PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2006) **364**, 3357–3373
doi:10.1098/rsta.2006.1900
Published online 19 October 2006

Mathematical virology: a novel approach to the structure and assembly of viruses

By Reidun Twarock*

Departments of Mathematics and Biology, University of York, York YO10 5DD, UK

Understanding the structure and life cycle of viruses is a fascinating challenge with a crucial impact on the public health sector. In the early 1960s, Caspar & Klug (Caspar & Klug 1962 *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24) established a theory for the prediction of the surface structures of the protein shells, called viral capsids, which encapsulate and hence provide protection for the viral genome. It is of fundamental importance in virology, with a broad spectrum of applications ranging from the image analysis and classification of experimental data to the construction of assembly models. However, experimental results have provided evidence for the fact that it is incomplete and, in particular, cannot account for the structures of *Papovaviridae*, which are of particular interest because they contain cancer-causing viruses. This gap has recently been closed by the viral tiling theory, which describes the locations of the protein subunits and inter-subunit bonds in viral capsids based on mathematical tools from the area of quasicrystals. The predictions and various recent applications of the new theory are presented, and it is discussed how further research along these lines may lead to new insights in virology and the design of anti-viral therapeutics.

Keywords: virus structure and assembly; viral tiling theory; Caspar–Klug theory; quasicrystals; Coxeter groups

1. Introduction

Viruses are fascinating micro-organisms consisting of a very compact genome and a protective protein shell or capsid, which hijack host cells typically 100 to 1000 times their size. Advances in virology and the design of anti-viral therapeutics rely strongly on an understanding of the viral replication cycle and, in particular, of the structures of the capsids, as well as of the mechanisms that trigger their assembly and disassembly.

Already, Crick & Watson (1956) observed that the majority of viruses exhibits symmetry in the structural organization of their capsids. It soon became clear that icosahedral symmetry occurs predominantly (see, for example, Casjens (1985) for a review), although an explanation for the origin of icosahedral symmetry in viruses was suggested only recently by Zadi *et al.* (2004). The icosahedral symmetry implies that the capsids exhibits 6 fivefold, 10 threefold and 15 twofold discrete rotational symmetry axes and that their surface structures are hence highly ordered. Experiments have provided evidence for the

*rt507@york.ac.uk

One contribution of 23 to a Triennial Issue ‘Mathematics and physics’.

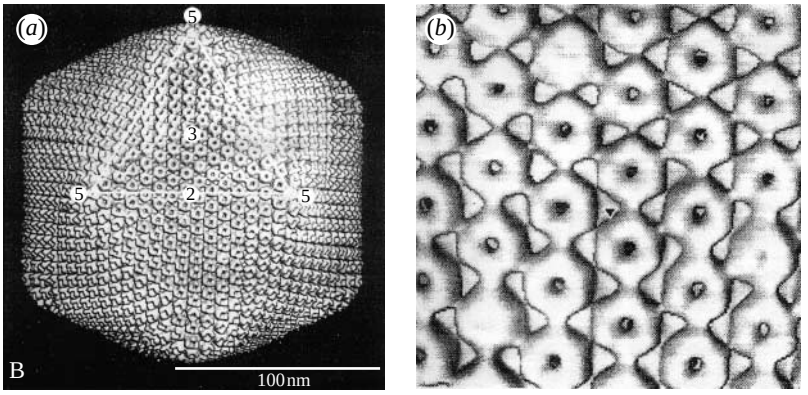

Figure 1. (*a*) Example of a viral capsid, with (*b*) capsomeres shown in magnification. Both figures are obtained from the Johnson Lab at the Scripps Research Institute.

fact that viruses follow particular types of surface lattices as blueprints for their structural organization. The proteins typically appear in clusters of three, five or six, called capsomeres that are organized as shown for example in figure 1*a*; capsomeres composed of six individual protein subunits are shown in magnification in figure 1*b*.

This suggests that it should be possible to pinpoint a uniform organizational principle for viral capsids in general. The first theory of this kind was proposed by Caspar & Klug (1962) in their seminal paper from the early 1960s. In this paper, they derive a series of polyhedra that encode the locations of the proteins in the viral capsids of icosahedral viral particles. This theory is one of the major tools in modern virology, universally accepted for the classification of viral capsids and the three-dimensional reconstructions of viral capsids from experimental data (e.g. Baker *et al.* (1999) and references within). Therefore, it is not surprising that the first experimental results deviating from Caspar–Klug theory have initially been thought to be incorrect, until it had been established that they rather have to be seen as pointers to the incompleteness of the theory. For example, the *Papovaviridae*, which contain cancer-causing viruses and are hence of prime importance for the public health sector, fall out of the scope of this theory (e.g. Rayment *et al.* 1982; Liddington *et al.* 1991).

The mathematical reason for this discrepancy lies in the fact that Caspar & Klug only consider the surface lattices that are induced by hexagonal lattices, so that pentamers, i.e. clusters of five individual protein subunits, located off the fivefold axes of icosahedral symmetry are excluded *a priori*. However, there are icosahedral viruses with a larger number of pentamers, such as polyomavirus with 72 pentamers. Its structural organization has therefore been a long-standing open problem in virology, formulated by Liddington *et al.* (1991) as follows: 'The puzzle is how do the (coloured) pentamers fit into the hexavalent holes?'.

Viral tiling theory has been introduced in 2004 to close this gap (e.g. Twarock 2004; featured in Science News by Weiss (2005)). By exploiting the concept of symmetry to the full, it makes use of generalized grids that are determined via the affinization of the non-crystallographic Coxeter group H_3 based on a method inspired by the projection formalism (e.g. Senechal 1996) known from the theory of quasicrystals (Shechtman *et al.* 1984) and Penrose tilings (Penrose 1974).

This more involved mathematical procedure is necessary because lattices invariant under the icosahedral group do not exist in two or three dimensions, so that a straightforward generalization of the Caspar–Klug construction is not possible. The new theory is well suited to the description of the capsid structures of *Papovaviridae* while still reproducing the tessellations relevant to the viruses covered by the Caspar–Klug classification. In particular, it has led to a new finite series of polyhedra, called the triacontahedral series, that complements the family of polyhedra in the Caspar–Klug theory (Keef & Twarock in press *a*). Moreover, its predictive power and scope of applications are significantly enhanced with respected to Caspar–Klug theory, because besides the locations of the proteins it also predicts the locations of the inter-subunit bonds.

The new tiling models have already sparked various applications, ranging from assembly models and the classification of tubular malformations to the description of crosslinking structures. But, by far, this does not exhaust the rich mathematical toolbox provided by this theory. For example, the generalized grids that have been derived for the construction of the polyhedra have interesting scaling properties that can be used in connection with a recent approach by Janner (2006) to derive encasing forms for the proteins in viral capsids. This departure from two-dimensional surface lattices to a three-dimensional representation of the capsid is a major step that opens up a plethora of novel applications, including the study of scaffold-mediated assembly, which is discussed in §5. Moreover, the connection between the locations of the protein subunits and the sites of a higher-dimensional periodic lattice that has been worked out by Keef & Twarock (in press *a*) allows to address the physical aspects of viral capsids, such as their elasticity, phonons and phason flips. In addition, these grids can be used to construct nested shell structures that lend themselves to the modelling of the genome packaging structure, as discussed in §5, with the potential to solve another open problem in virology.

2. Caspar–Klug theory

The theory developed by Caspar & Klug (1962) is the first approach to the prediction of viral capsid architecture. It is applicable to icosahedral viruses that exhibit protein subunits organized according to a hexagonal surface lattice, such as the viral capsid in figure 1. In order to predict the locations and relative orientations of the protein clusters (capsomeres) for viruses of this type, they consider embeddings of the surface of an icosahedron into a hexagonal lattice, as shown in figure 2*a*.

Every inequivalent embedding of this type—allowing for rotations and scalings of the icosahedral surface, such that vertices of the triangular faces meet centres of the hexagons in the underlying lattice—corresponds to a different blueprint for a viral capsid according to their theory. In particular, via a subsequent replacement of each hexagon by six triangles, triangulations are obtained that are compatible with the overall icosahedral symmetry. Each triangular facet in such a triangulation encodes the locations of three protein subunits in its corners. Hence, one obtains clusters of five protein subunits, called pentamers, at the fivefold axes of icosahedral symmetry and clusters of six protein subunits, called hexamers, otherwise. This is demonstrated in figure 2*b*,

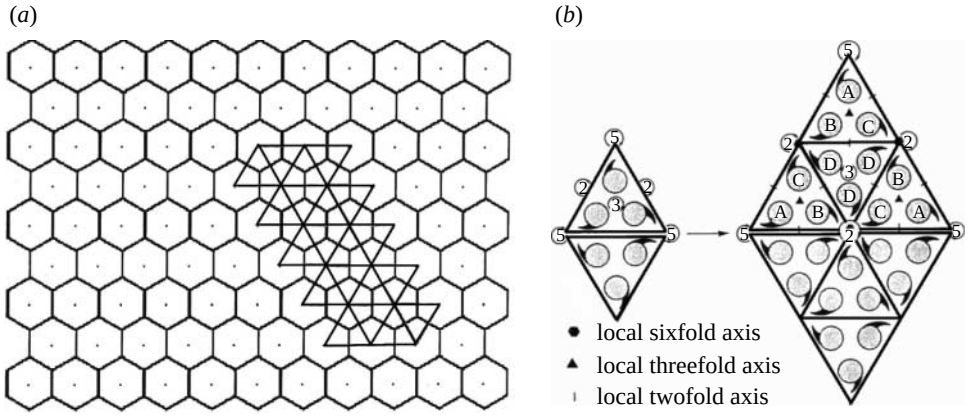

Figure 2. (*a*) An embedding of the surface of an icosahedron into a hexagonal lattice and (*b*) an example of a triangulation induced in this way. Adapted from Casjens (1985).

adapted from Casjens (1985). On the left, two of the 20 triangular faces of an icosahedron are shown, with a protein subunit marked schematically in each of their corners. By introducing further triangular facets via the method explained above, the faces are subdivided into triangular facets. A subdivision of each face into four triangular facets is shown on the right, each of them again encoding the locations of three protein subunits in their corners. In this way, the locations and orientations of the capsomeres are specified; for example, a hexamer has appeared on the edge between the two original faces.

The different surface lattices obtained in this way have been classified in terms of the *triangulation number* $T = h^2 + hk + k^2$, with $h, k \in \mathbb{Z} \geq 0$, which counts the number of triangular facets per face in the corresponding triangulation. They correspond to an infinite series of polyhedra, which we call the *Caspar–Klug series*, because—according to Caspar–Klug theory—viruses follow the structure of these polyhedra in the organization of their capsids.

It had subsequently been observed by Wrigley (1969, 1970) that superpositions such as in figure 2*a* are too restrictive and that certain types of viruses rather follow the structures of the Goldberg polyhedra (Goldberg 1937), which correspond to hexagonal close packings on the surface of an icosahedron. These surface structures are encoded by Goldberg diagrams (see, for example, fig. 2 in Wrigley (1969)) and have been used by Wrigley (1969, 1970) to explain the surface structures of *Sericestis*, and respectively *Tipula*, iridescent virus.

Moreover, Coxeter (1971) realized the connection between the structure of virus macromolecules and geodesic domes. However, all of these approaches, including Coxeter's, missed the intriguing connection between non-crystallographic Coxeter groups and generalized lattices and their stunning potential in unravelling virus structure.

3. Viral tiling theory

Viral tiling theory differs from the Caspar–Klug theory by the introduction of more general types of surface lattices. Most importantly, it considers grids suitable for the modelling of the local fivefold protein clusters off the global symmetry axes that

have been excluded *a priori* before. Moreover, it relaxes the strict interpretation of the hexagonal lattice picture in terms of triangulations, thus permitting tessellations in terms of other shapes, such as rhombs. These again encode the locations of protein subunits, but lead to different predictions for the protein stoichiometry as discussed below. In this way, the viral tiling theory incorporates the Caspar–Klug theory, but enhances its spectrum of applications crucially.

(*a*) *A generalized lattice approach inspired by quasicrystals*

Since pentagons cannot tessellate the plane without gaps or overlaps, a straightforward generalization of the Caspar–Klug construction and its spin-offs is not possible. However, generalized lattices with the desired properties can be inferred via projection from higher-dimensional lattices with icosahedral symmetry. This phenomenon is known from the study of quasicrystals, alloys with non-crystallographic symmetry and long-range order discovered by Shechtman *et al.* (1984). Their discovery sparked the development of new mathematical techniques beyond the tools of crystallography (Senechal 1996), among them tilings inferred via projections from higher-dimensional lattices. A prominent planar example is the Penrose (1974) tiling, which can be obtained from a regular lattice in five dimensions via projection.

In a similar way, 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. However, it is more convenient to exploit the connection between the six-dimensional root lattice D_6 and the non-crystallographic Coxeter group H_3, as in Keef & Twarock (in press *a*), and to infer generalized lattices directly from H_3 and its affine extension as explained below.

The approach relies on the fact that the rotational symmetries of the icosahedral group are generated by the reflections r_j in H_3, which are reflections at planes perpendicular to a set of vectors α_j that form the so-called root system of the group and encode its structure geometrically (Humphreys 1992). Hence, $r_j(x) = x - ((2(x|\alpha_j))/(\alpha_j|\alpha_j))\alpha_j$, for any $x \in \mathbb{R}^3$. This is illustrated in figure 3. The root system of H_3 corresponds to the vectors that point from the centre of an icosidodecahedron (figure 3*a*) to its vertices. Two planes characterized by two of these vectors are shown explicitly in figure 3*b*, while for all other planes, only their intersections with the surface of the sphere are indicated. The intersections of the planes of reflection mark the locations of the axes of rotational symmetry of the icosahedral group. For example, the intersection of the two planes in the figure corresponds to an axis of fivefold symmetry. Similarly, one can also reconstruct the locations of all other symmetry axes.

In order to exploit this correspondence for the construction of generalized lattices, one has to realize that the vectors in the root system of H_3 are related to a particular choice of basis vectors (simple roots) of the icosahedrally symmetric lattice D_6 via a projection formalism similar to the one used in the study of quasicrystals (Senechal 1996). Therefore, it is possible to work with the root system of H_3 directly, provided that a method can be found to extend this root system in a way compatible with the overall icosahedral symmetry. We have shown that this can be achieved by an affine extension of the group (Keef & Twarock in press *a*). Via an adaptation of a mathematical formalism known from

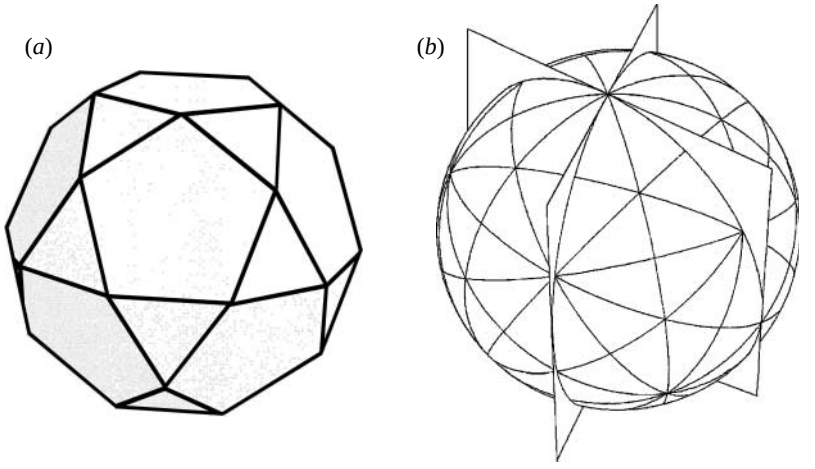

Figure 3. (*a*) The root polytope of H_3 and (*b*) examples of the reflection planes encoded by the root vectors.

Kac–Moody algebras (Patera & Twarock 2002), we have extended the root system of H_3 by a further vector. The method relies on the fact that the relations between the root vectors (simple roots) are encoded in a matrix, called the Cartan matrix, and that via an appropriate extension of this matrix, a further operation can be introduced. This operation describes an affine reflection, i.e. a reflection at a plane not centred at the origin, and acts as a translation T by a distinguished vector (the highest root) in the root system. Hence, the affine extended group is generated by three reflections, which correspond to reflections at planes through the origin, and are such that the products of any two of them correspond to a rotation in the icosahedral group, as well as T.

Via an iterated action of the generators of the extended group on the origin, point sets or generalized grids with icosahedral symmetry are constructed. If the action of the translation operator T were not restricted, $\mathbb{R}^3$ would be densely filled in this way and the point set would by construction correspond to a projection of the entire six-dimensional lattice D_6. However, if T acts only a finite number of times, say N, while the action of all other operations is not restricted, point sets $\mathcal{S}(N)$ are obtained that are finite subsets of generalized lattices known as cut-and-project quasicrystals or model sets. For increasing N, the point sets become larger and more dense, and they correspond to cut-and-project quasicrystals with increasingly larger acceptance windows. Examples have been worked out explicitly for the two-dimensional subgroup H_2 by Patera & Twarock (2002) and for H_3 by Twarock (2002), where applications to carbon cage structures have been considered.

(*b*) *A new series of polyhedra solving the structural puzzle for* Papovaviridae

The point sets $\mathcal{S}(N)$ form the starting point of the construction. They define nested point sets in three dimensions that are related to a six-dimensional lattice via the projection formalism mentioned before. By construction, they contain the vertex sets of the desired polyhedra or tilings. However, identifying these vertex sets within $\mathcal{S}(N)$ is not straightforward, because the polyhedra do not need to be

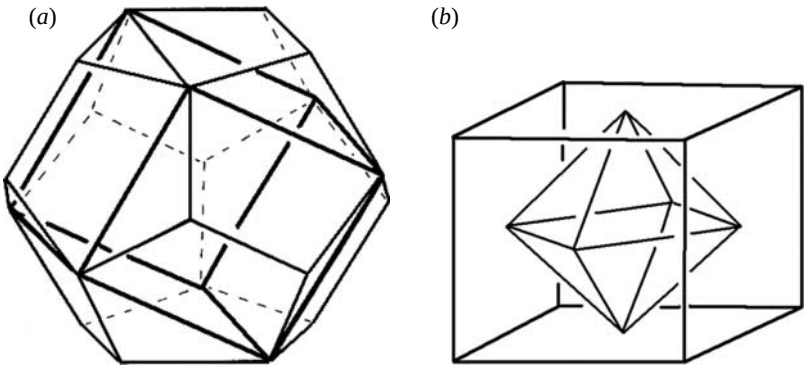

Figure 4. (*a*) A cube inscribed into a dodecahedron and (*b*) a tetrahedron inscribed into a cube.

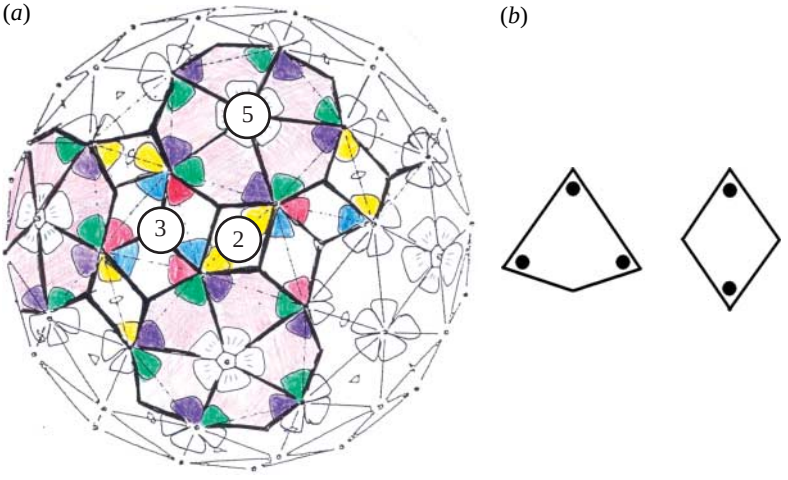

Figure 5. (*a*) The tiling representing the viral capsids of polyomavirus and simian virus 40 and (*b*) the corresponding tiles.

isometric and may also have octahedral or tetrahedral symmetry. The latter is due to the relations of the corresponding symmetry groups, which can be visualized geometrically as shown in figure 4.

For example, the polyhedra corresponding to all-pentamer capsids, i.e. capsids where all protein subunits cluster as five, have been worked out by Keef & Twarock (in press *a*). It has been shown that there is a (finite) series of three polyhedra of this type (apart from the icosahedron as a trivial case), which is called the *triacontahedral series*, because it starts with the triacontahedron. It contains the exceptional cases that complement the Caspar–Klug series discussed in §2. The large particle in the series corresponds to the tiling shown in figure 5*a*, where the locations of a twofold, a threefold and a fivefold symmetry axis are shown to facilitate visualization of the action of the icosahedral group. The tiles are rhombs and kites (figure 5*b*), and they encode the locations of the protein subunits similarly as in the Caspar–Klug theory as follows: all corners of the tiles meeting at five coordinated vertices in the tiling mark the locations of protein subunits. They are indicated schematically as dots in figure 5*b* and by colours in

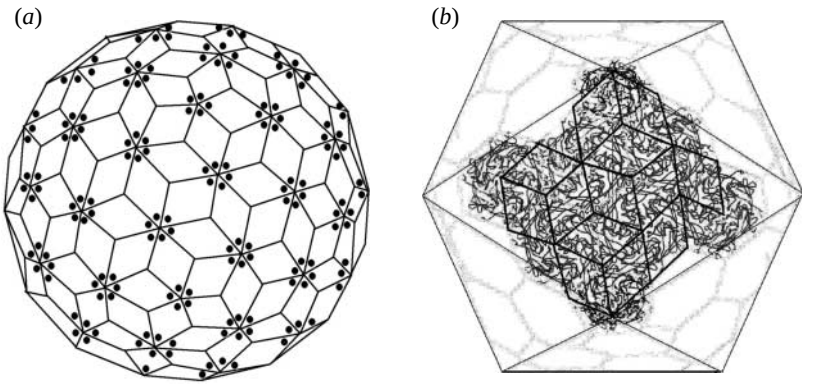

Figure 6. (*a*) The tiling representing the protein stoichiometry of bacteriophage HK97 and (*b*) in comparison with data from VIPER (Reddy *et al.* 2001). Both figures are adapted from Twarock & Hendrix (2006).

figure 5*a*, where sites of the same colour refer to the positions that are equivalent under the action of the icosahedral group. The surface structure encoded by this tiling provides the solution to the structural puzzle pointed out by Liddington *et al.* (1991).

(*c*) *Viral tiling theory for the case of hexagonal lattices*

Viral tiling theory includes the case of hexagonal surface lattices, so that the structures in the Caspar–Klug theory are also contained in this framework. However, it does not limit its scope to triangulations, but permits more general tessellations such as rhomb tilings (Twarock 2005*a*). An example is shown in figure 6; it encodes the surface structure of bacteriophage HK97, a micro-organism similar to a virus that infects bacteria. Experimental data on the locations of the proteins in the bacteriophage HK97 capsid are shown in figure 6*b*, together with a patch of the tiling to demonstrate how the information encoded in the tiling has to be translated into the real-life setting.

The distinction between a triangulation and a rhomb tiling is crucial. Both predict the locations of the capsomeres at the same places, but the prediction for their relative orientations is different. This is important, because it affects the ability of forming crosslinking structures, as will be discussed in more detail in §4.

(*d*) *Predictions of viral tiling theory*

By construction, viral tiling theory not only predicts the locations of the protein subunits as in the Caspar–Klug theory, but also specifies the locations of the inter-subunit bonds between proteins in different capsomeres. In particular, as discussed in §3*b*, the locations of the proteins are marked as dots on the tiles. In addition to that, tiles with two or three protein subunits represent an interaction between these two proteins (called dimer interaction) or three proteins (called trimer interaction), respectively, on that tile (Twarock 2005*a*).

This is demonstrated in figure 7 for the case of the tiling in figure 5, which corresponds, for example, to the polyomavirus and simian virus 40 in the family of *Papovaviridae*. The predicted locations of the C-terminal arm extensions of the

proteins, that are responsible for the dimer and trimer interactions, are visualized schematically as spiral arms. There are two types of dimer bonds represented by 'yellow–yellow' and 'blue–red' rhombs, respectively, and one type of trimer bond in the kite tiles. Moreover, there are two different types of local bonding environments around the pentamers in the capsid: one type around the pentamers located on the fivefold axes of icosahedral symmetry (represented by five kite tiles) and another one around all other 60 pentamers (represented by two kites and one rhomb tile). This distinction is important for the construction of assembly models as discussed in §4. The predictions agree well with the experimental results of Modis *et al.* (2002).

Another crucial feature of this approach is the fact that it predicts the relative radii of different particles in the triacontahedral series. Hence, there is only one scaling parameter that maps the overall mathematical set-up collectively onto its biological counterpart. For example, a comparison with experimental data (Kiselev & Klug 1969; Salunke *et al.* 1989) shows an excellent agreement with the radii predicted for *Papovaviridae* by our theory (Keef & Twarock in press *a*).

4. Recent applications of viral tiling theory

Owing to the fact that tiles have a biological interpretation in terms of interactions between the proteins they encode, the viral tiling theory lends itself to various applications. Recent work along these lines is discussed below in perspective with other works in these areas.

(*a*) *Assembly models*

Manipulating the assembly of viral capsids is one way of interfering with the viral replication cycle, hence a possible avenue for anti-viral drug design. The first model for the self-assembly of a small plant virus was pioneered by Zlotnick (1994), exploring the assembly of a dodecagonal shape by a cascade of single-order reactions. It has since been extended to more involved scenarios (Endres & Zlotnick 2002), including a recent study of the energy landscape underlying assembly (Endres *et al.* 2005) similar to the energy landscapes considered by Wales (2005). Moreover, Zandi *et al.* (2006) have generalized Zlotnick's approach and shown that there must be thermodynamically favourable intermediate states that produce a kinetic bottleneck to capsid formation. Related approaches include molecular dynamics studies of viral capsid assembly (Rapaport *et al.* 2004) and a molecular dynamics-like formalism that is implemented in connection with a 'local rules' mechanism, which regulates capsid assembly (Berger *et al.* 1994; Schwartz *et al.* 1998).

A characteristic feature of these models is the fact that the bonding structures of all building blocks are treated on an equal footing. While this is justified for many viruses, it is an inappropriate simplification for important families, such as the *Papovaviridae*. For example, the particles encoded by the tiling in figure 5*a* are composed of pentamers that attain two different conformations in the capsid: the 12 pentamers located on the fivefold axes of the capsid are surrounded by trimer interactions, while the 60 other pentamers are surrounded by a combination of dimer and trimer interactions. Viral tiling theory provides a mathematical tool to model these different bonding environments via the

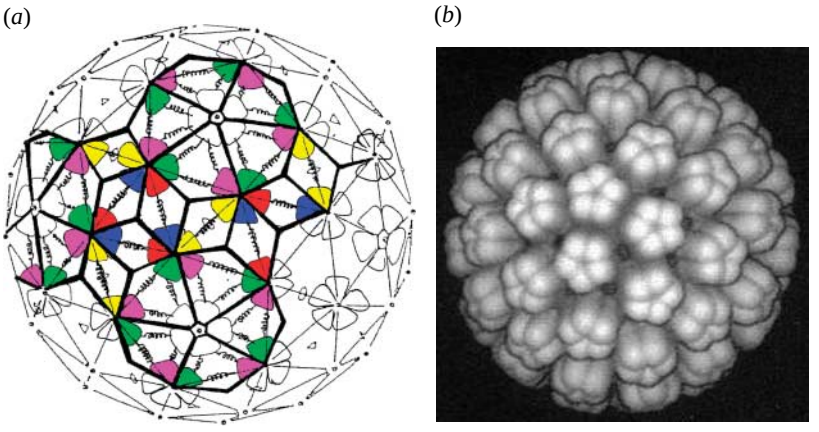

Figure 7. (*a*) Location of the dimer and trimer interactions in the capsids corresponding to the tiling in figure 5. (*b*) A computer simulation of the three-dimensional surface morphology of polyomavirus, adapted from Salunke *et al.* (1986), shown in order to demonstrate how the tilings translate into three-dimensional surface structures.

corresponding vertex stars, i.e. configurations of tiles around vertices, in the tiling. For example, the vertex stars of the tiling corresponding to the large particles in the family of *Papovaviridae*, which are shown in figure 8, have been used as the building blocks for the construction of assembly models (Keef *et al.* 2005). In particular, combinatorial structures called assembly graphs have been introduced, which encode the succession of the energetically favourable assembly intermediates under the assumption that assembly takes place by the addition of a single building block per iteration step. As discussed in §3*d*, there occur three different types of interactions in the capsids of these viruses, which are modelled by three association constants. The dependence of the assembly scenario on the values of these constants has been analysed by Keef *et al.* (2005) and the concentrations of the statistically dominant assembly intermediates have been computed. Moreover, the most probable assembly pathways have been determined based on a master equation approach (Keef *et al.* 2006*a*), and it has been shown that the assembly intermediates on the dominant pathways have characteristic geometrical properties that may potentially be exploited in the framework of anti-viral drug design.

(*b*) *Tubular malformations*

In vitro experiments on the self-assembly of viral capsid proteins have demonstrated that different types of viral particles may occur depending on boundary conditions, such as pH value and salt concentration in solution. For example, studies on polyomavirus have provided evidence for the occurrence of tubular malformations besides the spherical capsids (Salunke *et al.* 1989). Since the tiles in viral tiling theory have a biological interpretation as dimer and trimer interactions between protein subunits, they lend themselves as the building blocks for the modelling of these tubular structures. Indeed, based on the assumption that the same dimer and trimer interactions are responsible for the formation of both the spherical and the tubular particles, and that they can be represented by

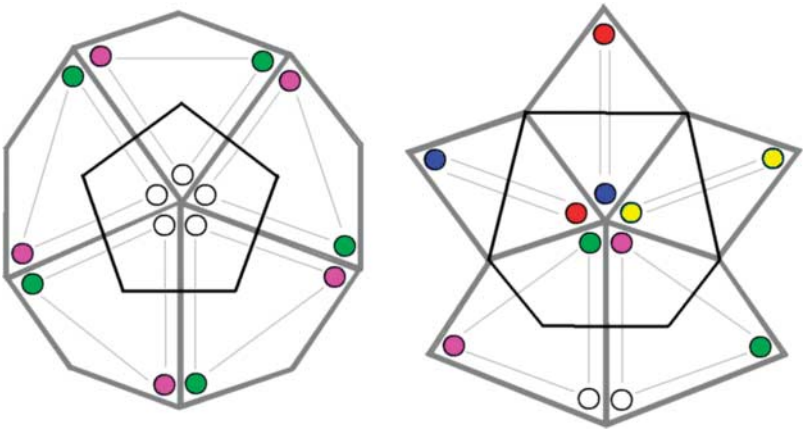

Figure 8. The building blocks for capsid assembly of the large particles in the family of *Papovaviridae*.

the same shapes of tiles which may assemble into either a compact or a planar surface, the classification of possible tubular malformations translates into the mathematical problem of classifying tubular lattices with a suitably chosen set of properties. For *Papovaviridae*, these lattices are given in terms of the rhomb and kite tiles in figure 5*b* and have been classified by Twarock (2005*b*). Moreover, some tubular particles exhibit end caps, which—from a mathematical point of view—imposes further constrains on the structures of their surface lattices. For the family of *Papovaviridae*, these have been classified by Keef *et al.* (2006*b*). It has been shown that the lattices obtained from first mathematical principles in this way describe more stable structures than the lattices previously suggested to interpret experimental data (Kiselev & Klug 1969).

(*c*) *Crosslinking structures*

Some types of viruses and bacteriophages have covalent bonds in addition to the dimer and trimer interactions that are encoded by the tiles in viral tiling theory. For example, in the bacteriophage HK97 capsid discussed in §3*c*, they have been shown to cover the entire capsid in a chainmail organization, as illustrated in figure 9.

Owing to the fact that such crosslinking structures affect an essential stabilization of the capsid, as demonstrated by Ross *et al.* (2005), it is important to develop a theoretical tool that allows the assessment of whether crosslinking is possible for a given type of virus. Since viral tiling theory encodes the structures of viral capsids in terms of tilings, it lends itself as a basis for such a theory. Indeed, Twarock & Hendrix (2006) have shown that probing for the possibility of crosslinking in a virus can be translated into the combinatorial problem of finding so-called higher-level tilings that are compatible with the tiling that describes its surface structure. For example, the higher-level tiling for the rhomb tiling representing bacteriophage HK97 in figure 6 is shown superimposed on the original tiling in figure 9*b*; it is a triangulation in terms of the tiles in figure 9*c*. Based on this approach, Keef (in press) derives a list of all types of crosslinking structures that are compatible with the surface structures

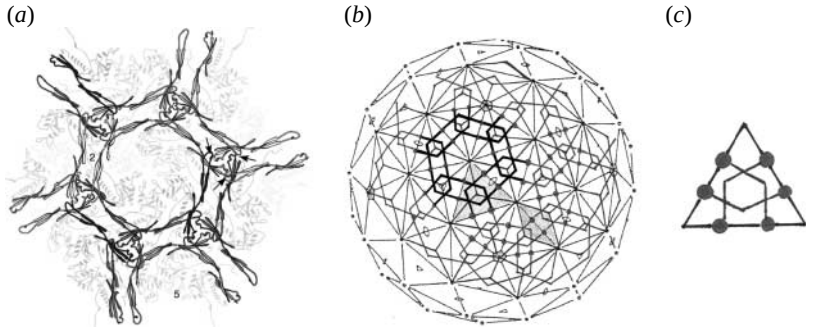

Figure 9. A cartoon, adapted from Helgstrand *et al.* (2003), showing the geometry of crosslinking in bacteriophage HK97 capsids. (*a*) The diagram shows one complete hexamer of protein subunits and portions of the surrounding capsomeres. These subunits crosslink to surrounding capsomeres, hence forming the chainmail organization encoded by the higher-level tiling (*b*) with the tiles in (*c*). The figures are adapted from Twarock & Hendrix (2006).

encoded in viral tiling theory. Apart from providing a tool for the analysis of cryo-electron microscopy data, these results also have potential applications in the design of capsids as containers for drug delivery, because it provides an overview of the crosslinking structures that may be engineered in these capsids in order to render them particularly stable.

5. Outlook: the journey has just begun...

The applications discussed in §4 are just the starting point. Viral tiling theory and, in particular, the generalized grids that have been used to construct the tilings provide very powerful mathematical tools with a plethora of possible applications in virology. Some of them have already been pointed out in §1. More details on some of the possible next steps are given below, and the potential for medical applications is discussed.

(*a*) *Assembly models*

The applications to assembly models discussed in §4*a* are still in their infancy. While we have good control over the assembly of a single species of viral particles at present, it is now important to extend the models to allow for the simultaneous assembly of different particles and tubular structures. This is essential to study the possibilities to misdirect capsid assembly and, hence, to find methods of interference with the viral assembly process that lead to medical applications. A starting point for this will be to combine our results on the classification of the tubular malformations with our approach based on assembly graphs for a single species.

In order to accurately reflect the situation *in vivo*, these models moreover need to incorporate the observed dependence on experimental conditions, such as pH value and salt concentration. This requires an additional input in the models that links the association constants of the inter-subunit bonds with parameters that capture these boundary conditions. A further challenge is to understand scaffold-mediated assembly: some larger viruses use a protein scaffold to guide the

assembly of their capsids (Prevelige *et al.* 1993). Inhibiting capsid assembly by inference with the scaffold may be another strategy for anti-viral drug design targeted to these types of viruses.

Another fascinating challenge is to better understand the assembly of RNA viruses. While most models so far concern the assembly of DNA viruses, where the capsid can be self-assembled from the capsid proteins alone and the viral genome is packaged inside the capsid after the completion of the assembly process, it is known that the coat proteins of RNA viruses interact with the genome during assembly. For example, enticing results on the assembly of an RNA virus have recently been obtained from the Stockley Lab at the University of Leeds (P. Stockley 2006, personal communication), and it is a challenge to come up with new models adapted to this situation. Modelling the role of RNA during capsid assembly could be achieved, for example, by imposing further combinatorial constrains in our assembly models for DNA viruses that reflect the effects of the interactions between RNA and the dimeric building blocks.

Finally, it is important to incorporate the shapes of the capsid proteins into the assembly models. Hagan & Chandler (2006) mimic the protein shapes by representing capsomeres with internal bond vectors in their models. Another alternative is to use our approach to obtain a three-dimensional representation of viral capsids, in which individual capsid proteins are compartmentalized into shapes that are compatible with the overall lattice structure. A proof of principle study for bacteriophage MS2 (Keef & Twarock in press *b*) is very promising.

(*b*) *Physical properties of viral capsids*

Now that the mathematical set-up of viral tiling theory has been worked out completely, it can be used to model various physical properties of viruses. For example, the generalized grids obtained via the affine extensions of non-crystallographic Coxeter groups allow us to model phonons and phason flips in viral capsids along similar lines as in the theory of quasicrystals (Janot 1992) based on the underlying projection picture.

The mechanical properties of icosahedral viral capsids have been studied by Zadi *et al.* (2004) and Zandi & Requera (2005), and they have shown that icosahedral symmetry provides a built-in genome release mechanism through inhomogeneous stress distribution. Moreover, the mechanical properties have been investigated with elastic network normal mode analysis by Tama & Brooks (2005), and elastic network theory has been used to investigate the global distortions of biological macromolecules by Tama *et al.* (2002). Moreover, buckling transitions have been considered via a model based on the nonlinear physics of thin elastic shells by Lidmar *et al.* (2003). Further details on these approaches can be found in a special issue devoted to mathematical virology (McLeish *et al.* 2005). It is conceivable that our generalized grids may also prove useful in this context.

(*c*) *Genome packaging*

Another challenging area of research in virology concerns the packaging of the viral genome. There are several competing models concerning the packaging structure (see Mullaney & Black (1998) and references within), but it has still not been possible to decide conclusively which one of them is correct. Again, mathematical models based on symmetry as a guiding principle should be able to

provide an answer to this open problem. Work in progress considers applications of the generalized grids in Keef & Twarock (in press *a*) to the modelling of the shell structures observed, for example, by Jiang *et al.* (2006).

Experiments have shown that there are RNA viruses that package part of their genome in a way that mirrors the icosahedral symmetry of the capsids (e.g. Rudnick & Bruinsma 2005). Since our generalized grids define a series of nested polyhedra that are compatible with the overall icosahedral symmetry by construction, they lend themselves also to a structural description of these packaging structures.

(*d*) *A road towards anti-viral drug design?*

Apart from their purely scientific interest, these approaches have the potential to lead to enticing medical applications. For example, as soon as the assembly process is better understood along the lines outlined above, it will be possible to study the methods of misdirecting capsid assembly, for example, by shifting the bias from the infectious spherical viral particles to the non-infectious tubular ones. Moreover, the assembly models may be used to investigate the possibility of inhibiting assembly via an anti-viral drug. For the case of herpes virus, this has already been done by Zlotnick *et al.* (2002), and with the new approach, similar studies are also conceivable for other viruses in the future.

Furthermore, viral capsids are currently under investigations as containers for drug delivery. For example, the capsid of simian virus 40, which is a member of the family of *Papovaviridae* that we are actively studying based on viral tiling theory (figure 5), is currently under investigation in the framework of gene therapy in the group of A. Oppenheim at the Hebrew University in Jerusalem.

Financial support via an EPSRC Advanced Research Fellowship is gratefully acknowledged. Some of the work has been done in collaboration with my PhD student Tom Keef, who is funded by the EPSRC LSI programme.

References

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

Berger, B., Shor, P. W., Tucker-Kellogg, L. & King, J. 1994 Local rule-based theory of virus shell assembly. *Proc. Natl Acad. Sci. USA* **91**, 7732–7736. (doi:10.1073/pnas.91.16.7732)

Casjens, S. 1985 *Virus structure and assembly*. Boston, MA: Jones and Bartlett.

Caspar, D. L. D. & Klug, A. 1962 Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

Coxeter, H. S. M. 1971 Virus macromolecules and geodesic domes. In *A spectrum of mathematics* (ed. J. C. Butcher), p. 238. Auckland, New Zealand: Auckland University Press.

Crick, F. H. C. & Watson, J. D. 1956 The structure of small viruses. *Nature* **177**, 473–475. (doi:10.1038/177473a0)

Endres, D. & Zlotnick, A. 2002 Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**, 1217–1230.

Endres, D., Miyahara, M., Moisant, P. & Zlotnick, A. 2005 A reaction landscape for identifying important intermediates in the self-assembly of virus capsids and other polyhedra. *Protein Sci.* **14**, 1518–1525. (doi:10.1110/ps.041314405)

Goldberg, M. 1937 A class of multi-symmetric polyhedra. *Tohoku Math. J.* **43**, 104–108.

Hagan, M. F. & Chandler, D. 2006 Dynamic pathways for viral capsid assembly. *Biophys. J.* **91**, 42–54. (doi:10.1529/biophysj.105.076851)

Helgstrand, Ch. *et al.* 2003 The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J. Mol. Biol.* **334**, 885–899. (doi:10.1016/j.jmb.2003.09.035)

Humphreys, J. E. 1992 *Reflection groups and coxeter groups, Cambridge studies in advanced mathematics*, vol. 29. Cambridge, UK: Cambridge University Press.

Janner, A. 2006 Crystallographic structural organization of human rhinovirus serotype 16,14,3,2 and 1A. *Acta Cryst. A* **62**, 270–286. (doi:10.1107/S010876730601748X)

Janot, C. 1992 *Quasicrystals: a primer*. Oxford, UK: Clarendon Press.

Jiang, W. *et al.* 2006 Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**, 612–616. (doi:10.1038/nature04487)

Keef, T. In press. A mathematical approach to cross-linked structures in viral capsids: predicting the architecture of novel containers for drug delivery. In *12th International Meeting on DNA Computing, Seoul, Korea, 5 June 2006, LNCS*. Berlin, Germany: Springer.

Keef, T. & Twarock, R. In press *a*. A novel series of polyhedra as blueprints for viral capsids in the family of *Papovaviridae*, q-bio.BM/0512047. *Bull. Math. Biol.*

Keef, T. & Twarock, R. In press *b*. The three-dimensional architecture of simple viruses is encoded by a novel symmetry principle based on self-similarity. *Nature*.

Keef, T., Taormina, A. & Twarock, R. 2005 Assembly models for *Papovaviridae* based on tiling theory. *J. Phys. Biol.* **2**, 175–188. (doi:10.1088/1478-3975/2/3/005)

Keef, T., Micheletti, C. & Twarock, R. 2006*a* Master equation approach to the assembly of viral capsids. *J. Theor. Biol.* **242**, 713–721. (doi:10.1016/j.jtbi.2006.04.023)

Keef, T., Taormina, A. & Twarock, R. 2006*b* Classification of capped tubular viral particles in the family of *Papovaviridae*. *J. Phys.: Condens. Matter* **18**, S375–S387. (doi:1088/0953-8984/18/14/S18)

Kiselev, N. A. & Klug, A. 1969 The structure of viruses of the papilloma–polyoma type. *J. Mol. Biol.* **40**, 155–171. (doi:10.1016/0022-2836(69)90465-3)

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. 1991 Structure of simian virus 40 at 3.8-Å resolution. *Nature* **354**, 278–284. (doi:10.1038/354278a0)

Lidmar, J., Mirny, L. & Nelson, D. R. 2003 Virus shapes and buckling transitions in spherical shells. *Phys. Rev. E* **68**, 051910–051919. (doi:10.1103/PhysRevE.68.051910)

McLeish, T., Stockley, P. & Twarock, R. (eds) 2005 Mathematical virology. *J. Theor. Med.* **6**, 67–141.

Modis, Y., Trus, B. L. & Harrison, S. C. 2002 Atomic model of the papillomavirus capsid. *EMBO J.* **21**, 4754–4762. (doi:10.1093/emboj/cdf494)

Mullaney, J. M. & Black, L. W. 1998 Activity of foreign proteins targeted with the bacteriophage T4 head and prohead: implications for packaged DNA structure. *J. Mol. Biol.* **283**, 913–929. (doi:10.1006/jmbi.1998.2126)

Patera, J. & Twarock, R. 2002 Affine extensions of noncrystallographic Coxeter groups and quasicrystals. *J. Phys. A* **35**, 1551–1574. (doi:10.1088/0305-4470/35/7/306)

Penrose, R. 1974 Pentaplexity. *Bull. Inst. Math. Appl.* **10**, 266–271.

Prevelige, P. E., Thomas, D. & King, J. 1993 Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**, 824–835.

Rapaport, D. C. *et al.* 2004 Self-assembly of polyhedral shells: a molecular dynamics study. *Phys. Rev. E* **70**, 051905–051917. (doi:10.1103/PhysRevE.70.051905)

Rayment, I., Baker, T. S., Caspar, D. L. & Murakami, W. T. 1982 Polyoma virus capsid structure at 22.5 Å resolution. *Nature* **295**, 110–115. (doi:10.1038/295110a0)

Reddy, V. S., Natarajan, Padmaja, Okerberg, Brian, Li, Kevin, Damodaran, K. V., Morton, T. Ryan & Johnson, John E. 2001 Virus particle explorer (viper), a website for virus capsid structures and their computational analyses. *J. Virol.* **75**, 11 943–11 947. (doi:10.1128/JVI.75.24.11943-11947.2001)

Ross, P. D., Cheng, N., Conway, J. F., Firek, B. A., Hendrix, R. W., Duda, R. L. & Steven, A. C. 2005 Crosslinking renders bacteriophage HK97 capsid maturation irreversible and effects an essential stabilization. *EMBO J.* **24**, 1352–1363. (doi:10.1038/sj.emboj.7600613)

Rudnick, J. & Bruinsma, R. 2005 Icosahedral packing of RNA viral genomes. *Phys. Rev.Lett.* **94**, 038101. (doi:10.1103/PhysRevLett.94.038101)

Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. 1986 Self-assembly of purified polyomavirus capsid protein VP1. *Cell* **46**, 895–904. (doi:10.1016/0092-8674(86)90071-1)

Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. 1989 Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* **56**, 887–900.

Schwartz, R., Shor, P. W., Prevelige Jr, P. E. & Berger, B. 1998 Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* **75**, 2626–2636.

Senechal, M. 1996 *Quasicrystals and geometry*. Cambridge, UK: Cambridge University Press.

Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. 1984 Metallic phase with long-range order and no translational symmetry. *Phys. Rev. Lett.* **53**, 1951–1953. (doi:10.1103/PhysRevLett.53.1951)

Tama, F. & Brooks III, C. L. 2005 Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.* **345**, 299–314. (doi:10.1016/j.jmb.2004.10.054)

Tama, F., Wriggers, W. & Brooks III, C. L. 2002 Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J. Mol. Biol.* **321**, 297–305. (doi:10.1016/S0022-2836(02)00627-7)

Twarock, R. 2002 New group structures for carbon onions and carbon nanotubes via affine extensions of noncrystallographic Coxeter groups. *Phys. Lett. A* **300**, 437–444. (doi:10.1016/S0375-9601(02)00856-3)

Twarock, R. 2004 A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* **226**, 477–482. (doi:10.1016/j.jtbi.2003.10.006)

Twarock, R. 2005*a* The architecture of viral capsids based on tiling theory. *J. Theor. Med.* **6**, 87–90. (doi:10.1080/10273660500148770)

Twarock, R. 2005*b* Mathematical models for tubular structures in the family of *Papovaviridae*. *Bull. Math. Biol.* **67**, 973–987. (doi:10.1016/j.bulm.2004.11.005)

Twarock, R. & Hendrix, R. 2006 Crosslinking in viral capsids via tiling theory. *J. Theor. Biol.* **240**, 419–424. (doi:10.1016/j.jtbi.2005.10.001)

Wales, D. J. 2005 The energy landscape as a unifying theme in molecular science. *Phil. Trans. R. Soc. A* **363**, 357–377. (doi:10.1098/rsta.2004.1497)

Weiss, P. 2005 Amor-plated puzzle. Deciphering the code of viral geometry. *Sci. News* **168**, 152–154.

Wrigley, N. G. 1969 An electron microscope study of the structure of *Sericesthis* iridescent virus. *J. Gen. Virol.* **5**, 123–134.

Wrigley, N. G. 1970 An electron microscope study of the structure of *Tipula* iridescent virus. *J. Gen. Virol.* **5**, 169–173.

Zandi, R. & Requera, D. 2005 Mechanical properties of viral capsids. *Phys. Rev. E* **72**, 021917. (doi:10.1103/PhysRevE.72.021917)

Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. 2004 Origin of icosahedral symmetry in viruses. *Proc. Natl Acad. Sci. USA* **101**, 15 556–15 560. (doi:10.1073/pnas.0405844101)

Zandi, R., van der Schoot, P., Reguera, D., Kegel, W. & Reiss, H. 2006 Classical nucleation theory of virus capsids. *Biophys. J.* **90**, 1939–1948. (doi:10.1529/biophysj.105.072975)

Zlotnick, A. 1994 To build a virus capsid: an equilibrium model of the self-assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59–67. (doi:10.1006/jmbi.1994.1473)

Zlotnick, A., Ceres, P., Singh, S & Johnson, J. M. 2002 A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *J. Virol.* **76**, 4848–4854. (doi:10.1128/JVI.76.10.4848-4854.2002)

AUTHOR PROFILE

Reidun Twarock

Dr. Reidun Twarock born in Cologne, Germany, Reidun Twarock studied mathematics and physics at the University of Cologne as a fellow of the Studienstiftung des Deutschen Volkes before moving to the University of Bath where she completed a Master of Science degree (with distinction) in nonlinear mathematics in 1993. She returned to Germany to the Technical University of Clausthal where she obtained a Ph.D. (with distinction) in mathematical physics in 1997. Reidun was a Dorothea Erxleben Fellow at the Arnold Sommerfeld Institute in Clausthal until 2000, when she moved to the University of York with a Marie Curie Fellowship. In 2001, she joined the Department of Mathematics at City University in London as a lecturer, and returned to York in 2005 as a reader jointly appointed by the Departments of Mathematics and Biology. Aged 36, Reidun is currently an EPSRC Advanced Research Fellow working on mathematical models for the structure and assembly of viruses, and she is one of the main organizers of a workshop series on mathematical virology, which started at the Centre for Mathematical Sciences in Oxford in June 2004, and will take place next at the International Centre for Mathematical Sciences in Edinburgh in August 2007. Scientific interests include algebraic and group theoretical methods in mathematical physics and group theory inspired models in virology. Reidun loves playing her flute and Celtic harp and enjoys painting.