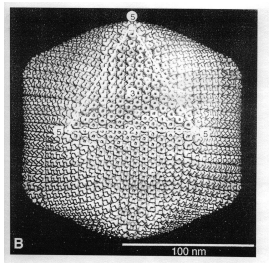
MATHEMATICAL MODELS FOR THE STRUCTURE AND SELF-ASSEMBLY OF VIRUSES

Reidun Twarock

Résumé. **—** Viruses have a protein shell, called the viral capsid, that encapsulates and hence provides protection for the viral genome. The distribution of the proteins in the capsids is highly structured and follows an organisational principle that can be described based on group theory and tiling theory. It provides a basis for mathematical models that address the self-assembly of the capsids from their capsid proteins, and may ultimately be used to assist the design of anti-viral therapeutics.

1. Introduction

Viral capsids are cage structures, formed from proteins, that are used by viruses to protect their genetic material. Most spherical viruses organise their protein subunits in clusters of 3, 5 or 6 proteins, that are called trimers, pentamers and hexamers, respectively, and are distributed with icosahedral symmetry in the capsid. An example is given in Fig. 1(a), with a magnified view in (b) showing a subset of the hexamers.

(a) 

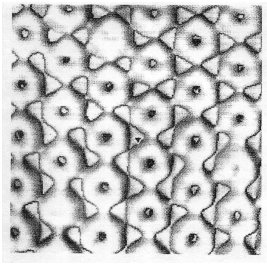
(b) 

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

In a landmark paper Caspar and Klug have established a theory that predicts the surface structures of viruses in terms of a family of polyhedra [**1**] that encode the locations and relative orientations of the protein clusters. It is fundamental in virology and has a broad spectrum of applications, ranging from image analysis of experimental data to the construction of models for the self-assembly of viral capsids. Despite its huge success, experimental results have provided evidence for the fact that this theory is incomplete, and in particular cannot account for the structure of viruses in the family of Papovaviridae, which are of special interest for the public health sector because they contain cancer-causing viruses. We have developed a theory based on group theory and tiling theory that closes this gap [**2, 3**]. It leads to a new series of polyhedra, the triacontahedral series [**4**], that corresponds to the particles observed during the self-assembly of the major capsid proteins of viruses in the family of Papovaviridae. Among others, the new theory allows to classify the malformations that may occur during self-assembly (e.g. [**5**]), and it has opened up various areas of application, most importantly the construction of models for the self-assembly of viral capsids.

2. The construction principle

Since all protein clusters in the capsids of Papovaviridae are composed of five individual protein subunits, including in particular also those not located at the 5-fold axes of icosahedral symmetry, the surface structures of these viruses cannot be modelled via hexagonal surface lattices and are hence a priori excluded by Caspar-Klug Theory. A straightforward generalisation of the Caspar-Klug construction to this case is not possible because there are no planar lattices composed only of pentagons which could be used instead of the hexagonal ones. However, appropriate surface lattices with the desired symmetry properties can be induced via projection from lattices in a higher-dimensional space. In particular, by exploiting the concept of symmetry to the full we make use of generalised grids that are determined via the affinisation of the non-crystallographic Coxeter group H_3 using a method inspired by the projection formalism (see e.g. [**6**]) known from the theory of quasicrystals [**7**] and Penrose tilings [**8**].

By construction, our method leads to finite three-dimensional nested point sets that are subsets of the vertex sets of such generalised lattices and by construction contain the vertices of polyhedra (or tessellations called tilings) that encode the surface structures of the viral capsids. For example, in order to classify the surface structures of Papovaviridae, one has to identify all (not necessarily isometric) polyhedra with vertices in this set that are such that all five-coordinated vertices are uniformly distributed. Each five-coordinated vertex then specifies the location and orientation

of a pentamer as illustrated in the example in Fig. 2. The tessellation in this figure is

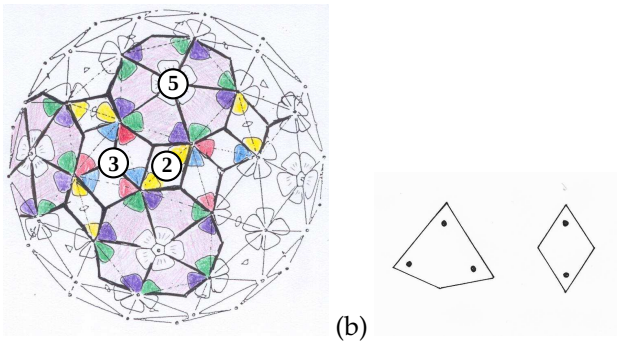


FIGURE 2. (a) The tiling representing the viral capsid of SV40, and (b) the corresponding tiles.

given in terms of two types of shapes, which are called tiles according to terminology in tiling theory. Since we are considering capsids formed from identical proteins their locations are represented by opening angles of equal magnitude on the tile set. There are thus precisely five proteins located around the 72 five-coordinated vertices (of which 12 are located on the five-fold axes of icosahedral symmetry indicated by a 5 in the figure), each specifying the locations of the proteins in a pentamer. A special feature of the theory is the fact that tiles are not only idealised mathematical objects, but have a biological interpretation in terms of interactions between protein subunits : each rhomb tile represents an interaction between the *two* protein subunits represented by that tile (called *dimer interaction*), while each kite tile represents an interaction between the *three* proteins it encodes (called *trimer interaction*).

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. Moreover, its predictive power and scope of applications is significantly enhanced with respected to Caspar-Klug Theory because it predicts besides the locations of the proteins also the locations of the intersubunit bonds between them.

3. Applications

Our theory for the structural description of viruses forms a basis for the construction of assembly models, because it specifies the locations of the bonds between the protein subunits that form an essential input in the modelling of the assembly process. In particular, we use this information to derive graphs that encode the structure and order of succession of the intermediate species (partly assembled capsids) that occur during self-assembly of the capsid proteins. These graphs are combinatorial

objects that are used to derive quantities of interest such as the concentrations of the assembly intermediates, and they hence characterize the assembly process [**9**]. Based on them, we have developed a method to quantify the change in assembly behaviour in dependence on changes in the association constants. Moreover, we have determined the dominant pathways of assembly via a master equation approach, and have analysed the geometric characteristics of the intermediate species represented by them, which has implications on potential strategies of interference with the assembly process for medical purposes [**10**].

With the same mathematical formalism we have furthermore characterised the additional covalent bonds in the capsids that are responsible for crosslinking, that is for the occurrence of networks of bonds that are organised in a chainmail construction and provide particular stability to the viral capsids [**11**]. We have shown that our approach can be used to classify these crosslinking structures, and that it provides a theoretical tool to probe whether crosslinking is possible for a given type of virus [**12**].

Many single-stranded (ss)RNA viruses organise a significant part of their genome in a dodecahedral cage as a RNA duplex structure that mirrors the symmetry of the capsid. We have further developed a model by Bruinsma and Rudnick for the structural organisation of the RNA in pariacoto virus based on results from graph theory and DNA network engineering [**13**]. We show that it is a representative of a whole family of cage structures that abide to the same construction principle, and derive the energetically optimal configurations.

Références

[1] Caspar, D.L.D & Klug, A. (1962) Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symp. Quant. Biol.* 27, pp. 1.

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

[3] R. Twarock (2005) The architecture of viral capsids based on tiling theory, *J. Theor. Medicine* **6**, pp. 87.

[4] Keef, T. and Twarock, R. (2005) A novel family of polyhedra as blueprints for viral capsids in the family of Papovaviridae, submitted to *J. Math. Biol.*.

[5] Twarock, R. (2005) Mathematical models for tubular structures in the family of Papovaviridae, *Bull. Math. Biol.* **67**, pp. 973.

[6] Senechal, M. 1996 *Quasicrystals and Geometry*, Cam. Univ. Press.

[7] Shechtman, D. *et al.* 1984 Metallic phase with long-range order and no translational symmetry. *Phys. Rev. Lett.* 53, 1951-1953.

[8] Penrose, R. 1974 Pentaplexity. *Bulletin of the Institute for Mathematics and Applications* 10, 266-271.

[9] Keef, T., Taormina, A. and Twarock, R. (2005) Assembly Models for Papovaviridae based on Tiling Theory, *Phys. Biol.* **2**, pp.175.

[10] Keef, T., Micheletti, C. and Twarock, R. (2005) Master equation approach to the assembly of viral capsids, to appear in *J. Theor. Biol*.

[11] Helgstrand, Ch. *et al.* 2003 The Refined Structure of a Protein Catenane : The HK97 Bacteriophage Capsid at 3.44Å Resolution. *J. Mol. Biol.* 334, 885-899.

[12] Twarock, R. and Hendrix, R. (2005) Crosslinking in Viral Capsids via Tiling Theory, *J. Theor. Biol.*, in press

[13] Jonoska, N. and Twarock, R. (2006) A Note on Genome Organisation in RNA Viruses with Icosahedral Symmetry, in preparation.

Reidun Twarock

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

E-mail : rt507@york.ac.uk