & Reddy, V. S. (2008). *Structure*, **16**, 1555–1561.