Models of viral capsid symmetry as a driver of discovery in virology and nanotechnology

Pierre-Philippe Dechant
(York St John University, U.K.)

Reidun Twarock
(University of York, U.K.)

Viruses are prominent examples of symmetry in biology. A better understanding of symmetry and symmetry breaking in virus structure via mathematical modelling opens up novel perspectives on how viruses form, evolve and infect their hosts. In particular, mathematical models of viral symmetry pave the way to novel forms of antiviral therapy and the exploitation of viral protein containers in bio-nanotechnology.

Symmetry is ubiquitous in virology

Viruses are evolutionarily highly optimized molecular machines. Understanding their inner workings sheds light on fundamental questions in molecular biology, biomedicine and nanotechnology. Viruses store their genetic material inside protective protein containers called viral capsids. Viral genomes consist of either DNA or RNA, which can both be single or double-stranded, and some viruses reverse transcribe between the two. Single-stranded viruses tend to have shorter genomes due to the relative flexibility of the nucleic acid molecule, and they often package their genomes into the capsid during its assembly in a co-assembly process. By contrast, genetically more complex viruses tend to store their genetic message in the form of the more stable dsDNA. Whilst some of these viruses have much more complex life cycles and less symmetric structures (e.g. poxviruses), a surprisingly large fraction of them still exhibit icosahedral and helical design principles (e.g. the tailed phages or many of the recently discovered giant viruses). In these viruses, the nucleic acid is often packaged into a preformed capsid using an energy-driven molecular motor.

In the vast majority of viruses, these capsid containers exhibit icosahedral symmetry, meaning that they look like tiny footballs at the nanoscale. From a mathematical point of view, this implies that the structural organization of the capsid building blocks, called capsomers, and their constituent protein subunits, displays a characteristic set of rotational symmetry axes with two-, three- and five-fold symmetry (Figure 1a). Denoting the locations of the protein subunits in the corners of the triangles gives rise to characteristic protein clusters (cf. clusters of grey spheres in Figure 1b). Crick and Watson provided

a biological explanation for this surprising degree of symmetry in virology. They argued that viruses encode only a small number of distinct protein building blocks, which are then repeatedly synthesized from the same gene, as this minimizes the part of the genome required to code for the capsid. For instance, hepatitis B virus and phage MS2 only encode one structural protein each and only have four genes altogether. At the same time, building a capsid from the largest known rotational symmetry, the icosahedral symmetry, ensures that the maximal possible number of subunits is used to form the capsid, thus optimizing its volume. This is known as the *principle of genetic economy*. It is a consequence of the selective pressure in viral evolution to generate capsid structures that make genome packaging as easy as possible, thus optimizing an essential step in any viral replication cycle.

Mathematical models of viral symmetry

Symmetry alone is not sufficient to explain all aspects of virus architecture as can be seen from the plethora of distinct capsid structures in nature that all obey icosahedral symmetry. Mathematics can play a key role in formulating the rules according to which viral capsids are organized. The first mathematical models of capsid architecture were introduced by Caspar and Klug in 1962. They are based on the *principle of quasi-equivalence,* which stipulates that protein subunits organize locally into equivalent environments. From a mathematical point of view, this implies that virus capsids should be describable by surface lattices. Caspar and Klug used triangulations of the capsid surface, in which protein organization in the triangular facets mimics that of the icosahedral faces in the simplest viruses. Such models can be built by drawing an
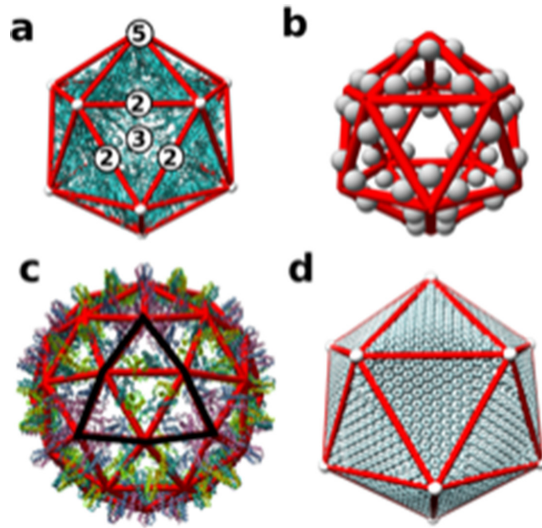
Figure 1. Historic mathematical models of virus architecture. (a) Viruses exhibit icosahedral symmetry, as exemplified for Satellite Tobacco Necrosis Virus (pdb-id 4bcu). Examples of particle 2-, 3-, and 5-fold symmetry axes are indicated on an icosahedral reference frame. (b) Protein positions (grey) in an icosahedral virus model. (c) Example of a model (here a T=4 triangulation for Hepatitis B virus based on pdb-id 3j2v) in Caspar and Klug's quasi-equivalence theory; an icosahedral face subdivided into 4 triangular facets is indicated by a black triangle. (d) Virus capsids with very large triangulation numbers have recently been discovered in giant viruses (here based on pdb-id 1m4x of Paramecium Bursaria Chlorella Virus, with a T-number of 169, shown together with an icosahedral reference frame).
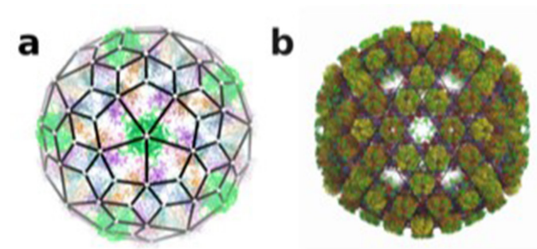

Figure 2. Generalized tiling models of virus architecture. (a) Viral Tiling theory model of Human Papillomavirus (HPV) (based on pdb-id 3j6r); interactions between three proteins (trimer interactions) are represented by kite-shaped tiles, and interactions between two proteins (dimer interactions) by rhombic tiles. (b) Herpes Simplex virus exhibits the architecture of an Archimedean surface lattice based on the generalised quasi-equivalence principle (based on pdb-id 6cgr).

icosahedral net on a hexagonal lattice and then folding this net up into an icosahedron. In their seminal theory, they provide a classification of virus architecture in terms of such triangulations, deriving polyhedral models that indicate the positions of individual capsid proteins in the surface lattice. Indeed, it is possible to break a triangle down into smaller triangles. An example of a virus structure that can be modelled in this way is the hepatitis B virus, where an icosahedral face is broken down into four triangular facets (Figure 1c). This idea leads to the triangulation number , which counts how many smaller triangles each icosahedral face consists of (e.g. four in the example above). The different triangles are then not necessarily symmetry-equivalent in a mathematical sense, but are in approximately equivalent local environments, which is why capsid architectures, according to Caspar and Klug, are called quasi-equivalent. A large fraction of the only recently discovered giant viruses also exhibit this design principle (Figure 1d). The largest precisely known structure is that occurs in *Cafeteria roenbergensis* virus. It is so large that it even has a virophage (the *Mavirus* virophage) associated with it. Mimiviruses are even larger, with an estimated -number around 1000.

Similarly, they have a virophage called Sputnik, which itself has a substantial triangulation number of 27. Caspar–Klug-type cage architectures also occur in other areas of science, where repeated building blocks are used to construct higher-order structures. Examples are carbon fullerenes, cellular compartments (such as carboxysomes), and they have even been engineered as geodesic domes in architecture. A major conclusion from Caspar–Klug theory is that only certain triangulation numbers, and thus numbers of capsid proteins, are possible.

With the advent of more refined imaging techniques, in particular, the recent revolution in cryo-electron microscopy (cryo-EM), it has become clear that these Caspar–Klug-type models are too restrictive in order to explain all known capsid architectures. Prominent examples are the cancer-causing papilloma viruses (e.g. *human papillomavirus* (HPV) in Figure 2a) that have capsids in which every protein subunit takes on one of two distinct types of local configuration. Viral tiling theory, a first generalization of Caspar–Klug theory, had been introduced to describe such non–quasi-equivalent capsid architectures, in which the proteins are not in approximately equivalent local environments, as the bonds they are forming with the surrounding proteins are not identical. This is done via tessellations akin to the famous Penrose tiling, in which distinct types of tiles represent the different types of biological interactions. In this case, kites represent three proteins forming a trimer interaction, and rhombs two proteins involved in a dimer interaction. Moreover, a *generalized principle of quasi-equivalence* has recently been introduced that also encompasses the architectures of viral capsids formed from more than one type of capsid protein such as herpes simplex virus (Figure 2b), or the dsDNA tailed phage Basilisk. This principle stipulates that local interactions between

identical proteins, as well as interactions between the same types of distinct proteins, must be the same across the entire capsid surface. In this framework, capsid architectures are modelled based on more general types of lattices called Archimedean lattices. This theory contains the hexagonal surface lattices from Caspar–Klug theory as a special case, and has a number of interesting implications for the geometric constraints on viral evolution. For example, it suggests that the size gaps between capsid architectures in Caspar–Klug theory may be bridged by capsid structures abiding to these more generalized lattice types. It even suggests a way in which larger capsid architectures may have evolved from smaller ones: the gyration of the surface lattice, whereby the relative sizes of pentagonal and triangular faces vary, resulting in a rotation of the protein subunits.

Fighting viruses with mathematics

The importance of symmetry is not only confined to the capsid surface itself, but also can manifest itself at different radial levels of a virus particle. In particular, if genome and capsid co-assemble, mediated by specific points of contact, the symmetry from the capsid impacts on the organization of the packaged genome. In order to formulate these arising mathematical constraints on genome organization, deeper mathematical concepts called root systems are required. By extending this concept to the specific example of icosahedral symmetry, the symmetry of virus capsids, it has been possible to derive a classification of nested shell arrangements that capture virus architecture at different radial levels (Figure 3a). These structures are not only relevant in virology, but also occur in the context of multi-shell fullerene structures in carbon chemistry, such as the nested carbon cages known as carbon onions (Figure 3d).

An important feature of these nested shell models is that, when applied to viruses, they pinpoint the positions between the genetic material of a virus and its capsid shell (Figure 3a, inset). Such information is important because it formulates constraints on where such contacts can be located in the genome as a travelling salesman problem: that is, as the combinatorial problem of how the nodes in a network can be visited precisely once along its edges. Indeed, by connecting all vertices corresponding to neighbouring binding sites into a polyhedral shell (Figure 3b), the order in which contacts are formed between secondary structure elements in the genome and the capsid shell can be represented as a path on a polyhedron (Figure 3c). Seen through this lens of geometry, and combined with bioinformatics and in collaboration with the experimental team led
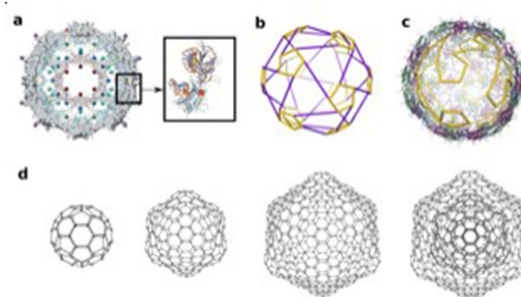

Figure 3. Radial models of virus architecture and packaging signals. (a) Virus organisation at different radial levels reveals a molecular scaling principle relating the positions and dimensions of different viral components; a 3D multi-shell model for bacteriophage MS2 is shown superimposed on a cryo-EM map from the Ranson lab (University of Leeds). Specific vertices of the model are positioned at the contact sites of genomic RNA and the inner capsid surface (inset). (b) These vertices form the corners of a polyhedron; (c) paths connecting vertices along its edges have been used as constraints in a bioinformatics approach (Hamiltonian Path Analysis) to identify secondary structure elements (packaging signals) in the viral genome in contact with the inner capsid shell. (d) Similar multi-shell models also occur in carbon chemistry, where several nested fullerene cages form a carbon onion whose structures are orchestrated collectively by an overarching symmetry principle.

by Peter Stockley at the University of Leeds, it has been possible to identify the molecular characteristics of these contacts between the genome and the capsid shell via an approach called *Hamiltonian path analysis*. This revealed an unsuspected phenomenon: the presence of multiple dispersed, sequence-specific contacts between capsid and genome (secondary structure features), which were termed *packaging signals*. These act collectively and cooperatively to orchestrate efficient co-assembly of the capsid around its genome, akin to clothing pegs on a washing line. Packaging signals constitute a second code, overlaid on top of the genetic code of the virus, that functions like a virus capsid assembly manual. Their discovery has opened up novel avenues for antiviral therapy that are based on both geometric and biophysical insights into capsid assembly.

The role of symmetry breaking

Viral capsids must perform different functions: package the viral genome efficiently, protect their cargoes whilst acting as a delivery vehicle, and finally, release it in response to cues from the host environment. This is in many cases facilitated by additional capsid components that break the capsid's overall icosahedral symmetry. Prominent examples are dsDNA phages with helical

tails, packaging motors that enable energy-driven internalization of the genomic DNA, or portals, such as the stargate in Mimivirus. At the other end of the size spectrum, one of the capsid protein dimers in the bacteriophage MS2 capsid is replaced by maturation protein, which enables attachment of the particle to the bacterial pilus at this distinguished site, and thereby its internalization into the bacterial host. Mathematical modelling can help better understand the consequences of such asymmetric capsid features and how they drive vital dynamic processes in the viral life cycle. The larger the viruses are, the more complex their structural organization becomes. Coronaviruses have one of the longest RNA genomes, presenting a challenge for its packaging into the confines of the particle volume; this is overcome by additional protein components, the nucleocapsid (N) protein, that aid compaction of the genome.
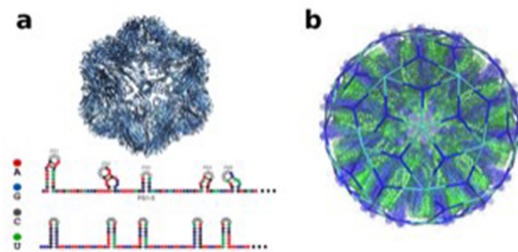

Figure 4. Artificial nanocages for applications in biomedicine and nanotechnology. (a) Knowledge of the packaging signal-encoded virus assembly instructions (top) enables the recoding of the nucleic acids (bottom) to enhance their packaging properties and the assembly efficiency of the surrounding capsid, as shown here for STNV. (b) Viral Tiling theory applied to the *de novo* engineering of self-assembling protein nanoparticles used in malaria vaccine design.

Turning tables on viruses – nanotechnology mining from nature

In addition to pointing the way to new types of antiviral strategies, a better understanding of viral geometry also opens up novel routes for drug delivery and vaccination: either by repurposing and optimizing viral protein containers or by *de novo* engineering containers based on similar geometric design principles. An example of the former would be exploiting and optimizing the virus assembly instructions as we are currently doing in collaboration with the Stockley lab, tuning it for optimal assembly efficiency as demonstrated for satellite tobacco necrosis virus (STNV; Figure 4a). An example of *de novo* design are nanoparticles that form from a protein building block with two different oligomerization domains that spontaneously form cages with local three- and five-fold axes (Figure 4b). These self-assembling protein nanoparticles (SAPNs) share structural similarities with papillomaviruses and can be modelled in terms of surface tessellations analogously to viral tiling theory. Such tiling models, indicating the positions of individual protein chains, have been used to analyse the particle morphologies of assembly products that arise experimentally and to mathematically reconstruct their surface architectures and properties. These SAPNs are currently used for the design of malaria vaccines.

A better understanding of virus capsids and their symmetries has, therefore, paved the way to new antiviral strategies and to the repurposing of the genome-encoded virus assembly instructions for engineering artificial virus-like particles. Such particles have a host of applications in nanotechnology, ranging from cargo storage, over drug delivery to diagnostics. Viruses are highly sophisticated molecular machines. We are just at the beginning of the tantalizing journey of unravelling how they work in detail, potentially with profound impacts in biomedicine and nanotechnology. ■

Further reading

- Crick, F.H.C. and Watson, J.D. (1956) Structure of small viruses. *Nature* **177**, 473–475. DOI: 10.1038/177473a0
- Caspar, D.L. and Klug, A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24. DOI: 10.1101/sqb.1962.027.001.005
- 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. and Luque, A. (2019) Structural puzzles in virology solved with an overarching icosahedral design principle. *Nat. Commun.* **10**, 1–9. DOI: 10.1038/s41467-019-12367-3
- Dechant, P-P. , Boehm, C. and Twarock, R. (2012) Novel Kac-Moody-type affine extensions of non-crystallographic Coxeter groups. *J. Phys. A.* **45**,285202. DOI: 10.1088/1751-8113/45/28/285202

Continued

- Keef, T. , Wardman, J.P. , Ranson, N.A., et al. (2013) Structural constraints on the three-dimensional geometry of simple viruses: case studies of a new predictive tool. *Acta Crystallogr. A.* **69**, 140–150. DOI: 10.1107/S0108767312047150
- Dechant, P.P., Wardman, J., Keef, T. and Twarock, R. (2014) Viruses and fullerenes - symmetry as a common thread? *Acta Crystallogr. A.* **70**, 162–167. DOI: 10.1107/S2053273313034220
- Twarock, R. , Leonov, G. and Stockley, P.G. (2018) Hamiltonian path analysis of viral genomes. *Nat. Commun.* **9**, 2021. DOI: 10.1038/s41467-018-03713-y
- Twarock, R. and Stockley, P.G. (2019) RNA-mediated virus assembly: mechanisms and consequences for viral evolution and therapy. *Annu. Rev. Biophys.* **48**, 495–514. DOI: 10.1146/annurev-biophys-052118-115611
- Indelicato, G., Wahome, N., Ringler, P. et al. (2016) Principles governing the self-assembly of coiled-coil protein nanoparticles. *Biophys. J.* **110**, 646–660. DOI: 10.1016/j.bpj.2015.10.057

Pierre-Philippe Dechant is a Senior Lecturer in Mathematical Sciences and the Programme Director for the Data Science Degree Apprenticeship at York St John University. Pierre received his PhD from Cambridge, where he worked on symmetry principles in gravitational and particle physics, before moving to York to start work in Mathematical Virology. His research combines computational and mathematical modelling, often involving symmetry applications that span biology, physics and algebra. Email: p.dechant@yorksj.ac.uk

Reidun Twarock is Professor of Mathematical Virology at the University of York. She is an EPSRC Established Career Fellow in Mathematics, a Royal Society Wolfson Fellow, and together with experimentalist Peter Stockley from the University of Leeds, a Wellcome Trust Investigator. Reidun's research in Mathematical Virology, an area pioneered by her, focuses on the development of mathematical and computational techniques to elucidate how viruses form, evolve and infect their hosts. She has won the Gold Medal of the Institute of Mathematics and Its Applications in 2018. Email: reidun.twarock@york.ac.uk