Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Principles of Virus Structure

Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†], Purdue University, West Lafayette, IN, United States

Glossary

Ångstrom A unit of length equal to one hundred-millionth of a centimeter to measure wavelengths and distances (1 Å = 0.1 nm).
Capsid The protein shell of a virus particle.
Capsomere The basic structural unit of a capsid which can self-assemble to form the complete capsid.
Envelope A host derived lipid membrane separating the genome from the viral capsid.
MCP Major capsid protein.
Nucleocapsid The protein capsid and the genome.
Spike Glycoprotein surface projections of varying length at regular intervals on a viral surface.
Structural proteins Proteins present in the virus particle.
Virion A virus particle capable of infecting a host cell.
VLP Virus like particle (VLP) consisting of the outer capsid without the genome.

Introduction

Viruses are nanomolecular assemblies tailored to parasitize a wide variety of hosts ranging from prokaryotes to eukaryotes. A virus particle consists of a genome, which can comprise either single- or double-stranded RNA or DNA molecule(s) protected by a protein or a proteolipid outer capsid. Viruses insert their genome into a host organism where they can replicate and assemble. They therefore have a strong influence on protein adaptation and are key drivers of molecular evolution (Enard *et al.*, 2016). Viruses occur in distinct shapes, sizes and symmetries designed for effective transmission of their genetic material. They are classified based on the type and organization of the encapsulated genome and the presence of a lipid envelope derived from the host cell surrounding the viral nucleocapsid.

Viruses occur in three predominant shapes: filaments, spheres and pleomorphic (**Fig. 1**). The spheres are 20-sided regular icosahedra. The size of icosahedral viruses varies from a diameter of 170 Å (porcine circovirus) to 5000 Å (mimivirus). However, pithovirus and pandoravirus, the largest known viruses, are oval and measure about 10,000 Å and 15,000 Å in size, respectively. Mimivirus, a double-stranded DNA (dsDNA) virus with a genome size of 1.2 megabase pairs has an icosahedral shape with a diameter of 5000 Å and is associated with 1200 Å-long fibers (Klose *et al.*, 2010; Raoult *et al.*, 2004; Xiao *et al.*, 2005,2009). Most plant viruses are small and are either filamentous or polygonal, as are also many bacterial viruses (Johnson and Speir, 1997; Rossmann, 2013; Solovyev and Makarov, 2016). The larger and more complex dsDNA containing bacteriophages, which infect *E. coli*, combine both filamentous and polygonal shapes (Ackermann, 1999). For example, the T4 bacteriophage is composed of an icosahedral head (1150 Å long, 210 Å in diameter) (Fokine *et al.*, 2004) with an associated 1000 Å rod-shaped long tail (Kostyuchenko *et al.*, 2005), a 460 Å-diameter baseplate and six 1450 Å-long tail fibers (Kostyuchenko *et al.*, 2003). Structures such as these are unique to bacteriophages. Animal viruses, however exhibit extreme variations in size and shape. For example, parvoviruses are icosahedrons and measure about 180–250 Å in diameter. In contrast, members of *Poxviridae* and *Togaviridae* are about 250–750 Å in diameter and are either polygons, filamentous or pleomorphic.

The viral genome encodes structural proteins, which constitute the virus particle and non-structural proteins essential for the replication of the virus inside the host cell. These proteins aid in several stages of a typical virus life cycle that includes host cell recognition, replication, assembly, maturation and the release of progeny virus particles. The capsid protects the viral genome against harsh conditions like variation in pH, temperature and degradation by enzymes during infection. The capsid also plays a role in host cell recognition and entry. Although the number and nature of the capsid building blocks vary among viruses, there are certain common principles of capsid organization that apply to most viruses.

Several excellent reviews on the principles of virus structure organization have been published over the years (Caspar and Klug, 1962; Harrison, 2007; Johnson and Speir, 1997; Klug and Caspar, 1960; Lee and Johnson, 2003; Prasad and Schmid, 2012; Rossmann, 2013; Rossmann and Johnson, 1989). Here, we give a short history of structural virology, a discussion of virus symmetry and the methods used to study virus structures followed by the current understanding of the principles of virus architecture. We will be discussing a series of example virus structures to show: (1) how virus capsid subunits are assembled, (2) what triggers their disassembly to release their genome, (3) why certain protein folds occur frequently in virus capsids, (4) how viruses recognize and enter host cells, and (5) how structures aid in the development of vaccines and anti-viral therapeutics.

General Principles of Capsid Architecture

The first significant development in structural virology was in 1935, when Stanley crystallized the infectious particles of tobacco mosaic virus (TMV) (Stanley, 1935) and subsequently Bawden and Pirie crystallized tomato bushy stunt virus (TBSV)

[†]Deceased.

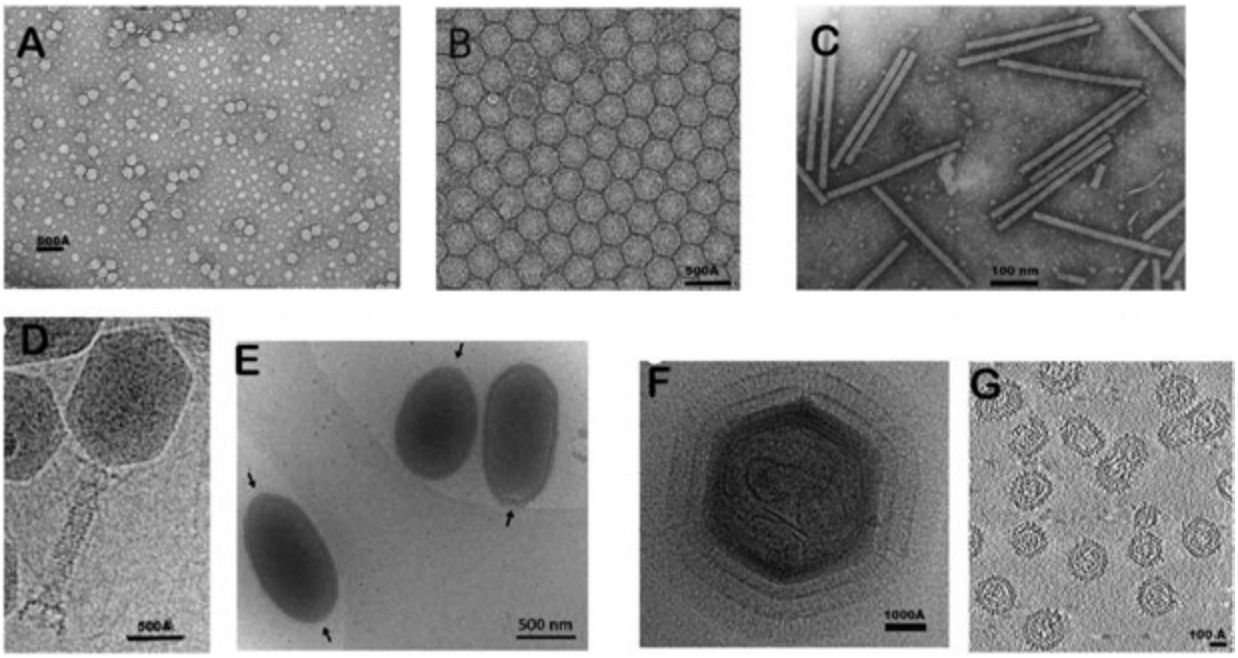

Fig. 1 *Morphology of viruses.* Micrographs showing icosahedral capsids of parvovirus (A), enterovirus 71 (B), the bacteriophage T4 head (D) and mimivirus (F). Micrographs of the filamentous structures in TMV (C) and the bacteriophage T4 tail (D). Micrographs of the oval-shaped capsids of pithoviruses are shown in (E) and of the pleomorphic capsids of rubella in (G). The scale bars in each figure are shown with their respective units. The micrographs were adapted from: (A) Hafenstein, S., Palermo, L.M., Kostyuchenko, V.A., *et al.*, 2007. Asymmetric binding of transferrin receptor to parvovirus capsids. Proceedings of the National Academy of Sciences of the United States of America 104, 6585–6589. (B) Plevka, P., Lim, P.Y., Perera, R., *et al.*, 2014. Neutralizing antibodies can initiate genome release from human enterovirus 71. Proceedings of the National Academy of Sciences of the United States of America 111, 2134–2139. (C) http://www.dpvweb.net/notes/showem.php?genus=Tobamovirus. (D) Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., Leiman, P.G., 2004. The bacteriophage T4 DNA injection machine. Current Opinion in Structural Biology 14, 171–180. (E) Okamoto, K., Miyazaki, N., Song, C.H., *et al.*, 2017. Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. Scientific Reports 7. (F) Xiao, C., Kuznetsov, Y.G., Sun, S., *et al.*, 2009. Structural studies of the giant Mimivirus, PLOS Biology 7, e1000092. (G) Mangala Prasad, V., Klose, T., Rossmann, M.G., 2017a. Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus. PLOS Pathogens 13, e1006377.

(Bawden and Pirie, 1938). In 1941 Bernal and Fankuchen used X-ray diffraction to establish the size and shape of TMV and TBSV particles and presented the first detailed description of any virus structure (Bernal and Fankuchen, 1941a,b). They established that the TMV virus particles consisted of a protein capsid that protected the viral genome. This work was then extended to near-atomic resolution by Franklin (Franklin, Holmes, 1958; Franklin *et al.*, 1957), Holmes (Holmes *et al.*, 1975), Stubbs (Stubbs *et al.*, 1977), and Bloomer (Bloomer *et al.*, 1978; Champness *et al.*, 1976). The description of the assembly of TMV capsid subunits around the RNA genome was among the foremost studies of virus capsid self-assembly (Butler and Klug, 1972).

In 1953, James Watson and Francis Crick reached their ground-breaking conclusion, using Franklin's X-ray diffraction images, that DNA exists in the form of a three-dimensional (3D) double helix with the two strands connected by hydrogen bonds between the nucleotides (Watson and Crick, 1953b). They also postulated the 'Central Dogma', which implied that the information flow in biological systems is from DNA to RNA to protein (Watson and Crick, 1953a). They observed that a much larger genome than observed in all known viruses would be required to code for a protein of sufficient size to completely cover the virus genome. Thus, the capsid must be constructed of multiple copies of the same protein. As each of these subunits have the same sequence and therefore the same structure, the environment of each subunit must be the same. Hence, they suggested that the capsid structure of viruses have to be regular polyhedrons. Of the five possible polyhedra, the icosahedron would have the greatest number of subunits and would, therefore, represent the most conservative way of using the least amount of genome to code for the largest possible virus (Crick and Watson, 1956; Crick and Watson, 1957). This was verified by Caspar (Caspar, 1956), who observed in an X-ray crystallographic study of TBSV that there appeared to be rotation axes at angles to each other consistent with icosahedral symmetry. The icosahedral symmetry in TBSV was also confirmed using electron microscopy (Horne *et al.*, 1959; Williams and Smith, 1958).

Icosahedral Symmetry

Caspar and Klug provided the first evidence for the presence of icosahedral symmetry by analyzing the X-ray diffraction patterns of TBSV and turnip yellow mosaic virus (TYMV). An icosahedron has 12 vertices with 5-fold rotational symmetry, 20 triangular faces

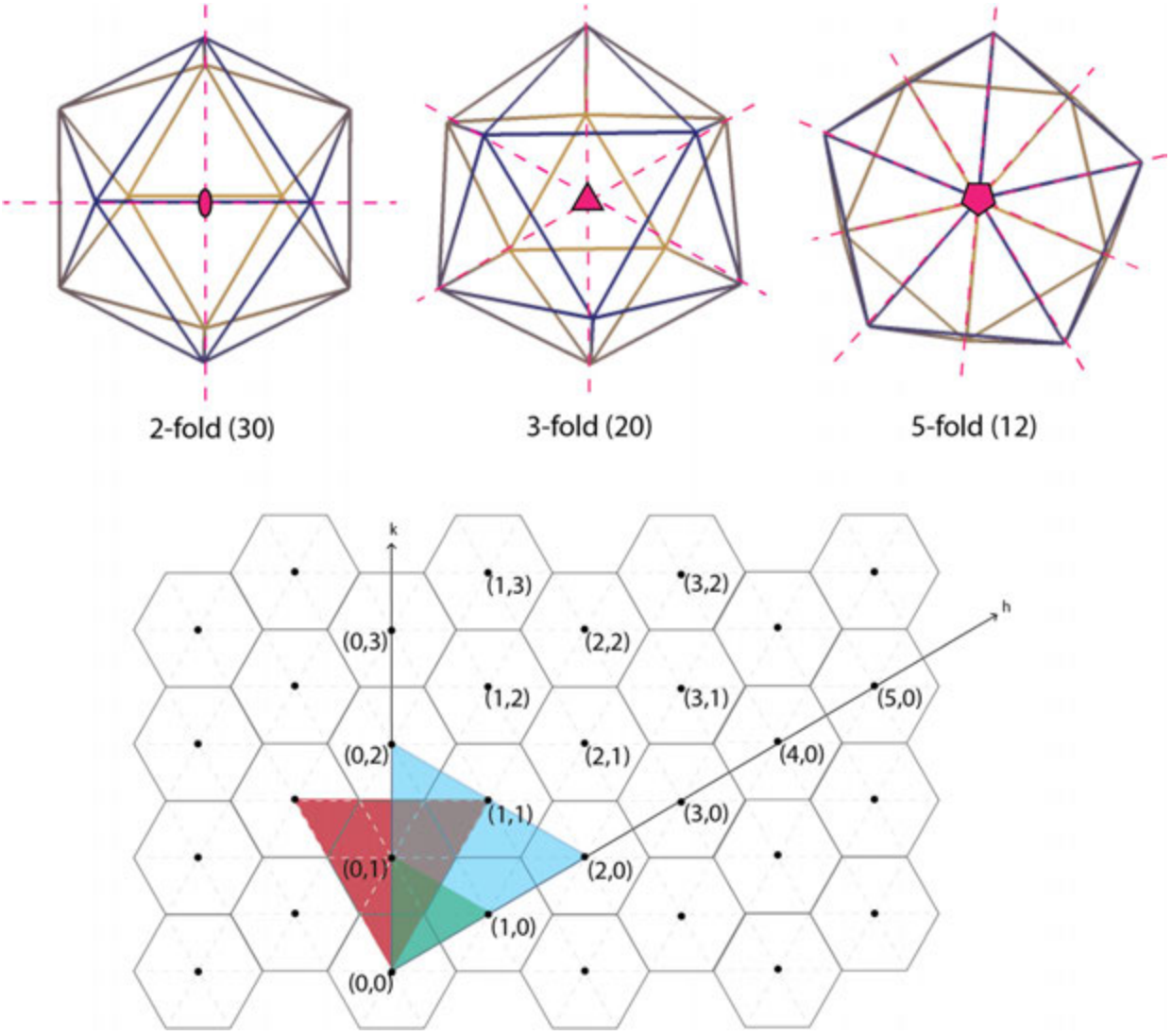

Fig. 2 *Icosahedral symmetry and triangulation.* An icosahedron displayed along the 2-fold (A), 3-fold (B) and 5-fold (C) axes of symmetry. An icosahedron has twelve 5-fold axes of symmetry along each of the vertices, twenty 3-fold axes of symmetry passing through the center of triangular faces and thirty 2-fold axes through the middle of each edge. (D) The hexagonal lattice system to define triangulation (T number) of an icosahedron. The hexagonal lattice is constructed with h and k axes crossing at 60° angles. Each hexagon can be decomposed into six equilateral triangles. The coordinates on the hexagonal lattice system (h, k) are labeled. Examples of three triangles with T numbers 1, 3 and 4 are shown in green, red, and blue respectively.

with 3-fold rotational symmetry and 30 edges with 2-fold rotational symmetry (Fig. 2). This allows the placement of 60 identical subunits decorated with capsid proteins with equivalent contacts between them. However, there are many spherical viruses where the number of capsid subunits far exceed the 60 subunits required by icosahedral symmetry. To address this question, Caspar and Klug proposed an extension to Watson and Crick's strict icosahedral symmetry concept in viruses (Caspar and Klug, 1962). They suggested that capsid subunits in an icosahedron might adapt to slightly different inter-subunit interfaces and introduced the principle of quasi-equivalence for larger virus structures with more than 60 capsid subunits. They predicted a virus assembly process in which the subunits assembled into hexagonal arrays. This would give rise to a scheme where an icosahedral virus can be constructed from pentamers of triangular faces (pentons) forming the 12 vertices and hexamers of triangular faces (hexons) covering the 20 icosahedral faces. The assumption that the monomers that make a hexamer would have quasi-equivalent environments increased the number of subunits in an icosahedron from 60 to 60 T, with T representing the triangulation number.

The triangulation number (T) is the number of quasi-similar subunits per icosahedral asymmetric unit with possible values of T, because of geometric restrictions, dictated by the equation $T = h^2 + hk + k^2$. Here, h and k are the number of hexons between pentons along the h and k axes of a hexagonal array, where h and k are either zero or positive integers. To further elaborate, the formation of an icosahedron can be visualized using a uniform hexagonal lattice, where each hexagon can be decomposed into six equilateral triangles. The hexagonal lattice is constructed with h and k axes crossing at 60° angles. A pentagon can be constructed in 3D by removing one triangle from the hexagon (Figs. 2 and 3). By arbitrarily choosing one of the 5-fold vertices as the origin (h, k = 0, 0),

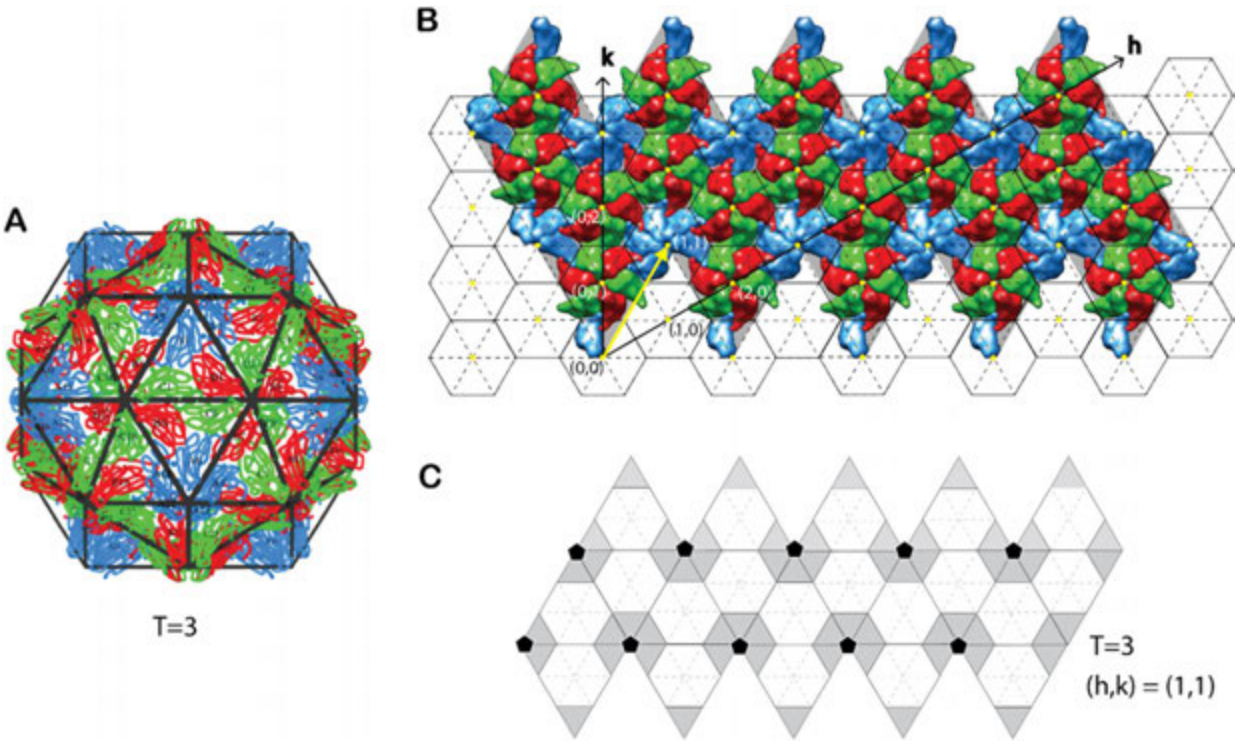

Fig. 3 *The construction of an icosahedron from arrays of hexagons.* T = 3 structure of turnip yellow mosaic virus, where (A) shows the icosahedral capsid assembly in TYMV (PDB ID: 2FZ2) and (B) the planar sheet of twenty equivalent triangles superposed on the hexagonal lattice. The T = 3 arrangement with h = 1 and k = 1 is indicated by a yellow arrow. (C) Schematic showing a planar sheet of 20 triangles which can then be folded up to form a closed icosahedron. Figures (A) and (B) were prepared using the Viper data base.

the positions h and k of the neighboring 5-fold vertex closest to the origin gives the T-number of that icosahedron. For example, T = 1 represents an icosahedron with the closest 5-fold vertex positioned at either h = 1 and k = 0 or at h = 0 and k = 1. T = 3, however, represents an icosahedron with the closest 5-fold vertex positioned at h=1 and k =1 (**Fig. 3**). Thus, a simple icosahedron with 60 identical subunits would have T=1 symmetry. The structures of tobacco bushy stunt virus (TBSV) (Harrison *et al.*, 1978) and southern bean mosaic virus (SBMV) (Abad-Zapatero *et al.*, 1980), the first near-atomic resolution structures of icosahedral viruses, had T = 3 quasi-symmetries consistent with the predictions of Caspar and Klug (subsection "Non-enveloped Icosahedral RNA Viruses"). Although the amino acid sequence of the capsid proteins was rather different, both structures were very much alike with three identical protein subunits constituting the icosahedral asymmetric unit. The first animal virus structure, that of human common cold virus strain 14 (HRV14) (Erickson *et al.*, 1983; Rossmann *et al.*, 1985), mimicked the structures of TBSV and SBMV plant viruses. However, in HRV14, the three independent subunits in the icosahedral asymmetric unit were different proteins, referred to as viral proteins VP1, VP2, and VP3 (subsection "Non-enveloped Icosahedral RNA Viruses").

Based on the equation $T = h^2 + hk + k^2$, only certain T-numbers are allowed. Therefore, the icosahedra can be further divided into three classes (Caspar and Klug, 1962; Prasad and Schmid, 2012). These classes are calculated using the formula $T = Pf^2$, where f is the largest common divisor between h and k. The first class, P = 1 (T = 1, 4, 9, 16, 25, 36 etc.), where h is greater than or equal to 1 and k = 0, the lattice lines run parallel to the edges of the triangular icosahedral facet. In the second class, P = 3 (T = 3, 12, 27, 48, 75, 108 etc.), the icosahedra have T numbers along the line that bisects the h and k axes with h = k. The third class, called the skew class, includes icosahedra with other allowed T numbers (7, 13, 19, 21, 28, 31 etc.) with the possibility of having enantiomorphic configurations. In T = 7 icosahedra, the asymmetric unit is made of 7 proteins giving a total of 420 capsid proteins per virus capsid. However, depending on the values of h and k, the capsid can either have a configuration of dextro (d) for h = 1, k = 2 or laevo (l) for h = 2, k = 1. For example, polyomavirus has a non-enveloped capsid with T = 7d pseudo-symmetry because it is only composed of pentamers and lacks the expected hexamers in its capsid whereas Rotavirus, a non-enveloped, icosahedral virion with a double capsid structure has a T = 13l symmetry for the outer capsid and T = 1 for the inner capsid. Another form of icosahedron, called the prolate icosahedron where the icosahedron is stretched along one axis, is found in viruses like T-even bacteriophages (Fokine *et al.*, 2004; Rao and Black, 2010). Here, the icosahedron has a central cylindrical body made of 10 triangular faces with two caps at the bottom and the top. The prolate icosahedra are characterized by the combination of a T number and an elongation number called Q according to the formula n = 30(T + Q), where T is the triangulation number of the central body, Q is any positive integer and n refers to the number of subunits. The caps follow the triangulation T, whereas the main cylindrical body follows the Q number. When T = Q, the icosahedra becomes isometric, when Q > T, the icosahedron is referred to as prolate and when Q < T, the icosahedra is referred to as oblate.

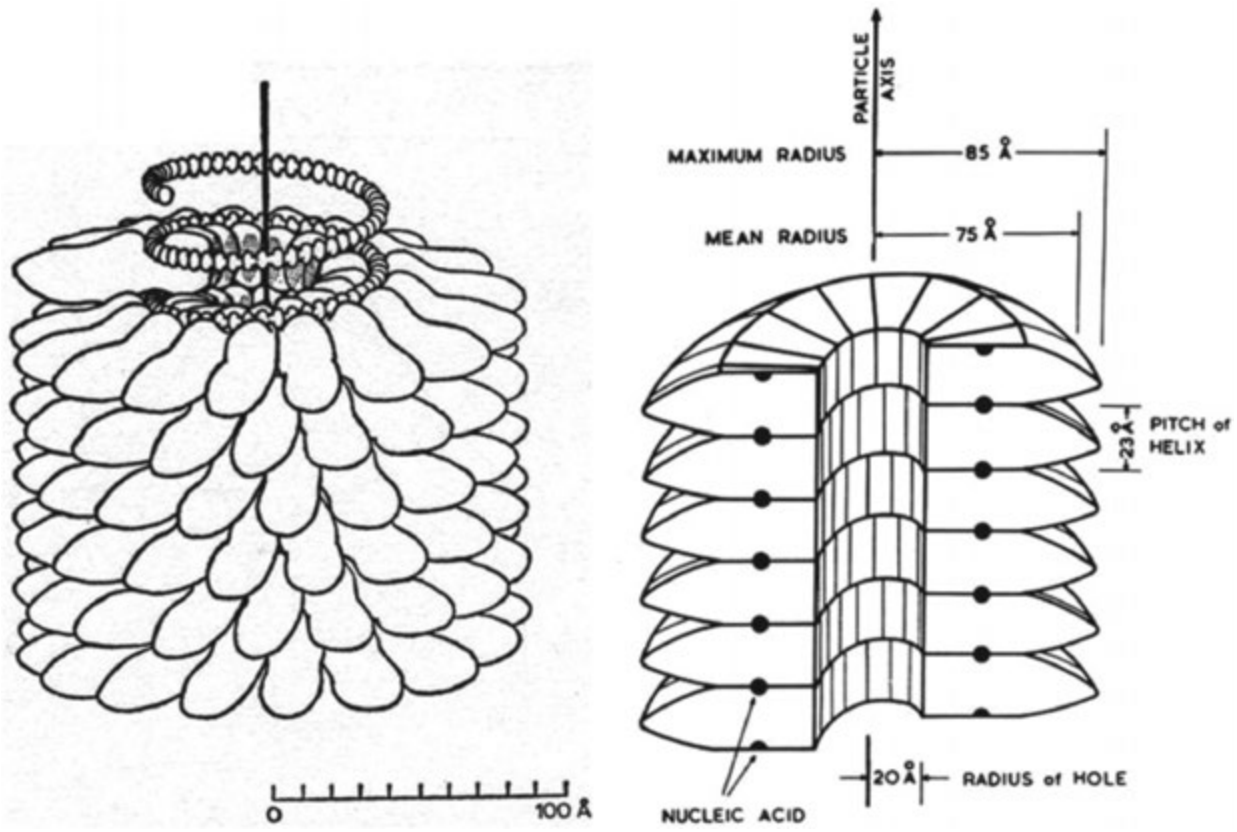

Fig. 4 Structure of tobacco mosaic virus (TMV). Figure adapted from Bernal, J.D., Fankuchen, I., 1941a. X-ray and crystallographic studies of plant virus preparations: I. Introduction and preparation of specimens. II. Modes of aggregation of the virus particles. III. (1) The structure of the particles and (2) biological implications. Journal of General Physiology 25, 111–165.

Helical Symmetry

Helical symmetry is another prevalent symmetry occurring in the capsids of numerous plant viruses, tails of bacteriophages, and flexible capsids of negative-stranded RNA genome containing viruses. Helical organization of capsids in plant viruses has the advantage that it imposes no limitations on the size of the packed RNA genome compared to icosahedral viruses which are limited by their internal volume. Helical symmetry combines rotation and translation, giving rise to a screw axis where the pitch of the helix (P) is defined as the axis rise. A helical virus structure is described by the location of subunits with respect to the helix axis, the number of subunits per helix turn (n) and the axial rise per subunit (h).

In TMV, one of the earliest and best characterized helical virus structures, the genomic RNA of 6400 bases is encapsidated by 2140 identical viral capsid proteins giving rise to a rod-like virion, 3000 Å long and 180 Å in diameter with a central hole of 20 Å in radius (**Fig. 4**). After the preliminary structural studies of TMV using X-ray fiber diffraction to 2.9 Å (Namba *et al.*, 1989a), a more recent cryoEM structure has been determined to a resolution of 3.3 Å (Ge and Zhou, 2011). In TMV the capsid protein subunits form a right-handed helix with a pitch of 23 Å and 16 and 1/3rd subunits in each turn. Each capsid subunit interacts with three nucleotides. The structure of the TMV genome follows the symmetry of the capsid providing details on capsid-RNA interactions and the assembly of the virus.

Methods for Studying Virus Structures

The principal techniques used to study intact virus structures are X-ray crystallography, X-ray fiber diffraction, cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryo-ET). To date, according to the statistics on the VIPERdb database (Carrillo-Tripp *et al.*, 2009) about 404 virus structures have been determined using X-ray crystallography and 345 structures have been determined using cryoEM. Major accomplishments in understanding the structure of viruses were closely associated with methodological, technological and computational advances in X-ray crystallography and cryoEM and developments in molecular virology techniques.

X-ray crystallography has been the most successful technique for determining the structure of individual icosahedral viruses since the 1970s. The first of the structures determined using crystallography were of icosahedral plant viruses (Abad-Zapatero *et al.*, 1980;

Harrison *et al.*, 1978) followed by more complex animal and human viruses. X-ray fiber diffraction was used to study viruses with helical symmetry. A key issue in X-ray crystallography is the solution of the phase problem, which in the case of viruses is helped enormously by the presence of non-crystallographic symmetry (NCS). Several crystallographic milestones that have also assisted in solving many significant virus structures have been (1) the use of high intensity synchrotron radiation for automated collection of diffraction images (Rossmann, 1979,1984,1999; Rossmann *et al.*,1979); (2) determining the orientation of virions in the unit cell (Rossmann and Blow, 1962; Tong and Rossmann, 1990); (3) using non crystallographic symmetry based methods to solve the phase problem (Åkervall *et al.*, 1972; Rossmann and Blow, 1963) and the use of NCS averaging for improving the phases (Buehner *et al.*, 1974; Champness *et al.*, 1976). If, for instance, isomorphous replacement had been used to obtain a 3.5 Å structure, then averaging that structure will increase the quality of the phases of the current map which can then be used to extend the resolution of the map (usually about one reciprocal lattice point in all directions). The slightly improved map can be used again for averaging and so forth.

The above real space considerations can also be expressed in reciprocal space and by doing so give insight into why phasing can be very successful by gradual resolution increase at each stage by less than one reciprocal lattice point given knowledge of the NCS. This had been under considerable dispute until the structure of HRV14 in 1985 (Arnold and Rossmann, 1986; Rossmann *et al.*, 1985) when the initial 5 Å resolution isomorphous replacement phases were extended to a beautiful 3 Å resolution map using NCS (Rossmann *et al.*, 1992a,b).

Electron microscopy of negatively stained virus samples was among the first techniques used to recognize virus shapes and sizes (Brenner and Horne, 1959; Horne *et al.*, 1959). Subsequent development of computational techniques to reconstruct the 3D structure of viruses from 2D projection images (Crowther *et al.*, 1970) and sample freezing techniques (Adrian *et al.*, 1984; Dubochet *et al.*, 1988) to preserve the native structure, revolutionized intact virus structure determination using cryoEM (Haas and Rossmann, 1970). For the next three decades, many virus structures were determined to a resolution of 25 Å. However, the more recent development of direct electron detectors as well as other technological advances in cryoEM has made it easier and more common to obtain cryoEM electron potential maps of virus samples to a resolution better than 3 Å. Prior to this, pseudo-atomic models of the virus capsid were derived by exploiting cryoEM and X-ray crystallography-based hybrid techniques. This required high resolution structures of viral components, obtained by using X-ray crystallography to be fitted into the low resolution cryoEM capsid map of the whole virus (Kuhn *et al.*, 2002; Zhang *et al.*, 2002). CryoEM has the advantage that it can be used where crystal growth is limiting especially with heterogeneous samples of enveloped viruses and viruses in complex with antibodies, receptors or anti-viral compounds. For example, crystals of West Nile virus diffracted to about 16 Å resolution (Kaufmann *et al.*, 2010), whereas cryoEM gave a map to a resolution of 10 Å (Zhang *et al.*, 2013). In the case of pleomorphic viruses such as Rubella (Battisti *et al.*, 2012; Mangala Prasad *et al.*, 2013; 2017a), cryo-electron tomography (cryo-ET) has been the method of choice to study the 3D structure. Tomography uses a series of collected images, where the microscope stage has been tilted to give a series of 2D projections that can be combined to form a 3D image (Frank, 1992; Lucic *et al.*, 2005; Subramaniam *et al.*, 2007). However, the resolution limitation for this technique (in the absence of averaging) has been about 20 Å because of the need to reduce the dose of each image to be able to collect a 3D data set before the sample has been destroyed by repeated exposure to electrons.

Structural Folds of Capsid Proteins

One of the characteristic features of viral capsid proteins lies in the folded topology of the capsid monomers, allowing them to assemble into large, symmetrical and geometrically sophisticated architectures (Chapman and Liljas, 2003; Cheng and Brooks III, 2013). The canonical jelly-roll β-barrel, as a core structural motif, is the most common capsid fold occurring in virus shells of varying sizes (Abad-Zapatero *et al.*, 1980; Chelvanayagam *et al.*, 1992; Harrison *et al.*, 1978; Johnson and Chiu, 2000). Other common protein folds occurring in viruses are the Greek key β-barrel with six strands (Choi *et al.*, 1991a; Tong *et al.*, 1992,1993), the helix bundle (Bloomer *et al.*, 1978; Conway *et al.*, 1997; Gamble *et al.*, 1997), the immunoglobulin-like fold (Rey *et al.*, 1995) and the HK97 fold (Suhanovsky and Teschke, 2015). The structurally unique capsid proteins are also functionally unique in terms of their folding topology because of the evolutionary constraints imposed by protein-protein interactions and are most likely a consequence of geometric requirements of the building blocks to form cage-like assemblies. Here, we describe the commonly occurring virus capsid folds.

Jelly-Roll β-Barrel

Both TBSV and SBMV, the first determined icosahedral virus structures, were similar with respect to their capsid structures and organization. It was then obvious that the two virus capsid structures with no apparent sequence similarity must have diverged from a common ancestral fold, later termed as a 'jelly-roll' fold by Jane S. Richardson in 1981 (Richardson, 1981), reflecting its resemblance to a Swiss roll cake. The jelly-roll fold is also referred to as a wedge shaped fold, an RNA virus capsid β-barrel or a β-sandwich (Cheng and Brooks, 2013).

A jelly-roll fold consists of eight antiparallel β-strands that are named from B to I along the polypeptide sequence (**Fig. 5**). The two four-stranded sheets that form the opposite sides of a β-barrel consist of strands BIDG and CHEF, folded such that strand B packs opposite strand C, I opposite H, etc. The end strands almost form a closed barrel structure, however there is no hydrogen-bonding between the end strands of facing sheets. A single tight turn between strands E and F leads to the pairing of strands D and G, C and H, and B and I, which then roll up, Swiss cake style. Therefore, the BI and DG pairs combine to form the BIDG sheet, and

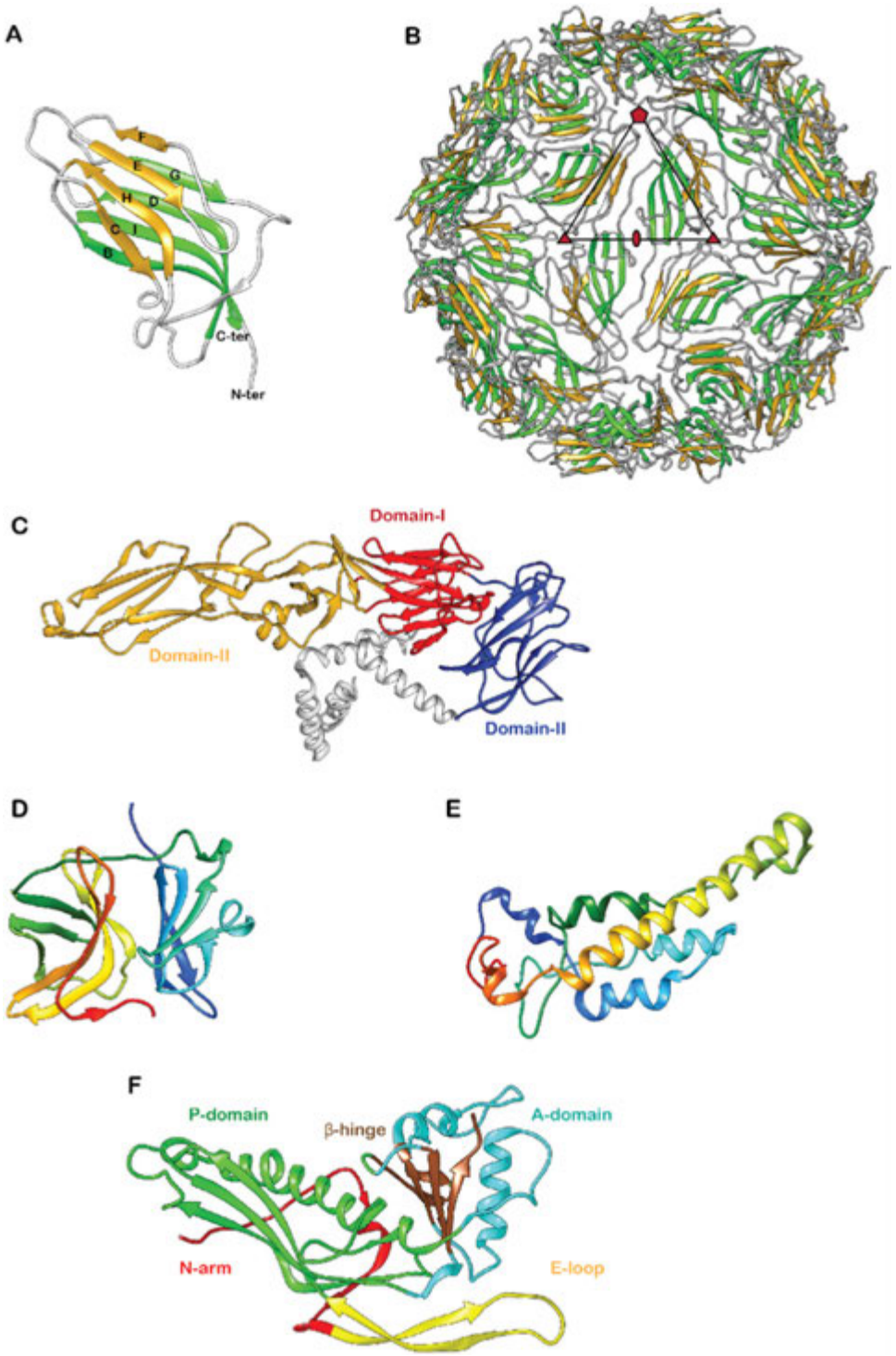

Fig. 5 *Virus capsid structure folds.* (A) and (B) show the jelly-roll β-barrel fold in satellite tobacco mosaic virus (STMV) (PDB ID: 4OQ9). The β-sheets BIDG are colored green and the β-sheets CHEF are colored golden yellow. (B) The T = 1 icosahedral assembly of the subunits to form the STMV virus capsid. The icosahedral asymmetric unit is drawn as a black triangle while the 2-, 3- and 5-fold symmetry axes are drawn in red. (C) Immunoglobulin fold in the domain-III (blue) of ZIKV envelope protein E (PDB ID: 6C08). Domain I, domain II and the membrane associated helices are colored red, yellow, and gray, respectively. (D) The serine protease-like fold in the capsid of Sindbis virus (PDB ID: 1SVP) and (E) the four-helix bundle in TMV (PDB ID: 1VTM) are shown in rainbow colors (blue to red) from N- to C-terminus. (F) The HK97 coat protein in the immature state (PDB ID: 3E8K). The common secondary structural elements observed in HK97 fold are labeled.

the CH and EF pairs combine to form the CHEF sheet. However, variations to the common core might occur in different viruses. There may be one or two additional strands like the β-strand A in TBSV outside the capsid's common fold. The length of the strands might also differ depending on triangulation number and capsid architecture. The largest variations occur in the size of the loops connecting the β-strands. In some viruses, these external loops come together to form antigenic sites (Rossmann *et al.*, 1985) and in others the loops account for the increase in capsid size (McKenna *et al.*, 1992a,b). Most of the jelly-roll folds are about 180 amino acids and can be as big as 580 amino acids in parvoviruses (Tsao *et al.*, 1991).

A large number of viruses build their capsids using either single or double jelly-roll fold capsid subunits. Double jelly-roll proteins have presumably evolved from single jelly-roll proteins by gene duplication. Single jelly-roll capsid folds are found primarily in many icosahedral RNA and DNA viruses. However, the majority are positive-sense single-stranded RNA viruses and the only dsDNA virus with single jelly-roll folds are the members of *Papillomaviridae* and *Polyomaviridae*. The double jelly-roll capsid proteins are found exclusively in dsDNA viruses spanning a large capsid size range. Most members of the double jelly-roll fold group are icosahedral, a few families such as *Poxviridae* and *Ascoviridae* have oval or brick-shaped mature virions.

The Immunoglobulin-like Fold

The immunoglobulin fold is found in a vast variety of proteins of different functions such as antibodies, cellular adhesion molecules, the receptor tyrosine kinase and many other molecules. The canonical immunoglobulin fold consists of a 7-stranded antiparallel β-sheet sandwich within which there is extensive hydrogen bonding. One example in a virus is one of the domains of a flavivirus envelope glycoprotein (E) (Rey *et al.*, 1995). The flavivirus protein E consists of three domains: The N-terminal domain-I, which is an antiparallel β-barrel quite different from the jelly-roll fold, a β-sheet rich domain-II and the C-terminal domain-III with an immunoglobulin fold. In flavivirus E protein domain-III, the three- (β-sheet CFG) and four- (β-sheet ABED) member sheets face each other and are commonly linked by a disulfide bond. The domain topology resembles a Greek key (**Fig. 5**).

The Serine Protease Fold

The serine protease fold is observed in chymotrypsin, many other proteases and viral proteins. It is again an anti-parallel Greek key ß-barrel with six-strands. Examples of viruses with this domain as core protein can be found in the capsid proteins of the members of *Togaviridae* such as Chikungunya virus (CHIKV) (Sun *et al.*, 2013b), Sindbis virus (Choi *et al.*, 1991a), Ross River virus, Rubella (Mangala Prasad *et al.*, 2013) and Eastern equine encephalitis virus (Hasan *et al.*, 2018b) (**Fig. 5**). In Sindbis virus, the capsid polypeptide is folded into two similar 'Greek key' ß-barrel domains with the substrate site situated between the domains. The capsid structure of Sindbis virus identified Ser215, His141 and Asp163 as the essential catalytic triad as in other serine proteases (Choi *et al.*, 1991a).

The Four-helix Bundle

The four-helix bundle is a common motif in which antiparallel α-helices are packed side by side with a close packed hydrophobic core (**Fig. 5**). It is also found in a wide range of proteins with diverse functions like ferritin, various cytochromes, cytokines, growth hormones etc. Viral examples include the coat protein of TMV, capsid of hepatitis B and the C-terminal domain of the HIV-1 capsid (Bloomer *et al.*, 1978; Conway *et al.*, 1997; Gamble *et al.*, 1997). Each capsid subunit of TMV consists of 158 amino acids assembled into a four-helix bundle, which are joined by a loop proximal to the virion axis. Each coat protein interacts with 3 RNA nucleotides (Namba *et al.*, 1989b).

The HK97 Fold

HK97 capsid fold is termed after the first capsid protein structure solved by X-ray crystallography of bacteriophage HK97 (Helgstrand *et al.*, 2003; Suhanovsky and Teschke, 2015; Wikoff *et al.*, 2000). Most of the viral capsids using the HK97-fold are found in bacteriophages. Additionally, this fold is also found in the capsids of Herpes Simplex virus type 1, murine cytomegalovirus, pseudorabies virus and the rhesus monkey rhadinovirus. Viruses using the HK97-fold share unexpected similarities beyond the fold of their coat proteins including the architecture of their virions, an internal scaffolding protein-mediated assembly, active dsDNA packaging, and capsid maturation events (Suhanovsky and Teschke, 2015).

The HK-97 fold is characterized by having an N-arm (sometimes α-helical rich), an E-loop with variable lengths that is a two-stranded anti-parallel β-sheet (β2 and β3), a peripheral P-domain with a long helix called the "spine helix" and an unusually long β-sheet; and an A-domain (axial domain) with a central β-sheet referred to as the β-hinge surrounded by short helices and loops (**Fig. 5**).

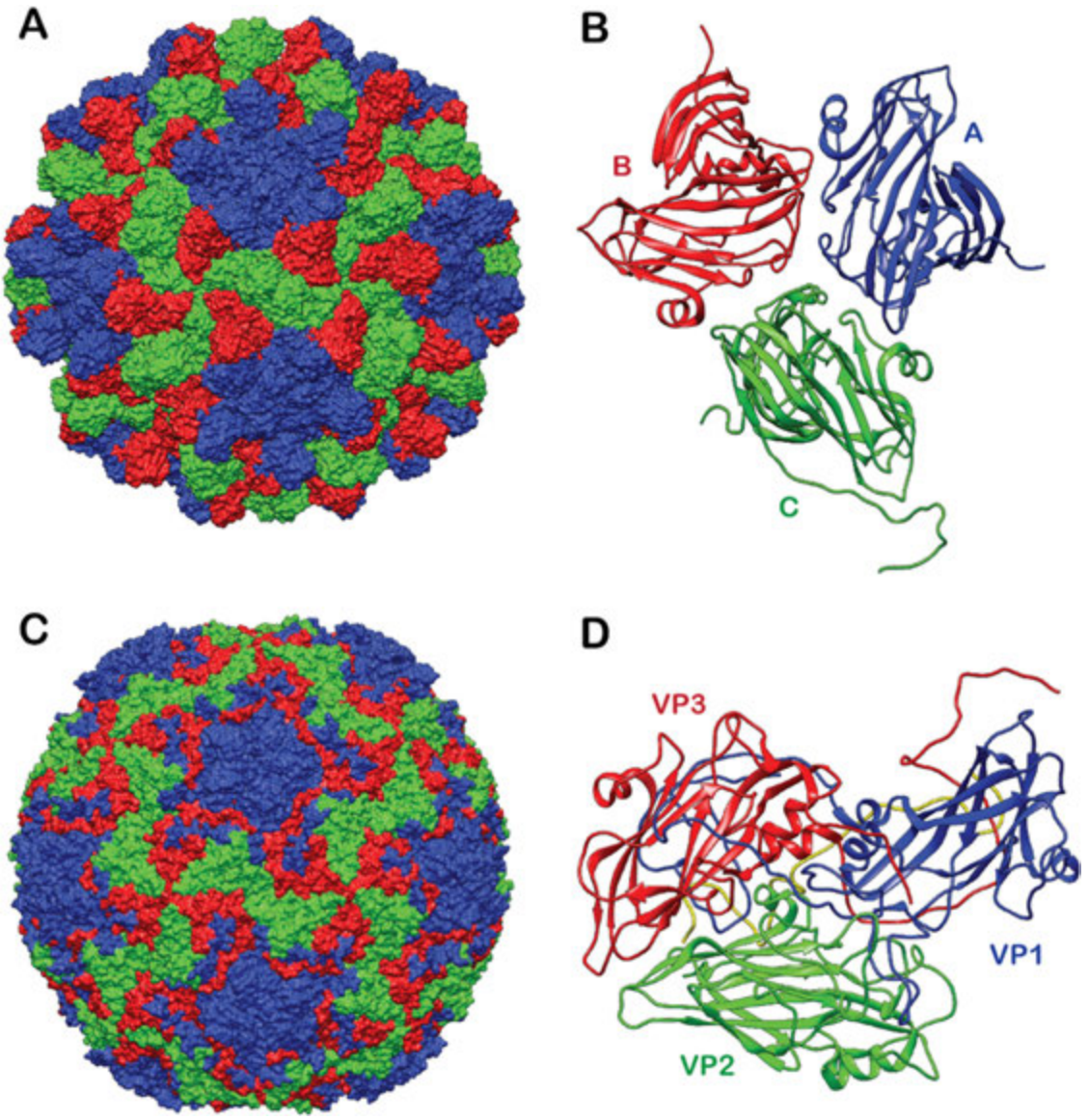

Fig. 6 *T = 3 icosahedral capsid structure of non-enveloped RNA viruses.* (A) The capsid structure of tomato bushy stunt virus (TBSV) (PDB ID: 2TBV). (C) A surface representation of human rhinovirus 14 (HRV14) (PDB ID: 4RHV). The icosahedral asymmetric unit of TBSV (B) consists of three similar subunits colored in red, blue and green, whereas in HRV14 (D) the subunits are formed by different proteins and labeled as VP1 (blue), VP2 (red), VP3 (red), and VP4 (yellow).

Capsid Assembly

Non-Enveloped Icosahedral RNA Viruses

The first icosahedral RNA plant virus structures, TBSV (Harrison *et al.*, 1978; Winkler *et al.*, 1977) and SBMV (Abad-Zapatero *et al.*, 1981), had similar capsid folds (jelly-roll fold) and organization with three independent subunits (A, B, and C) in the icosahedral asymmetric unit arranged around a quasi-3-fold axis (as predicted by Caspar and Klug). These structures clearly established that virus capsids with more than 60 subunits were organized based on a triangulated icosahedral lattice as described by Caspar and Klug (Caspar and Klug, 1962). The capsid is composed of 180 capsid subunits and exhibited a T = 3 capsid organization with rings of pentons and hexons. The assembly of these viruses occurred via AB and CC dimers. However, the AB dimer had a disordered amino-terminal β-strand compared to the one of the CC dimer, which was ordered, giving the two types of dimers a different curvature on the virus surface (**Fig. 6**).

The first animal virus structure was that of HRV14 (Arnold *et al.*, 1987; Erickson *et al.*, 1983), which mimicked the structures of TBSV and SBMV plant viruses. HRV14 belongs to the family of picornaviruses, which are small icosahedral animal viruses of about 300 Å diameter with a positive-sense eight thousand base RNA genome. However, while in the plant viruses the three protein

subunits in the icosahedral asymmetric unit were identical, the icosahedral asymmetric unit of HRV14 was composed of three independent subunits of the viral proteins: VP1, VP2, and VP3 (and VP4 which is internal in the mature virus) (**Fig. 6**). The capsid is composed of 60 copies of VP1, VP2, and VP3 (and internal VP4) arranged with pseudo T = 3 icosahedral symmetry. Nevertheless, each subunit had the same fold referred to as the "jelly-roll" fold. Shortly after the determination of the HRV14 structure and its extraordinary evolutionary implication, Hogle was able to determine the structure of poliovirus serotype 1, using the same phase extension technique as had been pioneered for HRV14 (Hogle *et al.*, 1985). This further confirmed the remarkable similarity of simple plant and animal RNA viruses.

The structure determination of HRV14 was not only a significant advancement in the method used for the analysis of X-ray diffraction data but was also informative of the functions required by the virus for infecting a cell and for replication. Viral escape mutations for a series of neutralizing monoclonal antibodies were all found to be on the virus surface and thus were likely the site of antibody binding (Sherry *et al.*, 1986; Sherry and Rueckert, 1985). The surface of HRV14 had a deep depression ("canyon") around each pentameric vertex (Rossmann *et al.*, 1985). None of the MAb binding sites were in the canyon. This, therefore, suggested that the canyon might be the site where cellular receptor molecules could bind without experiencing any inhibiting viral mutations (Rossmann *et al.*, 1985; Rossmann and Palmenberg, 1988). Thus, by hiding the receptor binding site in the canyon (which would require a long thin receptor molecule), the virus could evade the host's immune system while still being able to bind to the same receptor molecule (Rossmann, 1989). Subsequently it was established that most HRVs use intercellular adhesion molecule 1 (ICAM1) as receptor (Greve *et al.*, 1989; Staunton *et al.*, 1989). In one of the earlier applications of cryoEM, Rossmann and colleagues verified that indeed ICAM1 bound into the canyon as predicted (Olson *et al.*, 1993).

Another significant discovery related to the structure of HRV14 was that a series of anti-rhino-viral compounds bound into a hydrophobic pocket within VP1 (Smith *et al.*, 1986). This pocket was below and close to the canyon. It was shown that binding one of the antiviral compounds being developed by the Sterling Winthrop Company caused an enlargement of the pocket and a change of structure in the canyon, explaining why the binding of these compounds inhibited cellular attachment. Furthermore, it was found that this pocket was occupied by a small stabilizing lipid in HRVs and poliovirus. Thus, binding of the infectious virus to ICAM1 would displace the stabilizing "pocket factor" and prepare the virus for the release of its genome into the cytoplasm of the cell containing the ICAM1 (Filman *et al.*, 1989; Oliveira *et al.*, 1993; Rossmann, 1994; Smyth *et al.*, 1995). These results stimulated extensive work by the ViroPharma Company to create an inhibitor that was effective over a wide range of HRV serotypes, resulting in the development of "Pleconaril" (Hayden *et al.*, 2003; Xiao *et al.*, 2011). Although this compound was successful in clinical trials it put women on contraceptive hormones at risk of conception. Thus, the drug never reached the pharmaceutical market.

Enveloped Icosahedral RNA Viruses

Many major viral pathogens such as human immunodeficiency virus (HIV), hepatitis B virus (HBV), influenza viruses, herpes simplex virus (HSV), coronaviruses as well as various poxviruses, alphaviruses and flaviviruses contain host-derived lipid membranes studded with glycoproteins. Among these, members of *Hepadnaviridae* (ex: HBV) and *Herpesviridae* have an internal icosahedral capsid shell surrounded by an amorphous tegument and a lipid envelope embedded with glycoproteins. Flaviviruses and alphaviruses, however, have an icosahedral outer shell and an internal lipid membrane separating the outer glycoprotein shell from the genome (Mukhopadhyay *et al.*, 2005).

Alphaviruses have a positive-sense ssRNA genome that is about 11 kb in length and codes for 9 proteins (Kuhn, 2007). The 3′ end of the genome is transcribed for subsequent translation into a polyprotein precursor containing the three structural proteins PE2 (the precursor of E3 and E2), E1, and the capsid protein. The structures of a number of alphaviruses have been determined using cryoEM to resolutions better than 10 Å, including the structure of Chikungunya virus-like particles to 5.3 Å resolution (Sun *et al.*, 2013b), Venezuelan equine encephalitis virus to 4.6 Å resolution (Zhang *et al.*, 2011a), Eastern equine encephalitis virus to 4.4 Å resolution (Hasan *et al.*, 2018b) and Sindbis virus to 3.5 Å resolution (Chen *et al.*, 2018). Alphaviruses have an external diameter of about 700 Å and are icosahedral with quasi-T = 4 symmetry (Fuller, 1987; von Bonsdorff and Harrison, 1975; von Bonsdorff and Harrison, 1978). They have a nucleocapsid core that is completely surrounded by a lipid envelope, derived from a host membrane, into which is embedded an icosahedral array of glycoproteins. A single virus particle contains 240 copies each of the E1 and E2 glycoproteins in the mature virus (in addition 240 copies of E3 in the immature virus) that form 20 "i3" spikes situated on the icosahedral 3-fold axes and 60 quasi-3-fold "q3" spikes at general positions (Cheng *et al.*, 1995; Zhang *et al.*, 2002). There are 240 copies of the capsid protein on the internal cytoplasmic side of the viral lipid membrane, arranged as 12 pentamers about the 5-fold vertices and 30 hexamers about the icosahedral 2-fold vertices, consistent with the T = 4 symmetry of the glycoprotein on the external side of the membrane. The capsid protein of alphaviruses is similar to trypsin-like proteases and participates in the assembly process (Choi *et al.*, 1991b) (**Fig. 7**).

The E1 glycoprotein of CHIKV forms three β-barrel domains (I, II, and III). The CHIKV E2 glycoprotein is arranged into three immunoglobulin-like domains A, B and C (Gibbons and Kielian, 2002; Li *et al.*, 2010; Voss *et al.*, 2010). Domain A contains the receptor binding site, domain B is at the distal end of each spike protecting the fusion loop on DII of E1 and domain C is situated closest to the viral membrane. The E1 glycoprotein contains a hydrophobic fusion loop that is responsible for membrane fusion with an endosomal membrane when initiating infection. The E2 protein can bind to cellular receptors and protects the E1 fusion loop at neutral pH. During the initial stages of infection, the host cellular receptor is recognized by the surface glycoprotein E2.

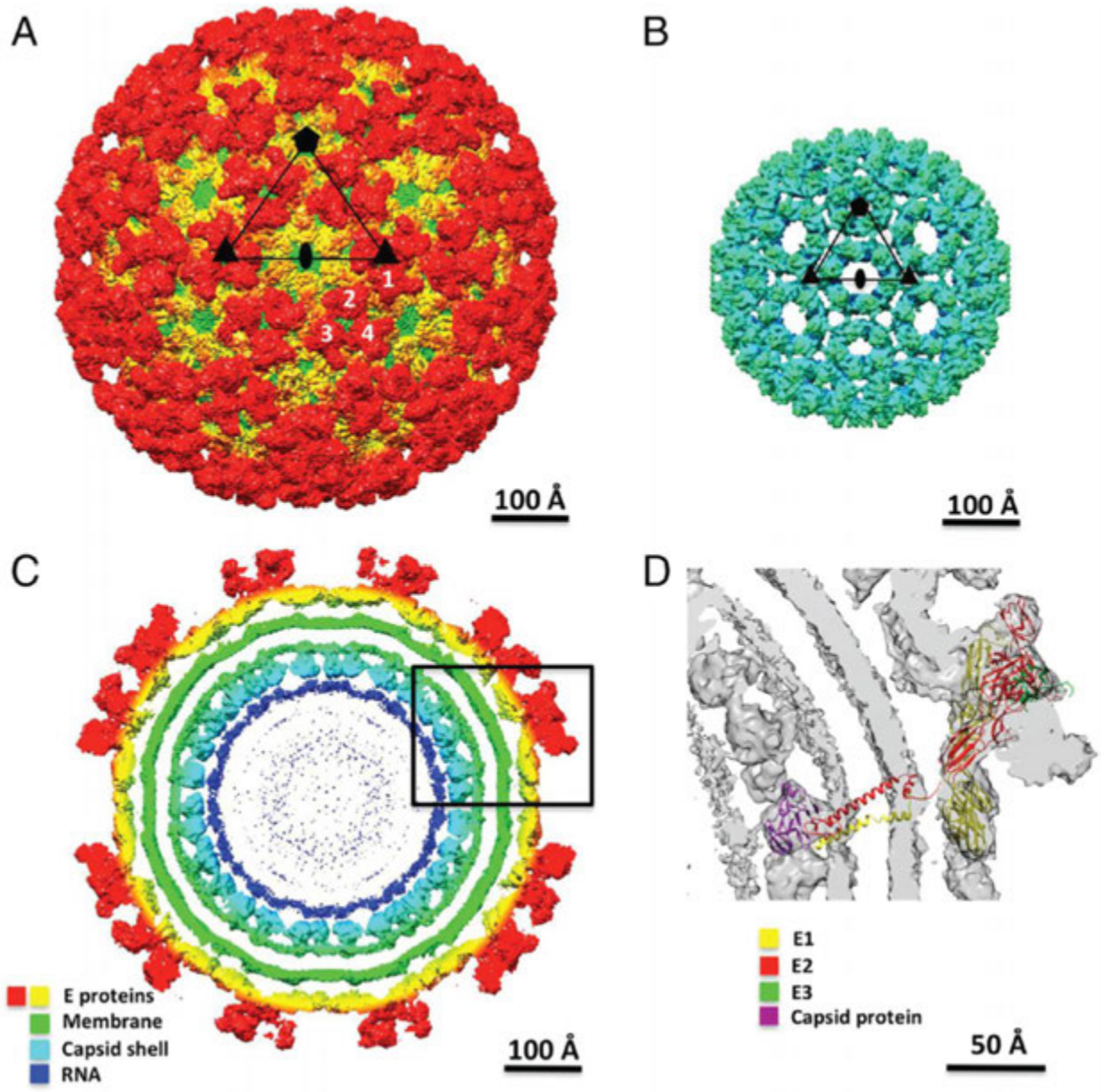

Fig. 7 *T = 4 icosahedral capsid architecture in immature Chikungunya virus (CHIKV).* (A) Electron potential map of immature CHIKV viewed down an icosahedral 2-fold axis. The icosahedral asymmetric unit is drawn as a black triangle and 2-, 3- and 5-fold symmetry elements are represented as oval, triangle and pentagon, respectively. The four subunits of icosahedral asymmetric units are labeled from 1 to 4 in white text. (B) Internal capsid shell of immature CHIKV. (C) Central cross-section of (A) with the represented components colored as labeled. (D) The fitting of E1-E2-E3 and C coordinates into the region outlined in (C). Figure adapted from Yap, M.L., Klose, T., Urakami, A., *et al.*, 2017. Structural studies of Chikungunya virus maturation. Proceedings of the National Academy of Sciences of the United States of America 114, 13703–13707.

Once the virus has been internalized, the low pH of the endosome causes the virion to undergo an irreversible conformational change resulting in the disassociation of E2 from E1 and the formation of E1 trimers. Upon E2 disassociation, the fusion loop then binds to and fuses with the endosomal membrane. Finally, the viral genome is released into the host cytoplasm, where replication of new viral particles can begin (Kielian and Rey, 2006; Lescar *et al.*, 2001; Lu *et al.*, 1999; Marsh and Helenius, 1989).

The presence of the lipid causes some heterogeneity, making it often difficult to crystallize these viruses (Harrison *et al.*, 1992). Instead the structure of many of these larger, lipid containing viruses have been determined by using cryoEM to provide a lower resolution structure of the virus and higher resolution X-ray crystal structures of the components. The latter could then be fitted into the EM structure of the whole virus. A particularly good early example was the determination of the mature dengue virus structure (Kuhn *et al.*, 2002). However, more recently it has been possible to determine the structure of lipid containing viruses to about 3 Å resolution with the development of direct electron counting detectors for electron microscopes and general

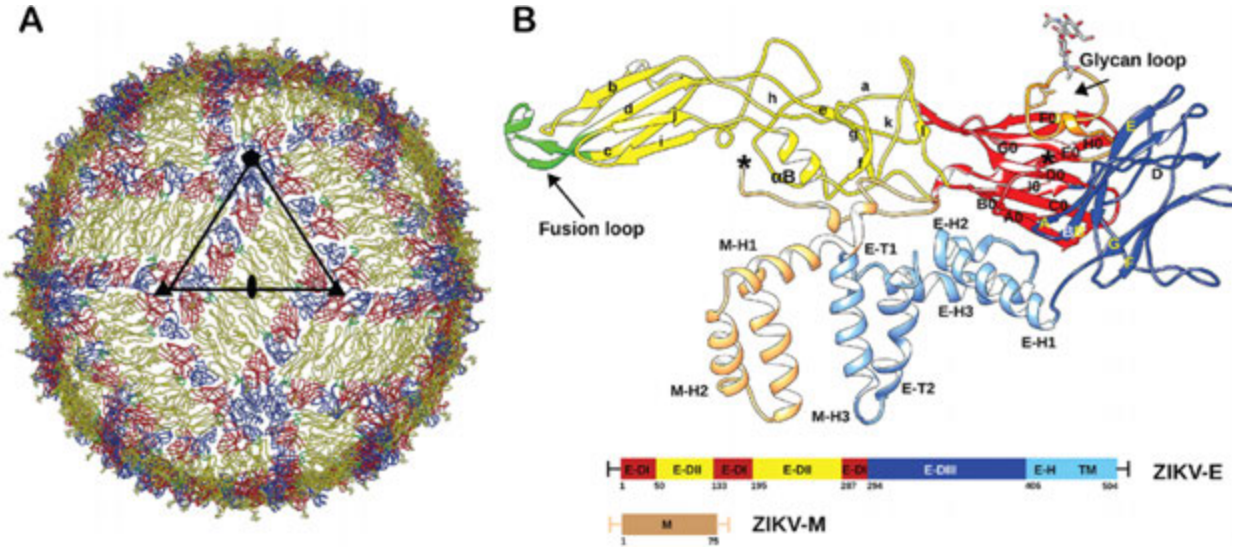

Fig. 8 *T = 3 icosahedral capsid architecture in flaviviruses.* (A) ZIKV capsid formed by herringbone pattern of 6 E-M heterodimers. The E proteins are shown as Cα backbone and the icosahedral asymmetric unit is outlined by a black triangle. The E protein domains E-DI, E-DII, E-DIII and the fusion loop are colored in red, yellow, blue and green, respectively. (B) E-M heterodimer with labeled secondary structure elements. The stem and transmembrane helices (E-H1-3, E-T1-T2) of the E protein and the M protein (M-H1-3) are colored in light blue and light brown, respectively. Residue numbering and domain definitions are shown as a linear peptide. Figure adapted from Sevvana, M., Long, F., Miller, A.S., *et al.*, 2018. Refinement and analysis of the mature Zika virus cryo-EM structure at 3.1 Å resolution. Structure 26, 1169–1177.

improvements to microscopes and techniques. For instance, the structure of mature Zika virus (ZIKV) was determined to about 3 Å resolution using cryoEM (Sevvana *et al.*, 2018), thus also confirming the earlier hybrid techniques.

Similar to other flaviviruses, ZIKV is an enveloped, single-stranded, positive-sense RNA virus (Hasan *et al.*, 2018a; Sevvana *et al.*, 2018; Sirohi *et al.*, 2016). The 11 kb RNA genome is translated into a long polyprotein. It is post-translationally cleaved by host and viral proteases into three structural proteins: pre-membrane (prM), envelope (E), and the capsid (C) as well as seven non-structural proteins (Sirohi and Kuhn, 2017). The E and prM proteins form a protective capsid around the genome. The E-protein mediates the assembly of virus, virus entry and fusion with host membrane and contains putative receptor binding sites. It forms a complex with prM in the endoplasmic reticulum shortly after its synthesis. The E-prM complex is arranged as 60 trimeric spikes (T = 3) on the surface of the immature virus (Mangala Prasad *et al.*, 2017b). The pr domain of prM prevents premature fusion to membranes during virus production. Immature virions undergo pH-induced conformational changes to form the mature virion. Mature virus particles are formed after the pr domain is cleaved from the prM protein by furin in the trans-Golgi network. The icosahedral mature ZIKV consists of 180 copies of E and M proteins arranged in 60 asymmetric units (quasi-T = 3). Each icosahedral asymmetric unit consists of three E-M oligomers. Three parallel dimers from adjacent asymmetric units form a "raft". Therefore, there are 30 rafts arranged in a herringbone pattern on the surface of the mature virus (**Fig. 8**). Flavivirus membrane fusion is initiated by E-protein interaction with a potential receptor molecule and the endocytosis of the virus. This is followed by low pH-induced conformational changes in the endosome, exposure of the fusion loop and its interaction with the endosomal membrane. The E-proteins then rearrange into trimeric structures (fusogenic trimer) leading to the fusion of viral and endosomal membrane and subsequent release of viral genome into the host cytosol.

Though, most of the flavivirus structures have icosahedral symmetry, a recent study has shown that immature Kunjin and Zika viruses are not accurately icosahedral (Therkelsen *et al.*, 2019). Their nucleocapsid core approaches close to one side of the external glycoprotein shell, indicating the presence of a special, unique region that may have been created while interacting with glycoproteins to promote budding from the endoplasmic reticulum (ER) in infected cells and may be required for genome packaging in flaviviruses. Furthermore, on the opposite face on the exterior glycoprotein surface, there is a departure from exact icosahedral symmetry that may reflect the final step in budding. Similar asymmetric properties may have been overlooked in the usual averaging procedures assumed in the study of most essentially icosahedral viruses.

Small Icosahedral ssDNA Viruses

Two examples of small icosahedral ssDNA viruses are ΦX174 (McKenna *et al.*, 1992b), which belongs to the *Microviridae* family, with a single-stranded circular DNA and parvoviruses with a linear genome. Parvoviruses have a 5000 nucleotide ssDNA genome, and typically have a T = 1 icosahedral capsid that is about 260 Å in diameter (**Fig. 9**). Most mammalian parvoviruses have 60 subunits of VP1, VP2, and VP3 in the viral capsid. Many parvovirus structures such as human B19 virus-like particles, canine parvovirus, feline parvovirus, porcine VLP's and murine VLP's were determined to near atomic resolution using X-ray crystallography (Agbandje *et al.*, 1993; Kaufmann *et al.*, 2005; 2008; 2004; Simpson *et al.*, 2000; Simpson *et al.*, 2002; Tsao *et al.*, 1991;

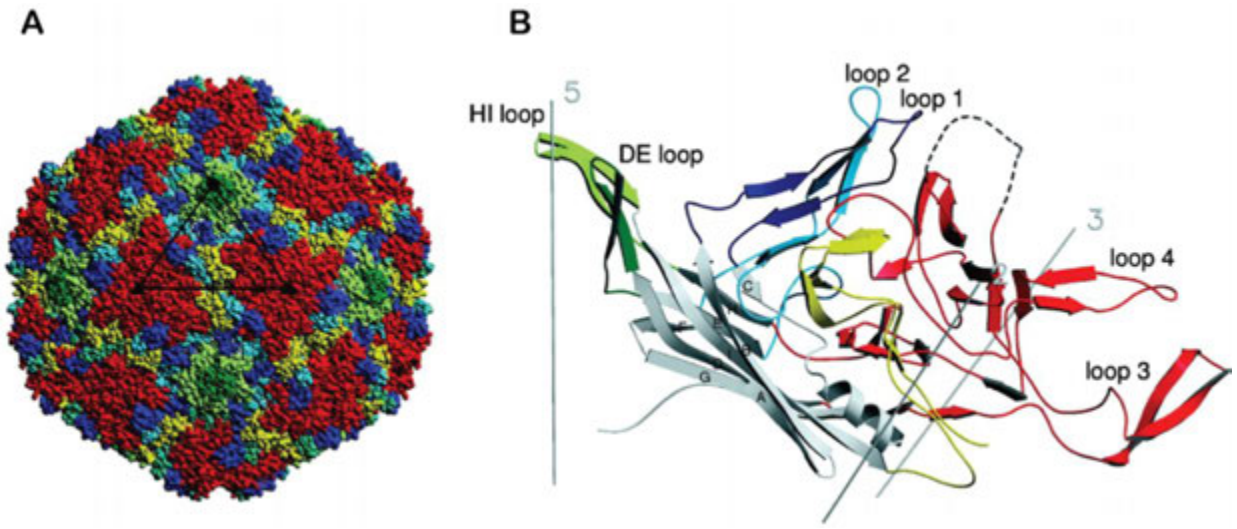

Fig. 9 *Structure of icosahedral (T = 1), human parvovirus B19 VP2 particles.* (A) Surface topography of B19 color-coded according to the ribbon diagram in (B). (B) Secondary structure of parvovirus B19 VP2 showing the jelly-roll β-barrel fold. The surface loops connecting the strands of the β-barrel are colored as follows: dark blue, BC loop; dark green, DE loop; light blue, EF loop; red, GH loop; light green, HI loop; yellow, C-terminal amino acids. Figure adapted from Kaufmann, B., Simpson, A.A., Rossmann, M.G., 2004. The structure of human parvovirus B19. Proceedings of the National Academy of Sciences of the United States of America 101, 11628–11633.

Wu and Rossmann, 1993; Xie and Chapman, 1996). Adeno-associated parvoviruses (AAV-2) are of particular interest because of their potential use in gene therapies (DiMattia *et al.*, 2012; Hueffer *et al.*, 2003b; Lerch *et al.*, 2012; Padron *et al.*, 2005; Walters *et al.*, 2004; Xie *et al.*, 2002). The capsid proteins of parvoviruses have large insertions on their jelly-roll structure. These account for most of the, interactions between neighboring subunits as well as receptor interactions (Chapman and Rossmann, 1996; Tsao *et al.*, 1992). For example, canine and murine parvoviruses can use transferrin (TfR) receptor for binding and infection (Hueffer *et al.*, 2003a). However, the TfR molecule only binds to a few of the 60 available binding sites on the virus thereby inducing asymmetry in the icosahedral virus (Hafenstein *et al.*, 2007).

Bacteriophage ΦX174 is a tailless phage that contains a single-stranded, closed and circular DNA with about 5000 nucleotides (Dokland *et al.*, 1999; 1998; McKenna *et al.*, 1992b). The T = 1 mature icosahedral virus consists of four structural proteins, J, F, G, and H. The capsid of 260 Å diameter is formed by 60 F subunits. The 12 pentameric spikes formed by 60 copies of the G protein rise about 30 Å above the capsid surface and both F and G subunits have jelly-roll folds. There are 12 copies of the H protein, which form a tail tube similar to the tail tube of the common tailed dsDNA phage during infection for the insertion of DNA into the host cell (Sun *et al.*, 2013a). Structural studies shed light on the assembly of ΦX174, which proceeds via virally encoded scaffolding proteins (B and D) both of which are discarded when assembly is complete (**Fig. 10**).

Large Icosahedral dsDNA Viruses

Nucleocytoplasmic large dsDNA viruses (NCDLVs) infecting a wide variety of organisms' span members of the *Asfaviridae*, *Ascoviridae*, *Marseilleviridae*, *Mimiviridae*, *Pandoraviridae*, *Phycodnaviridae*, *Iridoviridae*, and *Poxviridae*. Most NCDLVs are roughly icosahedral in shape. However, members of Poxviruses, Ascoviruses, and Pandoraviruses are pleomorphic. CryoEM reconstructions of several icosahedral NCDLVs infer similar capsid organization with large T numbers (Andreani *et al.*, 2017; Cherrier *et al.*, 2009; Klose *et al.*, 2016; Okamoto *et al.*, 2018; Reteno *et al.*, 2015; Xiao *et al.*, 2017; Yan *et al.*, 2005; 2000; 2009; Zhang, *et al.*, 2011b). An inner membrane that separates the genome from the external capsid is observed in most NCDLVs except Faustoviruses, which have an internal capsid.

The capsomeres in icosahedral NCDLVs are arranged in trisymmetrons (20 triangular arrays) and pentasymmetrons (12 pentagonal arrays) (**Fig. 11**) (Klose and Rossmann, 2014). The capsid is assembled from pseudo-hexagonal, closely arranged trimeric capsomers where each subunit is formed by a double jelly-roll fold protein. Each capsomere has a diameter of about 75–85 Å and a height of 75 Å. Because of their size range, the capsomeres can have two orientations related by a six-fold rotation. First, the capsomeres within each trisymmetron can have the same orientation or, second, they can be rotated by 60° relative to neighboring capsomeres which creates cleavage planes between neighboring trisymmetrons (Wrigley, 1969). The five-fold vertices are formed by a pentameric single jelly-roll fold protein. The pentasymmetrons have a constant size and consist of 30 trimeric double jelly-roll capsomeres and one pentameric single jelly-roll protein. It was postulated that this constant size is ideal to relieve the strain induced around the five-fold axis for icosahedral NCDLVs with large diameters.

Another characteristic feature of NCDLVs is the presence of minor capsid proteins found underneath the major capsid proteins towards the inside of the virus capsid as can be seen in high resolution structures of faustovirus (Klose *et al.*, 2016), Chilo iridescent virus (CIV) and Paramecium bursaria chlorella virus 1 (PBCV-1) (Zhang *et al.*, 2011b). The minor capsid proteins

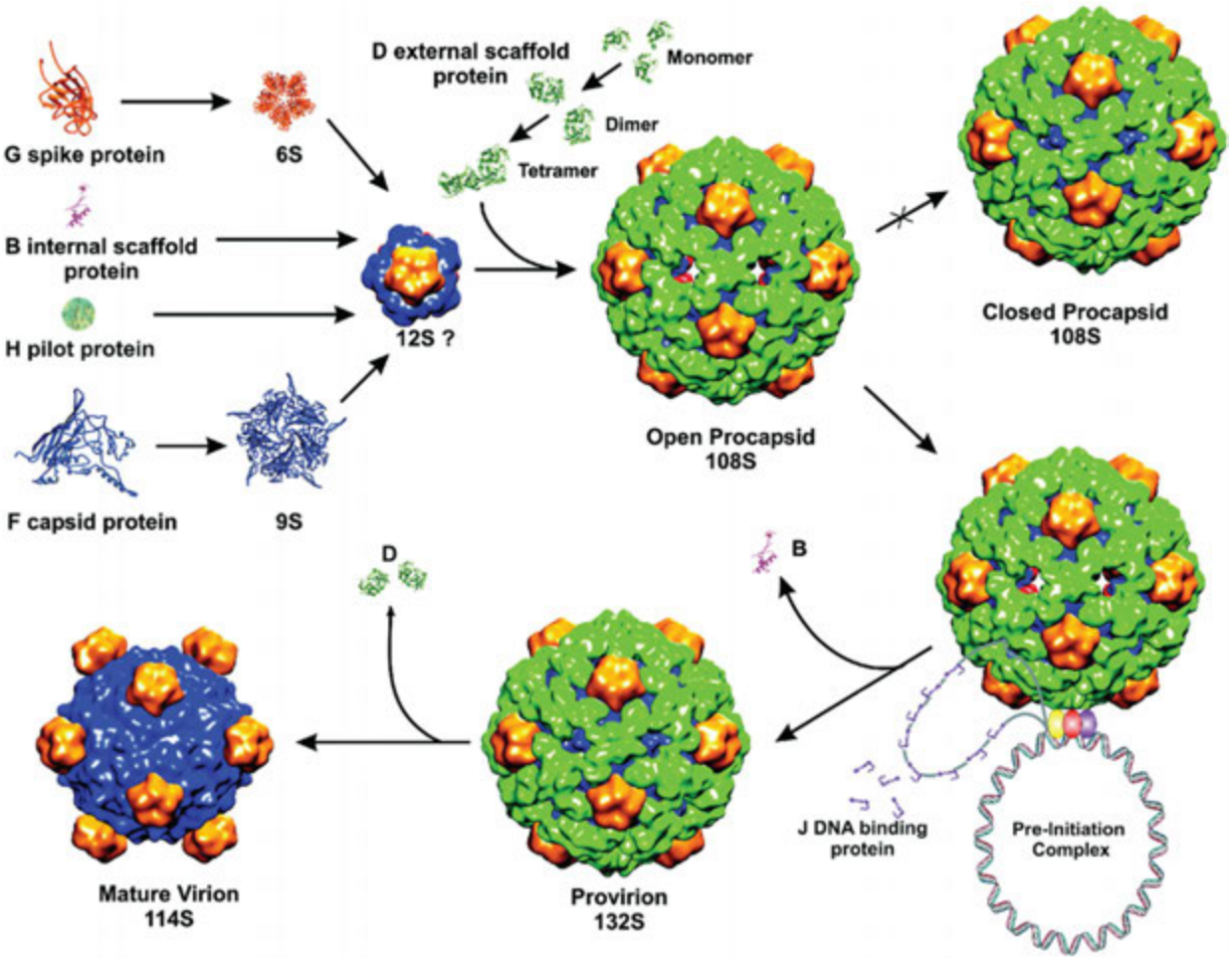

Fig. 10 *Capsid assembly pathway in Microviridae.* The first intermediates are formed by pentameric assemblies of F capsid in 9 s and G spike proteins in 6 s particles. With the help of internal and external scaffolding proteins (B, D, and H), these are later assembled into an empty protein capsid shell called the procapsid. The ssDNA is concurrently synthesized and packaged into the procapsid in complex with DNA-binding protein J, therefore forming the mature virion. Figure adapted from Bernal, R.A., Hafenstein, S., Olson, N.H., *et al.*, 2003. Structural studies of bacteriophage a3 assembly. Journal of Molecular Biology 325, 11–24.

support the outer capsid shell, connect the neighboring trisymmetrons and/or connect the outer capsid to the inner membrane. Some NCLDVs like mimivirus (Xiao *et al.*, 2005,2009; Klose *et al.*, 2010) and PBCV-1 also have a unique vertex where the membrane becomes separated from the capsid. Many NCDLVs have decoration proteins and surface fibers attached to the viral capsid. Mimivirus is covered by a dense layer of surface fibers except near the unique vertex called the stargate, which is tightly associated with the adjacent capsomers.

One of the most studied NCDLVs is PBCV-1, where a recent near-atomic resolution structure identified several of the minor capsid proteins and cast light on the capsid assembly pathway (Fang *et al.*, 2019). PBCV-1 has a 330 kbp genome that encodes 416 proteins, of which 149 different proteins were identified in the mature virion. The nucleocapsid is surrounded by a membrane, which in turn is surrounded by a roughly icosahedral capsid shell giving it an overall size of 190 nm. It also has a unique vertex with a spike structure required for host entry. The outer shell consists of the MCP (Vp54) and the penton protein P1. Each tri- and pentasymmetron contains 66 (MCP) and 31 (30 MCP + 1 P1) capsid proteins respectively, arranged in an icosahedral lattice with a triangulation number of T = 169d. The high resolution cryoEM structure led to the identification of the penton protein, P1, and 13 minor capsid proteins (named P2 through P14). The minor capsid proteins form a hexagonal network below the outer capsid shell. Most of the minor capsid proteins are primarily located at the interface between neighboring capsomeres. The size of the viral capsid is determined by a tape-measure protein, P2, and the zip protein P11 which plays a crucial role in cementing neighboring symmetrons (**Fig. 11**).

Faustovirus, on the other hand uses two concentric capsid shells to protect its genome (Klose *et al.*, 2016). Consequently, two distinct types of particles were observed in cryoEM images of purified virus. The larger icosahedral particle constituting the outer and inner capsids measured about 2600 Å in diameter, whereas the smaller particles with only an inner capsid measured between 1600–1900 Å. The capsomere on the outer capsid is arranged with a triangulation number of T = 277 (h = 7 and k = 12). The

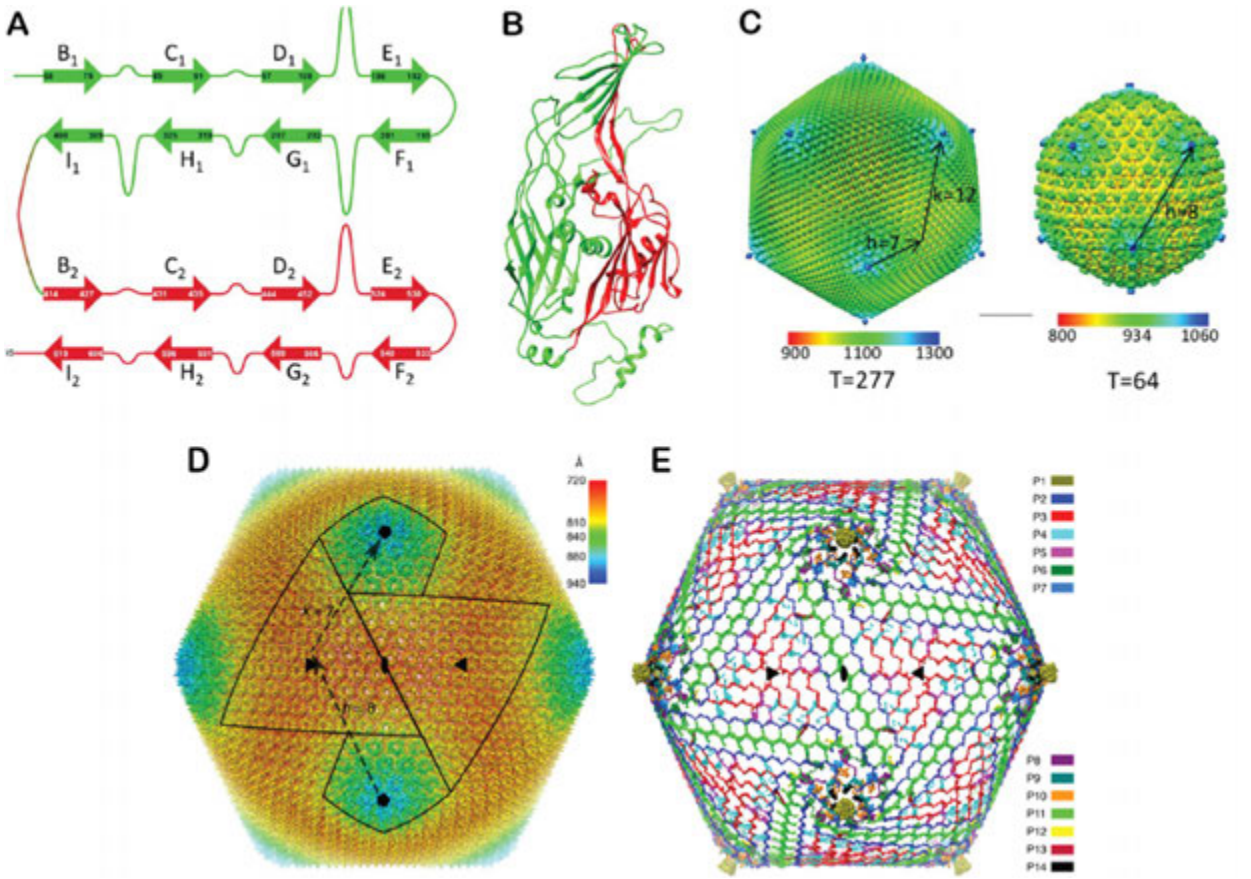

Fig. 11 *Capsid organization in NCLDVs.* (A) Topology plot and (B) cartoon representation of the double jelly-roll fold of the MCP in faustovirus. The secondary structure elements are labeled. (C) T = 277 and T = 64 capsid organizations of faustovirus external and internal capsid shells respectively. (D) CryoEM map of the overall structure of the PBCV-1 capsid. The boundaries of neighboring trisymmetrons and pentasymmetrons are outlined in black. (E) The cryoEM density of the minor capsid proteins and the penton proteins after subtracting the density for MCP. Each protein is depicted in a different color as indicated to the right of the image. The h and k indices are labeled on the icosahedron in panels (C) and (D). The radial color gradient scale bar is in Å. Figures adapted from: (A), (B), and (C) Klose, T., Reteno, D.G., Benamar, S., *et al.*, 2016. Structure of faustovirus, a large dsDNA virus. Proceedings of the National Academy of Sciences of the United States of America 113, 6206–6211. (D) and (E) Fang, Q.L., Zhu, D.J., Agarkova, I., *et al.*, 2019. Near-atomic structure of a giant virus. Nature Communications 10, 388.

icosahedral asymmetric unit is composed of one penton protein and 276 MCPs. Each tri- and pentasymmetron of the outer capsid shell contains 360 (MCP) and 31 (30 MCP + 1 P1) proteins, respectively (Fig. 11).

Tailed Bacteriophages

The structure of dsDNA bacteriophages is unique in that most of them have tails attached to a special pentameric vertex of their capsids (Fokine and Rossmann, 2014). The unique vertex in most bacteriophages is occupied by a dodecameric portal protein, which forms a channel for genome packaging during virion assembly and subsequent exit during infection. Tailed bacteriophages belong to the order *Caudovirales*, which are further divided into three families: *Myoviridae* (i.e., T4 and ϕ92 have long, straight contractile tails), *Siphoviridae* (i.e., lambda and HK97) have long flexible non-contractile tails, and *Podoviridae* (i.e., T7) have short non-contractile tails.

Capsid heads of tailed bacteriophages come in a large variety of sizes and range in diameter from about 400–1700 Å with either icosahedral or prolate heads. The capsid contains linear double-stranded DNA packaged to a high density of about 500 g/l and exerts an internal pressure of tens of atmospheres on the capsid walls. The capsid precursor, called the procapsid or prohead, is assembled at the portal vertex by copolymerization of the MCP and scaffolding proteins. The scaffolding protein regulates the correct geometry of the capsid. The prohead consists of the portal protein, the outer MCP, the internal scaffolding protein and in some cases other minor capsid proteins. The scaffolding protein and the internal core are degraded in the procapsids of many phages by head maturation proteases, therefore making space for DNA packaging. The genome is packaged by a DNA translocation motor via the portal vertex and is ATP-dependent. DNA packaging is followed by large structural rearrangements, an increase in internal volume and attachment of other proteins (for example decoration proteins) to form the mature capsid. After DNA packaging, the DNA translocation motor is removed and the capsid assembly is completed by sealing the portal gate, the

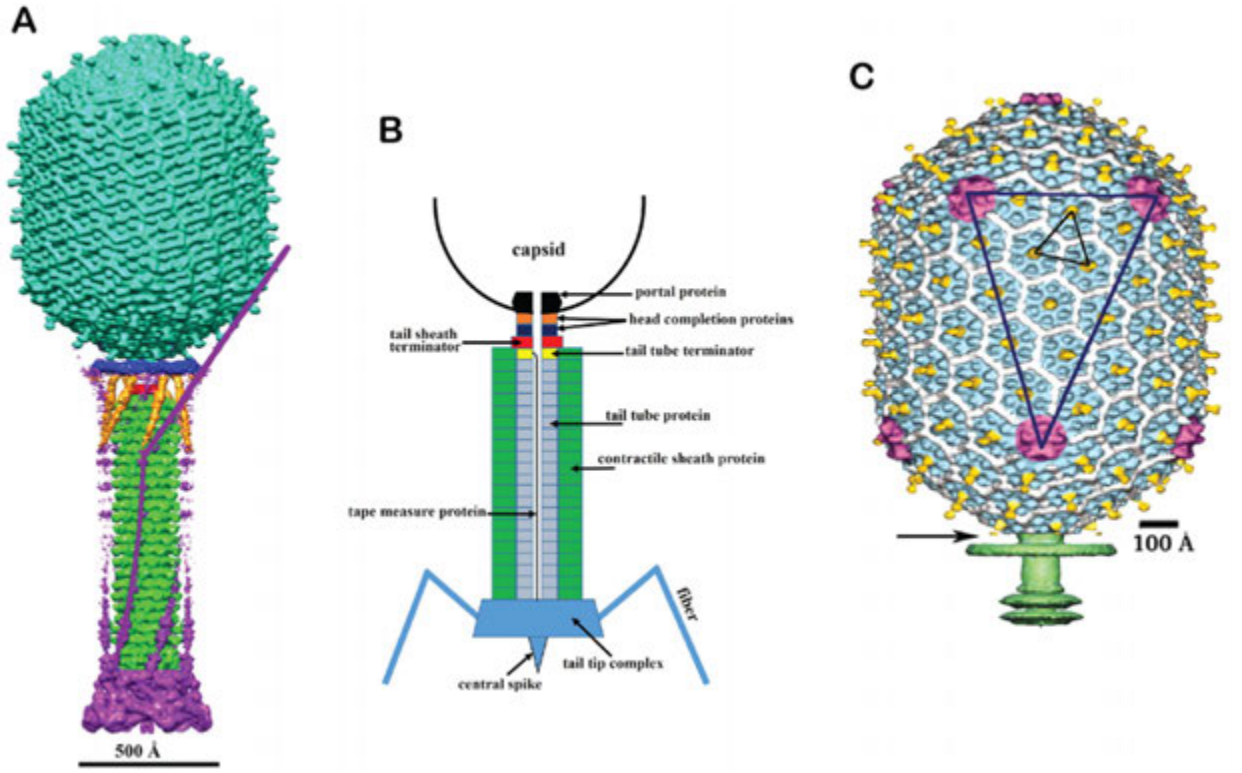

Fig. 12 *Structure of T4 bacteriophage (T = 13 laevo, Q = 20, h_1 = 3, k_1 = 1, h_2 = 4, and k_2 = 2).* (A) CryoEM reconstruction of the T4 capsid and extended tail. The capsid, tail sheath, phage collar and whiskers are colored cyan, green, blue and gold respectively. The density corresponding to the baseplate and long tail fibers is colored magenta. (B) Schematic representation of the bacteriophage T4 neck and tail components. (C) CryoEM reconstruction of the T4 head viewed perpendicular to the 5-fold axis. The capsid proteins, gp23 and gp24, are shown in blue and magenta respectively, *soc* is in white, *hoc* is in yellow, and the tail is in green. The facet triangle is shown in blue and the basic triangles in black. Adapted from: (A) Fokine, A., Zhang, Z., Kanamaru, S., *et al.*, 2013. The molecular architecture of the bacteriophage T4 neck. Journal of Molecular Biology 425, 1731–1744. (B) Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4, e28281. (C) Fokine, A., Chipman, P.R., Leiman, P.G., *et al.*, 2004. Molecular architecture of the prolate head of bacteriophage T4. Proceedings of the National Academy of Sciences of the United States of America 101, 6003–6008.

attachment of head completion proteins and the assembly of the tail. Bacteriophage tails are designed to recognize host cells, penetrate the cell envelope and deliver the genome. Their length varies from ~100 Å in *Podoviridae* to ~8000 Å in some *Siphoviridae*. Most tails of the members of *Myoviridae* and *Siphoviridae* consist of a tail tip complex for host cell recognition, a tail tube, which makes a conduit for genomic DNA with a contractile sheath around it in some phages and the terminator proteins near the tail-head interface (**Fig. 12**).

One of the most studied bacteriophage structures is of T4 bacteriophage, which infects *E. coli*. The dsDNA genome encodes about 40 structural proteins. The mature virus consists of an 1150 Å-long and 850 Å-wide prolate head encapsidating the genomic DNA. The head is attached to a 1000 Å-long and 210 Å-diameter cylindrical contractile tail and is terminated with a 460 Å-diameter baseplate (Fokine *et al.*, 2004; Kostyuchenko *et al.*, 2003) (**Fig. 12**). The baseplate is associated with six 1450 Å-long fibers (Aksyuk *et al.*, 2009; Kostyuchenko *et al.*, 2005). The distal ends of the long tail fibers (made of gp37) recognize the host cell (Bartual *et al.*, 2010) and then transfer the signal to the baseplate to unfold the short tail fibers (gp12) (Thomassen *et al.*, 2003; van Raaij *et al.*, 2001). The short tail fibers subsequently bind to the lipopolysaccharides on the surface of the E.*coli* host. The unfolding of the short tail fibers results in large conformational changes of the dome-shaped baseplate to a star-shape as the tail sheath (gp18) starts to contract (Aksyuk *et al.*, 2009; Aksyuk and Rossmann, 2011; Crowther *et al.*, 1977; Kostyuchenko *et al.*, 2003; Leiman *et al.*, 2004). The tip of the tail tube (gp5) punctures the cell, enters the periplasm and digests the cell wall (Browning *et al.*, 2012; Kanamaru *et al.*, 2002). The genome is then ejected into the host's cytoplasm by the tail tube. In phage T4 the pentameric protein forming 11 of the 12 vertices (gp24) has an HK97 fold, which has a high sequence similarity to T4 MCP (gp23) (Jiang *et al.*, 2003; Morais *et al.*, 2005).

Conclusion

Viruses are complex organisms that build extraordinary structures out of a limited number of building blocks. It is remarkable that the principles established by Watson and Crick (Watson and Crick, 1953b) and expanded by Caspar and Klug (Caspar and Klug,

1962) in the 1960s have been validated and still hold true for most virus structures discovered to date. Furthermore, it is worth noting that the large variety of virus structures that have been described in the scientific literature so far utilize a fairly small number of protein folds, most of which have been covered in this review. Improvements in sequencing technologies and less stringent isolation techniques have led to an immense increase in the number of viruses that have been described in the literature and it will be interesting to see if this increase eventually correlates with the discovery of new folds or new assemblies. It is already becoming clear that the future of structural virology will be dominated by cryoEM. Most of the recently determined virus structures were solved using this technique and one would expect this number to continue increasing. X-Ray crystallography will still play an important role, but will likely be confined to solving the structures of smaller components of virions or assemblies. In addition, cryoEM and cryoET will allow us to look into deviations of the strict symmetry requirements imposed onto most viral structures to date and help us to learn more about assembly pathways and unique structural components that might otherwise be overlooked. This might show us that viruses are not that perfect after all and may lead to important biological discoveries and potentially new therapeutic approaches.

Acknowledgments

MGR would like to thank all his many collaborators, post-docs, graduate students and colleagues who made it possible to accomplish many of the findings reported here. MGR is also grateful to the National Science Foundation, the National Institutes of Health and Purdue University for their generous support over many years. We thank Sheryl Kelly for her help in preparing the paper. MS and TK would like to thank MGR for the opportunity to co-write this article. This was the last paper MGR worked on and submitted a month before his tragic demise. It was an exciting time-travel through several decades of the history of structural virology.

References

Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., *et al.*, 1980. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 286, 33–39.
Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., *et al.*, 1981. A description of techniques used in the structure determination of southern bean mosaic virus at 2.8 Å resolution. Acta Crystallographica B 37, 2002–2018.
Ackermann, H.-W., 1999. Tailed bacteriophages: The order Caudovirales. Advances in Virus Research 51, 135–201.
Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W., 1984. Cryo-electron microscopy of viruses. Nature 308, 32–36.
Agbandje, M., McKenna, R., Rossmann, M.G., Strassheim, M.L., Parrish, C.R., 1993. Structure determination of feline panleukopenia virus empty particles. Proteins 16, 155–171.
Åkervall, K., Strandberg, B., Rossmann, M.G., *et al.*, 1972. X-ray diffraction studies of the structure of satellite tobacco necrosis virus. Cold Spring Harbor Symposia on Quantitative Biology 36, 469–483.
Aksyuk, A.A., Leiman, P.G., Kurochkina, L.P., *et al.*, 2009. The tail sheath structure of bacteriophage T4: A molecular machine for infecting bacteria. EMBO Journal 28, 821–829.
Aksyuk, A.A., Rossmann, M.G., 2011. Bacteriophage assembly. Viruses 3, 172–203.
Andreani, J., Khalil, J.Y.B., Sevvana, M., *et al.*, 2017. Pacmanvirus, a new giant icosahedral virus at the crossroads between *Asfarviridae* and Faustoviruses. Journal of Virology 91, e00212–e00217.
Arnold, E., Rossmann, M.G., 1986. Effect of errors, redundancy, and solvent content in the molecular replacement procedure for the structure determination of biological macromolecules. Proceedings of the National Academy of Sciences of the United States of America 83, 5489–5493.
Arnold, E., Vriend, G., Luo, M., *et al.*, 1987. The structure determination of a common cold virus, human rhinovirus 14. Acta Crystallographica Section A 43, 346–361.
Bartual, S.G., Otero, J.M., Garcia-Doval, C., *et al.*, 2010. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proceedings of the National Academy of Sciences of the United States of America 107, 20287–20292.
Battisti, A.J., Yoder, J.D., Plevka, P., *et al.*, 2012. Cryo-electron tomography of rubella virus. Journal of Virology 86, 11078–11085.
Bawden, F.C., Pirie, N.W., 1938. Crystalline preparations of tomato bushy stunt virus. British Journal of Experimental Pathology 19, 251–263.
Bernal, J.D., Fankuchen, I., 1941a. X-ray and crystallographic studies of plant virus preparations: I. Introduction and preparation of specimens. II. Modes of aggregation of the virus particles. III. (1) The structure of the particles and (2) biological implications. Journal of General Physiology 25, 111–165.
Bernal, J.D., Fankuchen, I., 1941b. X-ray and crystallographic studies of plant virus preparations. III. Journal of General Physiology 25, 147–165.
Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R., Klug, A., 1978. Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. Nature 276, 362–368.
Brenner, S., Horne, R.W., 1959. A negative staining method for high resolution electron microscopy of viruses. Biochimica et Biophysica Acta 34, 103–110.
Browning, C., Shneider, M.M., Bowman, V.D., Schwarzer, D., Leiman, P.G., 2012. Phage pierces the host cell membrane with the iron-loaded spike. Structure 20, 326–339.
Buehner, M., Ford, G.C., Moras, D., Olsen, K.W., Rossmann, M.G., 1974. Structure determination of crystalline lobster D-glyceraldehyde-3-phosphate dehydrogenase. Journal of Molecular Biology 82, 563–585.
Butler, P.J., Klug, A., 1972. Assembly of tobacco mosaic virus in vitro: Effect of state of polymerization of the protein component. Proceedings of the National Academy of Sciences of the United States of America 69, 2950–2953.
Carrillo-Tripp, M., Shepherd, C.M., Borelli, I.A., *et al.*, 2009. VIPERdb(2): An enhanced and web API enabled relational database for structural virology. Nucleic Acids Research 37, D436–D442.
Caspar, D.L.D., 1956. Structure of bushy stunt virus. Nature 177, 475–476.
Caspar, D.L.D., Klug, A., 1962. Physical principles in the construction of regular viruses. Cold Spring Harbor Symposia on Quantitative Biology 27, 1–24.
Champness, J.N., Bloomer, A.C., Bricogne, G., Butler, P.G., Klug, A., 1976. The structure of the protein disk of tobacco mosaic virus to 5 Å resolution. Nature 259, 20–24.
Chapman, M.S., Liljas, L., 2003. Structural folds of viral proteins. Advances in Protein Chemistry 64, 125–196.
Chapman, M.S., Rossmann, M.G., 1996. Structural refinement of the DNA-containing capsid of canine parvovirus using RSRef, a resolution-dependent stereochemically restrained real-space refinement method. Acta Crystallographica Section D: Biological Crystallography 52, 129–142.
Chelvanayagam, G., Heringa, J., Argos, P., 1992. Anatomy and evolution of proteins displaying the viral capsid jellyroll topology. Journal of Molecular Biology 228, 220–242.
Cheng, S., Brooks III, C.L., 2013. Viral capsid proteins are segregated in structural fold space. PLOS Computational Biology 9, e1002905.

Cheng, R.H., Kuhn, R.J., Olson, N.H., *et al.*, 1995. Nucleocapsid and glycoprotein organization in an enveloped virus. Cell 80, 621–630.
Chen, L.H., Wang, M., Zhu, D.J., *et al.*, 2018. Implication for alphavirus host-cell entry and assembly indicated by a 3.5 Å resolution cryo-EM structure. Nature Communications 9, 5326.
Cherrier, M.V., Kostyuchenko, V.A., Xiao, C., *et al.*, 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. Proceedings of the National Academy of Sciences of the United States of America 106, 11085–11089.
Choi, H.K., Tong, L., Minor, W., *et al.*, 1991a. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature 354, 37–43.
Choi, H.K., Tong, L., Minor, W., *et al.*, 1991b. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature 354, 37–43.
Conway, J.F., Cheng, N., Zlotnick, A., *et al.*, 1997. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. Nature 386, 91–94.
Crick, F.H.C., Watson, J.D., 1956. Structure of small viruses. Nature 177, 473–475.
Crick, F.H.C., Watson, J.D., 1957. Virus structure: General principles. In: Wolstenholme, G.E.W., Millar, E.C.P. (Eds.), Ciba Foundation Symposium on "The Nature of Viruses". Chichester, UK: John Wiley & Sons, Ltd, pp. 5–13.
Crowther, R.A., Amos, L.A., Finch, J.T., DeRosier, D.J., Klug, A., 1970. Three-dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. Nature 226, 421–425.
Crowther, R.A., Lenk, E.V., Kikuchi, Y., King, J., 1977. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. Journal of Molecular Biology 116, 489–523.
DiMattia, M.A., Nam, H.J., Van Vliet, K., *et al.*, 2012. Structural insight into the unique properties of adeno-associated virus serotype 9. Journal of Virology 86, 6947–6958.
Dokland, T., Bernal, R.A., Burch, A., *et al.*, 1999. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus ϕX174. Journal of Molecular Biology 288, 595–608.
Dokland, T., McKenna, R., Sherman, D.M., *et al.*, 1998. Structure determination of the ϕX174 closed procapsid. Acta Crystallographica Section D: Biological Crystallography 54, 878–890.
Dubochet, J., Adrian, M., Chang, J.J., *et al.*, 1988. Cryo-electron microscopy of vitrified specimens. Quarterly Reviews of Biophysics 21, 129–228.
Enard, D., Cai, L., Gwennap, C., Petrov, D.A., 2016. Viruses are a dominant driver of protein adaptation in mammals. eLife 5, e12469.
Erickson, J.W., Frankenberger, E.A., Rossmann, M.G., *et al.*, 1983. Crystallization of a common cold virus, human rhinovirus 14: "Isomorphism" with poliovirus crystals. Proceedings of the National Academy of Sciences of the United States of America 80, 931–934.
Fang, Q.L., Zhu, D.J., Agarkova, I., *et al.*, 2019. Near-atomic structure of a giant virus. Nature Communications 10, 388.
Filman, D.J., Syed, R., Chow, M., *et al.*, 1989. Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus. EMBO Journal 18, 1567–1579.
Fokine, A., Chipman, P.R., Leiman, P.G., *et al.*, 2004. Molecular architecture of the prolate head of bacteriophage T4. Proceedings of the National Academy of Sciences of the United States of America 101, 6003–6008.
Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4, e28281.
Frank, J., 1992. Electron Tomography: Methods for Three-dimensional Visualization of Structures in the Cell. New York: Springer Science + Business Media LLC.
Franklin, R.E., Holmes, K.C., 1958. Tobacco mosaic virus: Spplication of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution. Acta Crystallographica 11, 213–220.
Franklin, R.E., Klug, A., Holmes, K.C., 1957. X-ray diffraction studies of the structure and morphology of tobacco mosaic virus. In: Wolstenholme, G.E.W., Millar, E.C.P. (Eds.), Ciba Foundation Symposium on the Nature of Viruses. J. & A. Churchill, London: Ciba Foundation, pp. 39–55.
Fuller, S.D., 1987. The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. Cell 48, 923–934.
Gamble, T.R., Yoo, S., Vajdos, F.F., *et al.*, 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science 278, 849–853.
Ge, P., Zhou, Z.H., 2011. Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. Proceedings of the National Academy of Sciences of the United States of America 108, 9637–9642.
Gibbons, D.L., Kielian, M., 2002. Molecular dissection of the Semliki Forest virus homotrimer reveals two functionally distinct regions of the fusion protein. Journal of Virology 76, 1194–1205.
Greve, J.M., Davis, G., Meyer, A.M., *et al.*, 1989. The major human rhinovirus receptor is ICAM-1. Cell 56, 839–847.
Haas, D.J., Rossmann, M.G., 1970. Crystallographic studies on lactate dehydrogenase at -75°C. Acta Crystallographica Section B: Structural Science 26, 998–1004.
Hafenstein, S., Palermo, L.M., Kostyuchenko, V.A., *et al.*, 2007. Asymmetric binding of transferrin receptor to parvovirus capsids. Proceedings of the National Academy of Sciences of the United States of America 104, 6585–6589.
Harrison, S., 2001. Principles of virus structure. In: Fields Virology. Philadelphia: Lippincott Williams & Wilkins Publishers.
Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G., 1978. Tomato bushy stunt virus at 2.9 Å resolution. Nature 276, 368–373.
Harrison, S.C., Strong, R.K., Schlesinger, S., Schlesinger, M.J., 1992. Crystallization of Sindbis virus and its nucleocapsid. Journal of Molecular Biology 226, 277–280.
Hasan, S.S., Sevvana, M., Kuhn, R.J., Rossmann, M.G., 2018a. Structural biology of Zika virus and other flaviviruses. Nature Structural & Molecular Biology 25, 13–20.
Hasan, S.S., Sun, C., Kim, A.S., *et al.*, 2018b. Cryo-EM structures of Eastern equine encephalitis virus reveal mechanisms of virus disassembly and antibody neutralization. Cell Reports 25, 3136–3147.e5.
Hayden, F.G., Herrington, D.T., Coats, T.L., *et al.*, 2003. Efficacy and safety of oral pleconaril for treatment of picornavirus colds in adults: Results of two double-blind, randomized, placebo-controlled trials. Clinical Infectious Diseases 36, 1523–1532.
Helgstrand, C., Wikoff, W.R., Duda, R.L., *et al.*, 2003. The refined structure of a protein catenane: The HK97 bacteriophage capsid at 3.44 Å resolution. Journal of Molecular Biology 334, 885–899.
Hogle, J.M., Chow, M., Filman, D.J., 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. Science 229, 1358–1365.
Holmes, K.C., Stubbs, G.J., Mandelkow, E., Gallwitz, U., 1975. Structure of tobacco mosaic virus at 6.7 Å resolution. Nature 254, 192–196.
Horne, R.W., Brenner, S., Waterson, A.P., Wildy, P., 1959. The icosahedral form of an adenovirus. Journal of Molecular Biology 1, 84–86.
Hueffer, K., Govindasamy, L., Agbandje-McKenna, M., Parrish, C.R., 2003a. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. Journal of Virology 77, 10099–10105.
Hueffer, K., Parker, J.S.L., Weichert, W.S., *et al.*, 2003b. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. Journal of Virology 77, 1718–1726.
Jiang, W., Li, Z., Zhang, Z., *et al.*, 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. Nature Structural Biology 10, 131–135.
Johnson, J.E., Chiu, W., 2000. Structures of virus and virus-like particles. Current Opinion in Structural Biology 10, 229–235.
Johnson, J.E., Speir, J.A., 1997. Quasi-equivalent viruses: A paradigm for protein assemblies. Journal of Molecular Biology 269, 665–675.
Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., *et al.*, 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature 415, 553–557.
Kaufmann, B., Baxa, U., Chipman, P.R., *et al.*, 2005. Parvovirus B19 does not bind to membrane-associated globoside in vitro. Virology 332, 189–198.
Kaufmann, B., Chipman, P.R., Kostyuchenko, V.A., Modrow, S., Rossmann, M.G., 2008. Visualization of the externalized VP2 N-termini of infectious human parvovirus B19. Journal of Virology 82, 7306–7312.

Kaufmann, B., Plevka, P., Kuhn, R.J., Rossmann, M.G., 2010. Crystallization and preliminary X-ray diffraction analysis of West Nile virus. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 66, 558–562.
Kaufmann, B., Simpson, A.A., Rossmann, M.G., 2004. The structure of human parvovirus B19. Proceedings of the National Academy of Sciences of the United States of America 101, 11628–11633.
Kielian, M., Rey, F.A., 2006. Virus membrane-fusion proteins: More than one way to make a hairpin. Nature Reviews Microbiology 4, 67–76.
Klose, T., Kuznetsov, Y.G., Xiao, C., *et al*., 2010. The three-dimensional structure of Mimivirus. Intervirology 53, 268–273.
Klose, T., Reteno, D.G., Benamar, S., *et al*., 2016. Structure of faustovirus, a large dsDNA virus. Proceedings of the National Academy of Sciences of the United States of America 113, 6206–6211.
Klose, T., Rossmann, M.G., 2014. Structure of large dsDNA viruses. Biological Chemistry 395, 711–719.
Klug, A., Caspar, D.L., 1960. The structure of small viruses. Advances in Virus Research 7, 225–325.
Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., *et al*., 2005. The tail structure of bacteriophage T4 and its mechanism of contraction. Nature Structural & Molecular Biology 12, 810–813.
Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., *et al*., 2003. Three-dimensional structure of bacteriophage T4 baseplate. Nature Structural Biology 10, 688–693.
Kuhn, R.J., 2007. *Togaviridae*: The viruses and their replication. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, Fifth Lippincott Williams & Wilkins, pp. 1001–1021.
Kuhn, R.J., Zhang, W., Rossmann, M.G., *et al*., 2002. Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. Cell 108, 717–725.
Lee, K.K., Johnson, J.E., 2003. Complementary approaches to structure determination of icosahedral viruses. Current Opinion in Structural Biology 13, 558–569.
Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., Rossmann, M.G., 2004. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell 118, 419–429.
Lerch, T.F., O'Donnell, J.K., Meyer, N.L., *et al*., 2012. Structure of AAV-DJ, a retargeted gene therapy vector: Cryo-electron microscopy at 4.5 Å resolution. Structure 20, 1310–1320.
Lescar, J., Roussel, A., Wein, M.W., *et al*., 2001. The fusion glycoprotein shell of Semliki Forest virus: An icosahedral assembly primed for fusogenic activation at endosomal pH. Cell 105, 137–148.
Li, L., Jose, J., Xiang, Y., Kuhn, R.J., Rossmann, M.G., 2010. Structural changes of envelope proteins during alphavirus fusion. Nature 468, 705–708.
Lucic, V., F, F., Baumeister, W., 2005. Structural studies by electron tomography: From cells to molecules. Annual Review of Biochemistry 74, 833–865.
Lu, Y.E., Cassese, T., Kielian, M., 1999. The cholesterol requirement for Sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. Journal of Virology 73, 4272–4278.
Mangala Prasad, V., Fokine, A., Battisti, A.J., *et al*., 2013. Rubella virus capsid protein structure and its role in virus assembly and infection. Proceedings of the National Academy of Sciences of the United States of America 110, 20105–20110.
Mangala Prasad, V., Klose, T., Rossmann, M.G., 2017a. Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus. PLOS Pathogens 13, e1006377.
Mangala Prasad, V., Miller, A.S., Klose, T., *et al*., 2017b. Structure of the immature Zika virus at 9 Å resolution. Nature Structural & Molecular Biology 24, 184–186.
Marsh, M., Helenius, A., 1989. Virus entry into animal cells. Advances in Virus Research 36, 107–151.
McKenna, R., Xia, D., Willingmann, P., *et al*., 1992a. Atomic structure of single-stranded DNA bacteriophage ΦX174 and its functional implications. Nature 355, 137–143.
McKenna, R., Xia, D., Willingmann, P., Ilag, L.L., Rossmann, M.G., 1992b. Structure determination of the bacteriophage ΦX174. Acta Crystallographica Section B: Structural Science 48, 499–511.
Morais, M.C., Choi, K.H., Koti, J.S., *et al*., 2005. Conservation of the capsid structure in tailed dsDNA bacteriophages: The pseudoatomic structure of φ29. Molecular Cell 18, 149–159.
Mukhopadhyay, S., Kuhn, R.J., Rossmann, M.G., 2005. A structural perspective of the flavivirus life cycle. Nature Reviews Microbiology 3, 13–22.
Namba, K., Pattanayek, R., Stubbs, G., 1989a. Visualization of protein-nucleic acid interactions in a virus: Refined structure of intact tobacco mosaic-virus at 2.9-Å resolution by X-ray fiber diffraction. Journal of Molecular Biology 208, 307–325.
Okamoto, K., Miyazaki, N., Reddy, H.K.N., *et al*., 2018. Cryo-EM structure of a *Marseilleviridae* virus particle reveals a large internal microassembly. Virology 516, 239–245.
Oliveira, M.A., Zhao, R., Lee, W., *et al*., 1993. The structure of human rhinovirus 16. Structure 1, 51–68.
Olson, N.H., Kolatkar, P.R., Oliveira, M.A., *et al*., 1993. Structure of a human rhinovirus complexed with its receptor molecule. Proceedings of the National Academy of Sciences of the United States of America 90, 507–511.
Padron, E., Bowman, V., Kaludov, N., *et al*., 2005. Structure of adeno-associated virus type 4. Journal of Virology 79, 5047–5058.
Prasad, B.V.V., Schmid, M.F., 2012. Principles of virus structural organization. Viral Molecular Machines 726, 17–47.
Raoult, D., Audic, S., Robert, C., *et al*., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.
Rao, V.B., Black, L.W., 2010. Structure and assembly of bacteriophage T4 head. Virology Journal 7, 356.
Reteno, D.G., Benamar, S., Khalil, J.B., *et al*., 2015. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. Journal of Virology 89, 6585–6594.
Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. Nature 375, 291–298.
Richardson, J.S., 1981. The anatomy and taxonomy of protein structure. Advances in Protein Chemistry 34, 167–339.
Rossmann, M.G., 1979. Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. Journal of Applied Crystallography 12, 225–238.
Rossmann, M.G., 1984. Synchrotron radiation studies of large proteins and supramolecular structures. In: Diakun, G.P.G., D.C. (Eds.), Biological Systems: Structure and Analysis. Daresbury: Science and Engineering Research Council, pp. 28–40.
Rossmann, M.G., 1989. The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. Journal of Biological Chemistry 264, 14587–14590.
Rossmann, M.G., 1994. Viral cell recognition and entry. Protein Science 3, 1712–1725.
Rossmann, M.G., 1999. Synchrotron radiation as a tool for investigating virus structures. Journal of Synchrotron Radiation 6, 816–821.
Rossmann, M.G., 2013. Structure of viruses: A short history. Quarterly Reviews of Biophysics 46, 133–180.
Rossmann, M.G., Arnold, E., Erickson, J.W., *et al*., 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317, 145–153.
Rossmann, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallographica 15, 24–31.
Rossmann, M.G., Blow, D.M., 1963. Determination of phases by the conditions of non-crystallographic symmetry. Acta Crystallographica 16, 39–45.
Rossmann, M.G., Johnson, J.E., 1989. Icosahedral RNA virus structure. Annual Review of Biochemistry 58, 533–573.
Rossmann, M.G., Leslie, A.G.W., Abdel-Meguid, S.S., Tsukihara, T., 1979. Processing and post-refinement of oscillation camera data. Journal of Applied Crystallography 12, 570–581.
Rossmann, M.G., McKenna, R., Tong, L., *et al*., 1992a. Molecular replacement real-space averaging. Journal of Applied Crystallography 25, 166–180.
Rossmann, M.G., McKenna, R., Tong, L., *et al*., 1992b. Molecular replacement real space averaging. In: Dodson, E., Gover, S., Wolf, W. (Eds.), Molecular Replacement. Proceedings of the CCP4 Study Weekend, 31 January – 1 February 1992. Daresbury: Science and Engineering Research Council, pp. 33–48.
Rossmann, M.G., Palmenberg, A.C., 1988. Conservation of the putative receptor attachment site in picornaviruses. Virology 164, 373–382.
Sevvana, M., Long, F., Miller, A.S., *et al*., 2018. Refinement and analysis of the mature Zika virus cryo-EM structure at 3.1 Å resolution. Structure 26, 1169–1177.
Sherry, B., Mosser, A.G., Colonno, R.J., Rueckert, R.R., 1986. Use of monoclonal antibodies to identify four neutralization immunogens on a common cold picornavirus, human rhinovirus 14. Journal of Virology 57, 246–257.

Sherry, B., Rueckert, R., 1985. Evidence for at least two dominant neutralization antigens on human rhinovirus 14. Journal of Virology 53, 137–143.
Simpson, A.A., Chandrasekar, V., Hébert, B., *et al*., 2000. Host range and variability of calcium binding by surface loops in the capsids of canine and feline parvoviruses. Journal of Molecular Biology 300, 597–610.
Simpson, A.A., Hebert, B., Sullivan, G.M., *et al*., 2002. The structure of porcine parvovirus: Comparison with related viruses. Journal of Molecular Biology 315, 1189–1198.
Sirohi, D., Chen, Z., Sun, L., *et al*., 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. Science 352, 467–470.
Sirohi, D., Kuhn, R.J., 2017. Zika virus structure, maturation, and receptors. Journal of Infectious Diseases 216, S935–S944.
Smith, T.J., Kremer, M.J., Luo, M., *et al*., 1986. The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. Science 233, 1286–1293.
Smyth, M., Tate, J., Hoey, E., *et al*., 1995. Implications for viral uncoating from the structure of bovine enterovirus. Nature Structural Biology 2, 224–231.
Solovyev, A.G., Makarov, V.V., 2016. Helical capsids of plant viruses: Architecture with structural lability. Journal of General Virology 97, 1739–1754.
Stanley, W.M., 1935. Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. Science 81, 644–645.
Staunton, D.E., Merluzzi, V.J., Rothlein, R., *et al*., 1989. A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. Cell 56, 849–853.
Stubbs, G., Warren, S., Holmes, K., 1977. Structure of RNA and RNA binding site in tobacco mosaic virus from 4 Å map calculated from X-ray fibre diagrams. Nature 267, 216–221.
Subramaniam, S., Bartesaghi, A., Liu, J., Bennett, A.E., Sougrat, R., 2007. Electron tomography of viruses. Current Opinion in Structural Biology 17, 596–602.
Suhanovsky, M.M., Teschke, C.M., 2015. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. Virology 479–480, 487–497.
Sun, S., Xiang, Y., Akahata, W., *et al*., 2013b. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. eLife 2, e00435.
Sun, L., Young, L.N., Boudko, S.P., *et al*., 2013a. Icosahedral bacteriophage FX174 forms a tail for DNA transport during infection. Nature 505, 432–435.
Therkelsen, M.D., Klose, T., Vago, F., *et al*., 2019. Flaviviruses have imperfect icosahedral symmetry. Proceedings of the National Academy of Sciences of the United States of America 115, 11608–11612.
Thomassen, E., Gielen, G., Schutz, M., *et al*., 2003. The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. Journal of Molecular Biology 331, 361–373.
Tong, L., Choi, H.K., Minor, W., Rossmann, M.G., 1992. The structure determination of Sindbis virus core protein using isomorphous replacement and molecular replacement averaging between two crystal forms. Acta Crystallographica Section A 48 (Pt 4), 430–442.
Tong, L., Rossmann, M.G., 1990. The locked rotation function. Acta Crystallographica Section A 46, 783–792.
Tong, L., Wengler, G., Rossmann, M.G., 1993. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. Journal of Molecular Biology 230, 228–247.
Tsao, J., Chapman, M.S., Agbandje, M., *et al*., 1991. The three-dimensional structure of canine parvovirus and its functional implications. Science 251, 1456–1464.
Tsao, J., Chapman, M.S., Wu, H., *et al*., 1992. Structure determination of monoclinic canine parvovirus. Acta Crystallographica Section B 48, 75–88.
van Raaij, M.J., Schoehn, G., Burda, M.R., Miller, S., 2001. Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. Journal of Molecular Biology 314, 1137–1146.
von Bonsdorff, C.H., Harrison, S.C., 1975. Sindbis virus glycoproteins form a regular icosahedral surface lattice. Journal of Virology 16, 141–145.
von Bonsdorff, C.H., Harrison, S.C., 1978. Hexagonal glycoprotein arrays from Sindbis virus membranes. Journal of Virology 28, 578–583.
Voss, J.E., Vaney, M.-C., Duquerroy, S., *et al*., 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature 468, 709–712.
Walters, R.W., Agbandje-McKenna, M., Bowman, V.D., *et al*., 2004. Structure of adeno-associated virus serotype 5. Journal of Virology 78, 3361–3371.
Watson, J.D., Crick, F.H.C., 1953a. Genetical implications of the structure of deoxyribonucleic acid. Nature 171, 964–967.
Watson, J.D., Crick, F.H.C., 1953b. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 171, 737–738.
Wikoff, W.R., Liljas, L., Duda, R.L., *et al*., 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289, 2129–2133.
Williams, R.C., Smith, K.M., 1958. The polyhedral form of the Tipula iridescent virus. Biochimica et Biophysica Acta 28, 464–469.
Winkler, F.K., Schutt, C.E., Harrison, S.C., Bricogne, G., 1977. Tomato bushy stunt virus at 5.5-Å resolution. Nature 265, 509–513.
Wrigley, N.G., 1969. An electron microscope study of the structure of Sericesthis iridescent virus. Journal of General Virology. 5, 123–134.
Wu, H., Rossmann, M.G., 1993. The canine parvovirus empty capsid structure. Journal of Molecular Biology 233, 231–244.
Xiao, C., Chipman, P.R., Battisti, A.J., *et al*., 2005. Cryo-electron microscopy of the giant Mimivirus. Journal of Molecular Biology 353, 493–496.
Xiao, C., Fischer, M.G., Bolotaulo, D.M., *et al*., 2017. Cryo-EM reconstruction of the Cafeteria roenbergensis virus capsid suggests novel assembly pathway for giant viruses. Scientific Reports 7, 5484.
Xiao, C., Kuznetsov, Y.G., Sun, S., *et al*., 2009. Structural studies of the giant Mimivirus. PLOS Biology 7, e1000092.
Xiao, C., McKinlay, M.A., Rossmann, M.G., 2011. Design of capsid-binding antiviral agents against human rhinoviruses. In: Agbandje-McKenna, M., McKenna, R. (Eds.), RSC Biomolecular Sciences Series, No. 21, Structural Virology. London, England: Royal Society of Chemistry, pp. 321–339.
Xie, Q., Bu, W., Bhatia, S., *et al*., 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proceedings of the National Academy of Sciences of the United States of America 99, 10405–10410.
Xie, Q., Chapman, M.S., 1996. Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. Journal of Molecular Biology 264, 497–520.
Yan, X., Chipman, P.R., Castberg, T., Bratbak, G., Baker, T.S., 2005. The marine algal virus PpV01 has an icosahedral capsid with T=219 quasi-symmetry. Journal of Virology 79, 9236–9243.
Yan, X., Olson, N.H., Van Etten, J.L., *et al*., 2000. Structure and assembly of large lipid-containing dsDNA viruses. Nature Structural Biology 7, 101–103.
Yan, X., Yu, Z., Zhang, P., *et al*., 2009. The capsid proteins of a large, icosahedral dsDNA virus. Journal of Molecular Biology 385, 1287–1299.
Zhang, R., Hryc, C.F., Cong, Y., *et al*., 2011a. 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. EMBO Journal 30, 3854–3863.
Zhang, W., Kaufmann, B., Chipman, P.R., Kuhn, R.J., Rossmann, M.G., 2013. Membrane curvature in flaviviruses. Journal of Structural Biology 183, 86–94.
Zhang, W., Mukhopadhyay, S., Pletnev, S.V., *et al*., 2002. Placement of the structural proteins in Sindbis virus. Journal of Virology 76, 11645–11658.
Zhang, X., Xiang, Y., Dunigan, D.D., *et al*., 2011b. Three-dimensional structure and function of the *Paramecium bursaria* chlorella virus capsid. Proceedings of the National Academy of Sciences of the United States of America 108, 14837–14842.

Further Reading

Bernal, R.A., Hafenstein, S., Olson, N.H., *et al*., 2003. Structural studies of bacteriophage a3 assembly. Journal of Molecular Biology 325, 11–24.
Chiu, W., Burnett, R.M., Garcea, R.L., 1997. Structural Biology of Viruses. Oxford University Press.
Fokine, A., Zhang, Z., Kanamaru, S., *et al*., 2013. The molecular architecture of the bacteriophage T4 neck. Journal of Molecular Biology 425, 1731–1744.
Horne, R.W., 1974. Virus Structure. New York and London: Academic press.
Mateu, M.G., 2013. Structure and Physics of Viruses. Heidelberg, New York, London: Springer Dodrecht.
Nermut, M.V., Steven, A.C., 1987. Animal Virus Structure. Amsterdam, New York: Elsevier Academic Press.
Okamoto, K., Miyazaki, N., Song, C.H., *et al*., 2017. Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. Scientific Reports 7.

Plevka, P., Lim, P.Y., Perera, R., *et al.*, 2014. Neutralizing antibodies can initiate genome release from human enterovirus 71. Proceedings of the National Academy of Sciences of the United States of America 111, 2134–2139.
Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., Leiman, P.G., 2004. The bacteriophage T4 DNA injection machine. Current Opinion in Structural Biology 14, 171–180.
Roy, P., 2005. Virus Structure and Assembly. California: Elsevier Academic Press.
Yap, M.L., Klose, T., Urakami, A., *et al.*, 2017. Structural studies of Chikungunya virus maturation. Proceedings of the National Academy of Sciences of the United States of America 114, 13703–13707.

Relevant Websites

https://pdb101.rcsb.org/browse/viruses
Browse: Viruses.
PDB-101.
https://youtu.be/KoJWuWzVgqQ
Stephen Harrison (Harvard) Part 1: Virus structures: General principles.
https://www.ebi.ac.uk/pdbe/emdb/
The Electron Microscopy Data Bank (EMDB) at PDBe.
https://viralzone.expasy.org
ViralZone root.
https://www.viprbrc.org/brc/home.spg?decorator=vipr
Virus Pathogen Database and Analysis Resource.
http://viperdb.scripps.edu
Welcome to VIPERdb.
Scripps Research.