

A comparative analysis of the structural architecture of ssDNA viruses

Antonette Bennett, Robert McKenna and Mavis Agbandje-McKenna*

Department of Biochemistry and Molecular Biology, College of Medicine, The McKnight Brain Institute,
Center for Structural Biology, University of Florida, Gainesville, FL, USA

(Received 4 February 2008; final version received 2 April 2008)

Virus assembly, utilizing a limited number of viral coat protein (CP or VP) building blocks, is an excellent example of a directed macromolecular interaction occurring in nature. Two basic principles govern the assembly of spherical (icosahedral) viruses: (i) Genetic Economy – the encapsidated genome encodes a single or few CPs that assemble a protective shell (the viral capsid); and (ii) specificity – the CPs have to fold to recognize each other and form exact CP–CP interfacial interactions during the assembly pathway. Using a variety of biophysical techniques, including X-ray crystallography and cryo-electron microscopy, combined with homology model building, biochemistry and molecular biology, the nature of the interactions between protein–protein subunits and protein–nucleic acid that facilitate viral capsid assembly have been studied. This review discusses both the similarities and differences that have been elucidated for ssDNA *Microviridae*, *Geminiviridae*, and *Parvoviridae* virus families. The *Microviridae* represent a family of bacteriophages that utilize several CPs and scaffold proteins to assemble a $T = 1$ icosahedral capsid, the *Geminiviridae* plant viruses assemble a unique twinned quasi-isometric (geminiate) pseudo $T = 1$ virion assembled from a single CP and the *Parvoviridae* represent animal viruses whose $T = 1$ capsids are formed from the common overlapping region of two to four VPs that have unique N-terminal extensions. A survey of the three-dimensional (3D) data available for these viruses shows that they utilize structural commonalities, facilitated by disparate CP/VP amino acid sequences for the successful assembly of mature infectious virions.

Keywords: viral capsid protein; ssDNA viruses; $T = 1$ capsids; virus assembly

Introduction

Viruses are the causative pathogens of severe diseases and are ubiquitous in all organisms, infecting everything from bacteria to humans. To achieve such biodiversity they have evolved different strategies of host recognition, internalization, cellular trafficking, genome replication, capsid assembly and release of progeny to optimize their life cycle in their unique niche. This has resulted in viruses of different shapes and sizes, but invariably their viral coat protein(s) (CPs or VPs) form a protective shell (a viral capsid) around the infectious genomic nucleic acid, which can be ssDNA, ssRNA, dsDNA or dsRNA. The packaged viral genome encodes all the required structural CPs (or VPs) and auxiliary non-structural proteins that are required, in combination with host proteins, for host infection. Some viruses, termed enveloped viruses, in addition, incorporate their host's lipids as either an internal and/or external envelope during assembly. For a number of viruses, CP recognition and encapsidation of the genomic nucleic acid is a prerequisite for infectious capsid formation, while for others the genome is packaged into preformed capsids via interactions with viral or host encoded proteins. In addition to

*Corresponding author. Email: mckenna@ufl.edu

genome encapsidation and protection during cellular entry and trafficking, the CP can also dictate many other viral functions, including host receptor/vector recognition, transmission and the genomic transduction efficiency during infection. The formation of the assembled mature viral capsid from CP (or VP) monomers is the focus of this review.

For spherical virus assembly, the CP organization in the capsid architecture takes on the form of an icosahedron (a platonic solid with point group symmetry 5.3.2). This symmetric CP shell [10] is a consequence of it consisting of identical (or almost identical) gene products, consistent with the argument that there is insufficient volume inside a virus to accommodate a more complicated protein coding strategy. The exact 2-, 3- and 5-fold symmetry of the icosahedron permits the (quasi) equivalence symmetry [8] required to construct structures with 60 or multiples (denoted by a T number) of 60 subunits.

Significant effort has been dedicated toward understanding the driving forces behind icosahedral viral capsid assembly. Structural biology tools, such as X-ray crystallography and cryo-electron microscopy (cryo-EM) combined with 3D homology model building, enable the atomic visualization of interaction interfaces between protein–protein subunits and protein–nucleic acids in virus capsids (e.g. [1,20]). Combined with biochemical and molecular biology analysis, these studies indicate a high degree of fidelity in the steps that result in the assembly of mature infectious virus capsids. They also show that the fundamental principles governing successful viral capsid assembly, efficient polymerization of CP subunits utilizing specific interface interactions that spontaneously terminate, often employ structural polymorphisms to facilitate the required interactions (e.g. [20,21]). However, there are still many unanswered questions regarding the numerous steps that ensure the polymerization of protein subunits, with precise accuracy, in the successful assembly of infectious virions. This review will limit itself to only include the simplest single stranded (ss) DNA families, which assemble a $T = 1$ or twinned pseudo $T = 1$ icosahedral capsid.

The ssDNA virus families

There are six ssDNA virus families: *Microviridae*, *Inoviridae*, *Nanoviridae*, *Geminiviridae*, *Circoviridae* and *Parvoviridae* [42]. The first two families infect bacteria, the next two infect plants, and the last two infect mammals, with the *Parvoviridae* also infecting invertebrates. Despite this diverse host range and no genetic homology between them, the CPs/VPs for four of the six viruses (*Microviridae*, *Nanoviridae*, *Circoviridae* and *Parvoviridae*) assemble a $T = 1$ icosahedral capsid, the *Geminiviridae* assemble a unique twinned pseudo $T = 1$ icosahedral capsid, and the *Inoviridae* are filamentous. This review discusses what is known about the structural architecture of the CPs/VPs and assembled capsids of three ssDNA virus families, the *Microviridae*, *Geminiviridae* and *Parvoviridae*. These families provide three disparate examples of capsid assembly that result in a $T = 1$ capsid or twinned pseudo $T = 1$ capsids: the *Microviridae* requiring the involvement of scaffold proteins in the assembly of four different CPs; the *Geminiviridae* involving the assembly of two incomplete shells from a single CP to accommodate its genome; and the *Parvoviridae* which orchestrate interactions between the common region of two to four overlapping VPs.

Microviridae

The *Microviridae* are bacteriophages found in numerous sources populated by their hosts, such as soil, seawater, sewage and the intestine of animals [6,25]. Members of this virus family package a circular (+) sense ssDNA genome of 4.4–5.4 kb in length into capsids that are 240–270 Å in diameter. They are divided into four genera [32]: *Microvirus*, *Bdellomicrovirus*,

Chlamydiomicrovirus and *Spiromicrovirus*. Members of the *Microvirus* genus infect the host *Enterobacteria*, the type specie is *Enterobacteria phage* Φ X174; *Bdellovibrio* infect the host *Bdellovibrio*, the type specie is *Bdellovibrio phage MAC 1*; *Spiromicrovirus* infect the host *Spiroplasma*, with *Spiroplasma phage 4* as the type specie; and *Chlamydiomicrovirus* infect the host *Chlamydia* and the type specie is *Chlamydia phage 1*.

For the purpose of this review Φ X174 will be used as a representative member for the *Microviridae*, as it is the most studied and best characterized of the bacteriophages. Φ X174 infects its host via an interaction with a lipo-polysaccharide receptor on the host cell surface. Its 5.4 kb genome has open reading frames (ORFs) that encode for eleven proteins, four of which are structural proteins, F, G, H and J, present in the assembled mature virions and two other structural proteins, B and D, which serve as scaffolding proteins required for capsid assembly and are present in an intermediate procapsid [16,19,28]. The assembled mature Φ X174 virion contains 60 copies each of the F, G and J proteins and 12 copies of the pilot H protein. The major capsid protein, F (426 amino acids, 48.4 kDa), is important for host attachment and forms the integral viral capsid needed for genome encapsidation. The spike protein G (175 amino acids, 19.1 kDa), also 60 copies per virion, is important for host penetration and DNA ejection. The pilot protein H (328 amino acids, 34.4 kDa), of which there are only 12 copies in the capsid, is important for DNA ejection and infection of the host cell. The virion contains 60 copies of the DNA packaging protein J (37 amino acids, 4.2 kDa), which is highly basic at its amino terminus and is required for DNA packaging, playing a role in the neutralization of the negative charge of the ssDNA phosphate backbone. The internal scaffold protein B (120 amino acids, 13.8 kDa) and external protein D (152 amino acids, 16.9 kDa) are important for the assembly and maturation of the virion [16]. The remaining Φ X174 proteins, A, A, A*, C, E and K are essential for various steps of the life cycle including host cell lysis, genome replication and packing of progeny ssDNA [17,19,35].

Geminiviridae

The *Geminiviridae* are a family of highly pathogenic viruses that cause significant crop losses due to their infection of horticultural plants such as maize, tomatoes, beans, squash, cassava and cotton. They are normally transmitted from host to host via a whitefly or leafhopper vector. Efforts aimed at geminivirus control are mainly focused on identifying resistance/silencing genes that can be engineered into plant crops. The *Geminiviridae* package either one circular ssDNA genome (monopartite), of 2.5–2.8 kb, or two ssDNA molecules (bipartite, with component A and B), totalling ~5.2 kb, into one or two different twined pseudo-icosahedral capsids, respectively. The capsid is $\sim 200 \times 400 \text{ \AA}$ in diameter. The *Geminiviridae* are divided into four genera: *Mastrevirus*, *Curtovirus*, *Begomovirus* and *Topocuvirus*, based on the genome organization, host range and the insect vector used [31]. Members of the *Mastrevirus* have a monopartite genome and are transmitted by specific species of leafhoppers (*Cicadellidae* sp.). Members of this genus infect monocotyledonous plants; *Bean yellow dwarf virus* and *Tobacco yellow dwarf virus* infect dicotyledons. Maize streak virus (MSV) is the type member of this genus. The *curtoviruses* have monopartite genomes and are transmitted by leafhoppers to dicotyledonous plants. *Beet curly top virus* is the type member of the genus. The *begomoviruses* can have a single (e.g. *Tomato yellow leaf curl virus* (TYLCV) and *Bean yellow vein mosaic virus*) or bipartite genome (e.g. *Bean golden yellow mosaic virus*), and are transmitted by whiteflies (*Bemisia tabaci*), and infect dicots. *Bean golden mosaic virus* is the type species. The genus *Topocuvirus* has only one member, *Tomato pseudo-curly top virus*. This virus has a monopartite genome and is transmitted by treehoppers.

In this review, MSV will serve as the viral model for the discussion of the *Geminiviridae*. As for Φ X174, MSV is one of the best characterized members of its family. The virus is

transmitted specifically by the leafhopper *Cicadulina mbila*. The 2.7 kb genome encodes four gene products, CP, replication proteins Rep and RepA, and the movement protein, MP. Unlike the Φ X174 virion assembled from several CPs, the MSV virion is assembled from 110 copies of a single CP (244 amino acids, \sim 27 kDa). Hence, the CP is required for a variety of viral life cycle functions that include; insect transmission, systemic infection and encapsidation of ssDNA (reviewed in [5]). The expressed CP accumulates in the nucleus and facilitates nuclear and cell-to-cell transport of the genomic ssDNA via an interaction with the MP. In addition, the Rep and RepA proteins have multifunctional roles that include genome replication (reviewed in [5]).

Parvoviridae

The *Parvoviridae* also exhibit diverse host range and tissue tropism, with an infection outcome spectrum of non-pathogenicity to lethality. The *Parvoviridae* package linear (–) or (– and +) sense ssDNA that is \sim 5 kb into capsids that are \sim 260 Å in diameter. The family is divided into two subfamilies: *Parvovirinae*, which infect vertebrates, and the *Densovirinae*, which infects insects [36]. This review will focus specifically on the *Parvovirinae*, which infect a wide distribution of hosts including mice, cats, dogs, pigs, cows, monkeys and humans. This subfamily is subdivided into five genera: *Parvovirus*, *Erythrovirus*, *Dependovirus*, *Amdovirus* and *Bocavirus*. The type member for the *Parvovirus* genera is *Minute virus of mice* which infects mice, with other members infecting rats, cats, dogs and pigs. This genus contains both pathogenic and non-pathogenic strains. The genus *Erythrovirus* contains human parvovirus *B19* as the type member and includes viruses that infect non-human primates such as rhesus monkeys and pig-tailed macaques. The genus *Dependovirus* comprises the Adeno-associated viruses (AAVs), which require helper function from viruses such as Adenovirus and Herpes virus for replication, hence the name of the genus. These viruses, for which there has been no disease association, have been isolated from a wide range of hosts, including mice, rat, cows, snakes, non-human primates and humans. The type member for this genus is *AAV serotype 2* (AAV2). The *Amdovirus* genus currently contains one member, *Aleutian mink disease parvovirus* infecting minks. Members of the *Bocavirus* genus are bovine parvovirus, which infect cows and human bocavirus, recently discovered in children and is associated with respiratory disease and gastroenteritis.

For this review, AAV2 will serve as the parvovirus model. This virus is also the best-characterized member of the *Dependovirus* genus, mostly because of interest in its development and use for gene delivery applications. The AAV2 genome of 4.7 kb contains a replication (rep) and a capsid (cap) ORF. The rep ORF encodes four non-structural overlapping proteins Rep 78, Rep 68, Rep 52 and Rep 40 that are involved in the viral replication [29]. The cap ORF encodes three overlapping structural proteins, VP1 (735 amino acids, 87 kDa), VP2 (598 amino acids, 72 kDa) and VP3 (533 amino acids, 62 kDa), that are alternatively spliced. These proteins have a common C-terminal region (the VP3 sequence) with VP2 and VP1 having N-terminal extensions compared to VP3 of 66 and 202 amino acids, respectively. The AAV2 capsid, that is \sim 260 Å in diameter, is assembled from 60 copies (in total) of VP1:VP2:VP3 in a ratio of 1:1:8. The VP3 common region is known to be important for cell binding, antigenic recognition and possibly genomic DNA packaging [3]. VP1 and VP2 contain nuclear localization sequences and are suggested to play a role in transporting VP3 to the nucleus. VP1 has a unique N-terminal region (of \sim 130 amino acids) which displays a phospholipase A2 activity required for endosomal escape during cellular trafficking [18].

The assembled capsid and the CP architecture of the ssDNA viruses

An important component to elucidating the assembly of non-enveloped icosahedral capsids is the study of the tertiary structural architecture of their CPs and understanding of the

juxtaposition of CP interactions that configure the mature virion. In addition, structural information for other viral or host proteins transiently associated with the capsid CPs during assembly, such as scaffold or chaperone proteins, is also valuable. For the *Microviridae*, atomic and low resolution structures, determined by X-ray crystallography and cryo-EM, respectively, are available for several members, including the mature Φ X174 virion, containing proteins F, G, H and J proteins, and the assembly intermediate procapsid which additionally contains B and D scaffolding proteins and have no J proteins [12,13,26,27]. For the *Geminiviridae*, low resolution cryo-EM image reconstructed structures are available for MSV [41] and *African Cassava Mosaic Virus* [4], which belongs to the *Begomovirus* genus, into which pseudo-atomic models have been built for their CPs based on homology modelling. For the *Parvoviridae*, atomic and low-resolution structures are available for at least one member of each genera [9], with AAV2 being the first structure determined for the dependoviruses [23,39]. The wealth of structural information for these ssDNA viruses has enabled data arising from biochemical and genetic studies of their capsid assembly mechanisms to be interpreted within a structural context.

***Microviridae* Φ X174**

The structure of the Φ X174 virion has been determined by cryo-EM and X-ray crystallography to 21 and 2.7 Å resolution, respectively [26,27]. The mature capsid consists of 60 copies of the F protein forming a contiguous protein shell of ~ 260 Å in diameter. The internal volume of the F protein capsid into which the ~ 5.4 kb genome packages is $\sim 3.7 \times 10^6$ Å³. The F protein capsid surface features include protrusions at the 3-fold and depressions at the 2-fold icosahedral symmetry axes (Figure 1(A); [26,27]). The most prominent feature of the Φ X174 capsid exterior surface is the twelve spikes at the icosahedral 5-fold vertices. Each spike consists of a pentamer of G proteins positioned on a crater on the F protein capsid surface. The spikes are 70 Å wide in diameter and extend the surface of the capsid an additional 30 Å (Figure 1(A)). This arrangement of proteins (five F proteins and a pentamer of G proteins) produces a gated hydrophilic channel at each of the icosahedral vertices.

The fold of the F protein contains an eight-stranded (β B- β I) anti-parallel β -barrel (with two sheets BIDG and CHEF; [33]) found in most other virus proteins (Figure 1(B)). This motif is sometimes called a β -sandwich, because the two sheets curve towards each other in the shape of a wedge with the narrow end pointing towards the 5-fold vertex [33]. The F protein β -barrel is oriented tangential to the surface of the viral capsid (Figures 1(A) and (B)). The sheet containing strands B, I, D and G is closest to the interior of the capsid. There are two large loop insertions between the EF and the HI strands, the β EF and β HI loops, which are 163 and 112 amino acids in length, respectively. These loops, which represent $\sim 65\%$ of the F protein structure, contain 10 small α -helical ($\alpha 1$ – $\alpha 10$, Figure 1(B)) stretches. The icosahedral 2-fold symmetry related F protein interface interactions are mediated by helical regions $\alpha 1$, $\alpha 5$ and the $\alpha 5\beta$ F loop (Figures 1(B) and 2(A)). This region is the thinnest part of the F protein capsid, being only one polypeptide chain thick (Figures 1(B) and 2(A)). Small surface protrusions surround the icosahedral 3-fold axes formed primarily by helices $\alpha 4$, $\alpha 5$, $\alpha 7$ and $\alpha 8$ (Figures 1(B) and 2(D)). The icosahedral 5-fold symmetry related F protein interactions are extensive and are formed between strands in the BIDG β -sheet of one monomer and the large insertions extending from the narrow end of the β -barrel strands of the adjacent monomer (Figures 1(B) and 2(G)).

The G spike protein shares the same eight-stranded anti-parallel β barrel topology observed in the F protein but has no significant loop insertions and runs radially outwards from the F protein capsid parallel to the icosahedral 5-fold symmetry axis (Figures 1(A) and (B); [26,27]). The hydrophilic 5-fold channel formed by the F and G proteins utilizes the N-terminus, strands β F and β G, and the β F β G loop of the G protein and residues in the β FG loop of the F protein.

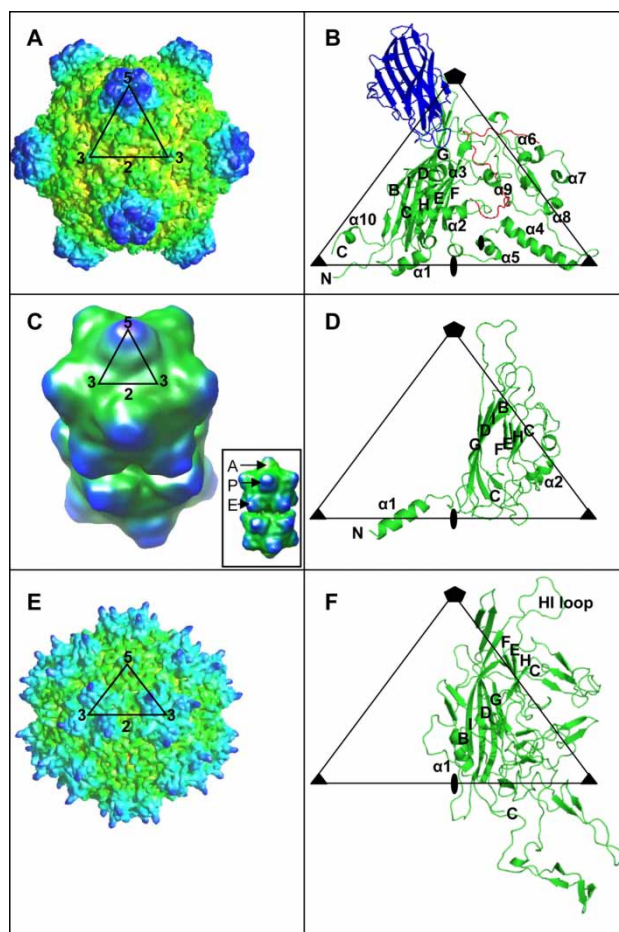


Figure 1. Virus capsids and CP/VP structures. Surface rendition of the virus capsids of (A) Φ X174, (C) MSV (insert; A = apical, B = peripentonal, E = equatorial pentamers, respectively), and (E) AAV2. The surfaces are radially coloured outwards from yellow to blue. Ribbon drawings of (B) Φ X174 F, G and J proteins in green, blue and red, respectively; (D) MSV CP in green, and (F) AAV2 VP3 in green. The virus asymmetric unit is drawn as an open triangle with the virus symmetry 2-, 3- and 5-fold axes depicted as a filled ellipsoid, triangle, and pentagon, respectively. Secondary structural elements are as labelled. Images (A), (C) and (E) were generated with the program Chimera [30] and images (B), (D) and (F) were generated with the program PyMol [11].

The J protein has an extended S-shaped topology and is located on the interior surface of the F protein capsid at an interface between 5-fold related β barrels, with its C-terminus interacting with four DNA nucleotides (Figure 1(B)). The structure and location of the H protein has not been observed in either X-ray or cryo-EM structures, although it has been postulated that some of the H protein resides in the 5-fold channel positioned ready for DNA ejection.

Geminiviridae MSV

The twinned quasi-isometric incomplete $T = 1$ icosahedral capsid architecture for the geminiviruses was determined from the cryo-EM image reconstruction of the MSV capsid at 25 Å resolution [41]. This geminate architecture is unique among all known virus families.

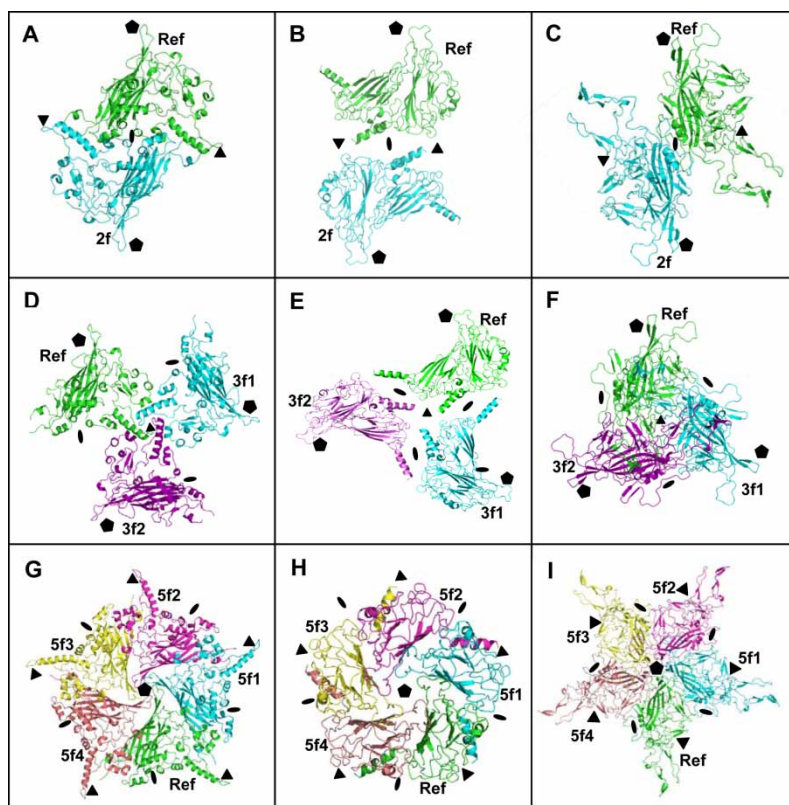


Figure 2. Capsid protein interfaces. Ribbon drawings of the icosahedral 2-, 3- and 5-fold symmetry related interfaces for Φ X174 F-protein [(A), (D) and (G)]; MSV CP [(B), (E) and (H)], and AAV2 VP3 [(C), (F) and (I)]. The virus symmetry 2-, 3- and 5-fold axes depicted as a filled ellipsoid, triangle, and pentagon, respectively. The reference CP or VP is labelled Ref, and symmetry related proteins are labelled [Example 5-fold related proteins labelled, Ref (green), 5f1 (cyan), 5f2 (magenta), 5f3 (yellow), and 5f4 (orange)]. This figure was generated with the program PyMol [11]. Available in colour online.

The MSV capsid dimensions are $\sim 220 \times 380 \text{ \AA}$ (Figure 1(C)). The 2.7 kb genome is packaged into an internal capsid volume of $\sim 1.2 \times 10^6 \text{ \AA}^3$. The capsid consists of 110 copies of the CP that are arranged as 22 pentameric capsomers with 5.2 point group symmetry. Each incomplete icosahedral ‘head’ (55 CPs) is missing 5 CPs (a pentamer) compared to a regular $T = 1$ capsid assembled with 60 CPs. The structure is built from distinct pentamers that protrude radially outwards from the capsid. Three ‘classes’ of pentameric building blocks are observed. They are: the two apical – that are located at each end of the capsid; the 10 peripentonal – that are adjacent to the apical pentamers; and the 10 equatorial – that line the waist of the capsid at the interface between the two fused incomplete icosahedra (Figure 1(C)). A gap of $\sim 25 \text{ \AA}$, containing fifteen regions of tubular density, connects the two heads. A small protrusion is observed at the icosahedral 3-fold axis and there is a small depression at the 2-fold interfaces between non-equatorial pentamers. Recently, a cryo-EM image reconstruction of ‘single’ head capsids of MSV with $T = 1$ icosahedral symmetry which package subgenomic ssDNA [7] has been determined to 10.2 \AA resolution (not shown; Reutzel *et al.* unpublished data). This structure shows a capsid constructed exclusively from pentameric building blocks that resemble the apical capsomers of the geminate capsid.

A pseudo-atomic model of the CP, residues 19–244, has been built into the cryo-EM envelope of MSV based on an amino acid sequence alignment with the CP of satellite tobacco necrosis virus (Figure 1(D); [41]). The first 18 amino acids were not built because their location could not be unambiguously assigned. As observed for the F and G proteins of Φ X174, the MSV CP model also contains an eight stranded anti-parallel β -barrel motif that forms the core contiguous capsid. For MSV, this core motif represents $\sim 60\%$ of the CP amino acids. In MSV, the BIDG sheet is also closest to the interior of the capsid, and the β -barrel runs radially outwards similar to the G protein orientation in Φ X174 (Figures 1(C) and (D)). The MSV CP model also contains two α -helical regions, residues 19–34 and 138–144, $\alpha 1$ and $\alpha 2$, respectively (Figure 1(D)). The N-terminal helix ($\alpha 1$) extends from the β -barrel motif and $\alpha 2$ is located in the β EF loop (Figure 1(D)). The 2-fold axis is formed by $\alpha 1$ and the C-termini of one CP and $\alpha 2$ from an adjacent CP between the apical and peripentonal capsomers (Figures 1(C), 2(B) and (E)). In each of the 10 equatorial pentamers, the $\alpha 1$ helix of the CP that forms the 2-fold head connection is rotated $\sim 70^\circ$ around the C α of residue 35 (compared to the other CPs) to satisfy ten inner equatorial connections (not shown). The $\alpha 1$ helix and the β CD and β GH loops form the 3-fold axis. As for the Φ X174 F protein icosahedral 5-fold symmetry related interactions, the β IDG sheet and N-terminal helix of one CP interacts with the loops extending from the narrow end of the β -barrel wedge, including $\alpha 2$, of an adjacent CP (Figures 2(G) and (H)). The MSV capsid also contains a channel that runs from the inside of the capsid to the outside at the icosahedral 5-fold axis, as observed in the Φ X174 capsid.

Parvoviridae AAV2

The AAV2 capsid structure has been determined by cryo-EM and X-ray crystallography to 10 and 3.2 Å resolution, respectively [23,39]. The characteristic surface features of the capsid include depressions at the icosahedral 2-fold axes, three spike-like protrusions around the icosahedral 3-fold axes (centred ~ 25 Å from the 3-fold axes), and a cylindrical channel at the icosahedral 5-fold axes surrounded by a moat (Figure 1(E)). This surface topology forms a capsid that is ~ 220 , ~ 230 and ~ 240 Å in diameter at the icosahedral 2-, 3- and 5-fold axes, respectively, and a diameter of ~ 290 Å at the maximal extremity of the protrusions surrounding the 3-fold axes (Figure 1(E)). The internal volume of the capsid that packages the ~ 4.7 kb genome is $\sim 3.5 \times 10^6$ Å³.

The crystal structure of the AAV2 capsid has provided information on the common VP3 amino acids 217–735 only, despite the fact that VP1, VP2 and VP3 were present in the samples studied [39]. The lack of ordering of the N-terminal regions of VP1–VP3 is consistent for all parvovirus structures determined to date (reviewed in [9]). The low copy numbers of VP1 and VP2 and a possible multiple conformation of the N-termini in all three capsid VPs, which is incompatible with the icosahedral symmetry imposed during the structure determination, are likely responsible for this apparent disorder. As described above for the F and G proteins of Φ X174 and the CP of MSV, the main structural motif of the AAV2 common VP3 region is an eight-stranded anti-parallel β -barrel motif which forms the core contiguous capsid with large inter-strand loops, reminiscent of the Φ X174 F protein insertions, forming the surface decorations on the capsid (Figure 1(F)). The AAV2 β -barrel is oriented tangential to the capsid surface, as discussed above for the Φ X174 F protein. The BIDG sheet also forms the interior surface of the AAV2 capsid. The inter-strand loop insertions constitute $\sim 70\%$ of the VP3 fold and contain small stretches of β -strand and α -helical structure (Figure 1(F)). A conserved loop region at the C-termini of the VP3 (residues 696–704), located after β I, invade neighbouring icosahedral 2-fold related VP3 monomers to form the floor of the depression at the 2-fold axis (Figures 1(E) and 2(C)). The 2-fold interface interactions also involve a conserved

(in all parvovirus structures) helix, αA , which forms the wall of the surface depression at this axis (Figures 1(F) and 2(C)). The 2-fold is the thinnest region of the capsid shell, being only one polypeptide chain thick, as is also the case in the $\Phi X174$ and MSV capsids. The interactions between the icosahedral symmetry 3-fold related monomers are the most extensive in AAV2 and involve residues within the βBC , βEF and βGH loops. The βGH loop has the most amino acid sequence variability among the parvoviruses and interdigitate to form the surface spike-like protrusions around the icosahedral 3-fold axes characteristic of members of the dependovirus genera of the *Parvoviridae* (Figure 2(F)). The spike-like protrusions surround a depression at the icosahedral 3-fold axis (Figure 1(E)). As observed for the $\Phi X174$ F protein and MSV CP, the loop extensions from the narrow end of the β -barrel wedge form the icosahedral 5-fold VP3 symmetry related interactions with elements of the BIDG sheet and the N-terminal residues in the adjacent monomer. In addition, the βHI loop (structurally conserved in all parvoviruses) is positioned above the β -strands of adjacent VP3 monomers forming the floor of the conserved depression that surrounds the icosahedral 5-fold axis (Figure 2(I)). As observed for $\Phi X174$ and MSV, AAV2 also has a channel that connects the inside and outside of the capsid at the 5-fold axis formed by the βDE ribbon which assembles a surface turret at this axis (Figures 1(E) and 2(I)).

Structure to function correlation for the ssDNA viruses: Capsid assembly

The ssDNA viruses families discussed have evolved coding schemes optimized in terms of genetic economy for the production of proteins important for functionality. This optimization includes the use of self-priming dsDNA regions within their genome and of overlapping transcripts for coding replication proteins. The matreviruses also use complementary transcripts for the coding of essential replication proteins and the parvoviruses use overlapping transcripts for the coding of VPs. Significantly, these viruses have evolved different CP/VP assembly strategies. The *Microviridae* utilize four different CPs to assemble a mature $T = 1$ icosahedral capsid which perform essential capsid-associated functions during the life cycle, the *Geminiviridae* utilize a single multifunctional CP to form a twinned pseudo $T = 1$ icosahedral capsid, and the *Parvoviridae* employ N-terminal extensions on essentially a single VP to expand the functional repertoire of a common overlapping polypeptide region in their $T = 1$ capsid. For $\Phi X174$, the F-protein capsid performs the essential function of genome protection in addition to host recognition, a specialized G protein turret permits the penetration of the host bacterial cell wall for the ejection of its genomic material and a J protein aids the packaging of its genome into preformed capsids. This virus ensures further replication fidelity by the use of a pilot protein H to guide its genome into the host cell. The *Geminiviridae* which are transmitted by a vector and have no need to recognize a receptor or evade the host immune system are assembled from a simple CP that functions to protect the genome and facilitates its transport through plants to the site of replication and the packaging of progeny genome. The overlapping VPs of the *Parvoviridae* contain regions that perform overlapping functions, such as receptor recognition, as well as unique roles, such as the phospholipase A2 function of the VP1 unique region that facilitates the trafficking of intact capsids through the endosome. The elaborate inter-strand insertions play a role in differential host recognition and antigenic reactivity including neutralization by host antibodies and adaptive host immune response evasion.

Despite the diversity in viral protein functionalization in these ssDNA viruses, the general topology of their capsid CPs/VPs and assembled virions share common features. All three viral models assemble similar thin shells at the icosahedral 2-fold axis resulting in a depression; protrusions at or surrounding a depression at the 3-fold axis and a channel at the 5-fold axis formed by β strands that connect the interior of the capsid to the external environment (Figure 1).

Further, commonalities are observed in the utilization of conserved (in their respective virus families) secondary structural motifs in the formation of symmetry related interfaces including helical regions for 2-fold interactions, helical regions in the 3-fold interactions in Φ X174 and MSV, and flexible loop insertions at the narrow end of the β -barrel wedge for the 5-fold axis interactions (Figure 2). For *Geminiviridae* and *Parvoviridae* the 5-fold channel is implicated in both DNA ejection and packaging [3,24]. However, for the *Microviridae*, DNA packaging is known to occur via the icosahedral 3-fold axis, while DNA ejection proceeds from the 5-fold channel [2,40]. It is possible that the use of flexible loop regions or helical structures that allow structural rearrangements at the symmetry axis have evolved to facilitate the dynamic capsid rearrangements required for encapsidation or events related to uncoating by these capsids. As an example, conformational differences at the 3-fold axis in the structures of the Φ X174 procapsid and the mature particle, which transitions from 'open' to 'closed' following DNA packaging, implicates this interface in virion morphogenesis.

The eight-stranded anti-parallel β barrel domain of the CPs/VPs of these ssDNA viruses form a conserved superimposable core scaffold adorned with surface loops (Figure 3) which confer host range diversity within each family. Superimposition of the Φ X174 F protein and MSV CP aligns 142 amino acids ($\sim 58\%$ of MSV CP sequence) with a root-mean-squared-deviation (RMSD) of 3.2 Å, despite a sequence identity of $\sim 7.2\%$. A similar number of residues (111 amino acids, $\sim 25\%$ of F protein sequence) are structurally aligned between the Φ X174 F protein and the AAV2 VP3 protein which are $\sim 10\%$ identical in sequence with an RMSD of 4.2 Å. The MSV CP and AAV2 VP3 are 6.2% identical, with 129 of the MSV CP 244 amino acids ($\sim 53\%$) aligning with an RMSD of 3.6 Å. These residues are located at the N- and C-termini of their polypeptide chains but knit together to assemble their capsid. With respect to the smaller MSV CP, the additional Φ X174 F protein and the AAV2 VP3 amino acids represent larger amino acid insertions between the core β strands and conserved α -helices. These additional amino acids are utilized by the latter viruses to expand their capsid volumes (3.7×10^6 and 3.5×10^6 Å³, respectively for Φ X174 and AAV2) compared MSV (1.2×10^6 Å³) consistent with their capacity to package larger genomes of ~ 5 kb compared to ~ 2.7 kb.

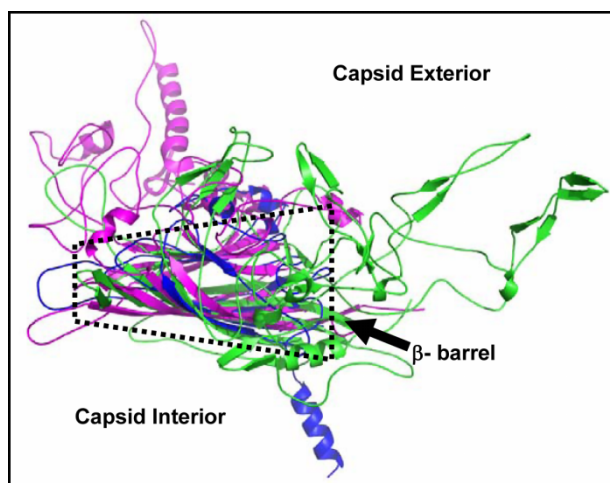


Figure 3. Structural overlay of the capsid protein β -barrels. Φ X174 F protein (magenta), MSV CP (blue) and AAV2 VP3 (green). The CP/VP proteins are orientated with the capsid exterior at the top and the interior at the bottom. The dotted open trapezoid delineates the superimposed β -barrel motif. This figure was generated by the program PyMol [11]. Available in colour online.

The calculation of the buried surface area for the formation of the Φ X174 and AAV2 symmetry related interfaces (Table 1), which can be used as a measure of the strength of association, provides a numerical value for the structural observations described above. These calculations were not carried out for MSV, since the current model is based on a low-resolution cryo-EM study. Significantly, the Φ X174 and AAV2 capsids bury similar areas at their 2- and 5-fold interfaces. On the other hand, the AAV2 3-fold interfaces association bury almost double the area for a 5-fold interaction and $\sim 15\%$ of the area buried by Φ X174 F protein 3-fold. The 5-fold interactions in both viruses involve the structurally conserved core β barrel domain amino acids residues, which are also conserved between familial genera members of ssDNA viruses being compared, suggesting an important role in the life cycle. The 5-fold interactions observed for the Φ X174 F and G proteins likely correspond to the pentameric 9S and 6S subassembly particles that have been biochemically and genetically isolated [15]. For AAV2 10-15S, subassembly complexes containing predominantly VP2 and VP3 consistent with pentamers have been isolated in addition to the disruption of capsid assembly for VP mutations that change amino acids involved in 5-fold interactions [14,37,38]. These observations support a role for pentamers in AAV2 assembly. It is also possible that trimeric intermediates are involved in AAV2 assembly, given the strong association between the icosahedral 3-fold symmetry related VP3s. However, the tolerance of variation in the residues that make up the 3-fold interfaces of the parvoviruses in general suggests that specific interactions between the residues involved are not essential for assembly. A potential explanation for the significant difference in the buried surface for the Φ X174 F protein and AAV2 VP3 likely lies with the fact that the 3-fold interface is held open for DNA packaging in Φ X174 while there is no evidence that the 3-fold axis of AAV2 plays a similar role. For MSV and other geminiviruses, biochemical and genetic studies indicate that pentamers are the building block, consistent with extensive 5-fold interactions with minimal 2- and 3-fold contacts (Bennett *et al.* unpublished data).

While the role of CP–genomic ssDNA interactions in the assembly of ssDNA viruses is still poorly understood, particularly at the structural level, like many RNA viruses, the interaction with genome is a requirement for the stable assembly of Φ X174 and the characteristic geminate capsid architecture. DNA ‘empty’ capsids are not assembled by Φ X174, rather it packages subgenomic DNA that is $\sim 20\%$ of the full length genome into non-infectious capsids if full length genome is not available [34]. For MSV, it is possible that an interaction of the full length genomic ssDNA with the CP is the control switch for geminate capsid assembly given that geminate capsids only contain full length genomes and ‘single’ head capsids package subgenomic ssDNA [7]. The N-terminal region of the MSV CP is implicated in the interaction of the CP with DNA for encapsidation, although specificity of genome recognition has not been proven. For AAV2, ‘wild-type’ empty capsids (devoid of DNA) can be assembled in heterologous systems, for example, a baculovirus/insect cell expression system. Consistently, there are reports that genomic ssDNA is packaged into preformed AAV2 capsids [14]. The AAV2 VPs can also package foreign DNA when small stretches of dsDNA recognized by its Rep proteins are present. These observations suggest that a requirement for full length genomic DNA is not an essential component of stable AAV2 capsid assembly.

Table 1. Buried surface area and number of inter–subunit interactions (NISI).

Virus/CP or VP	Buried surface area ^a (\AA^2)/NISI		
	I-2	I-3	I-5
AAV2/VP3	3060/56	10352/211	4949/90
Φ X174/F	3166/59	1777/37	6260/126

^a Solvent molecules were not included in calculations.

In summary, all the three families of ssDNA viruses discussed, the *Microviridae*, the *Geminiviridae* and the *Parvoviridae*, have evolved a functional assembled capsid to serve their life-cycle needs. Significantly, the need to encapsidate a ssDNA genome, of which there is a limited amount in most cells, appears to have driven the evolution of the CPs/VPs of these viruses to a common simple fold resulting in similarities in capsid architecture. It is anticipated that a better understanding of the CP–CP interactions and CP–DNA interactions which drive capsid assembly is required for the molecular characterization of viral life-cycles towards the elucidation of mechanisms of host infection and host adaptation as well as the development of treatment strategies. These studies aimed at molecular characterization have shown that viruses from the three domains of life, Archea, Bacteria and Eukarya, utilize commonalities in the ‘construction’ of their CPs, but evolve additional domains for functional optimization (e.g. [22]). Thus, the studies of simple capsids, such as those discussed here, serve as models for more complex systems which are not as easily tractable for structural, biochemical and genetic analysis and manipulations.

Acknowledgement

Our research on ssDNA virus capsid assembly is funded by NSF project DMS 0714912 (AB and MAM) and NIH project P01 HL51811 (MAM and RM).

References

- [1] T. Baker and J. Johnson, *Principles of virus structure determination*, in *Structural Biology of Viruses*, W. Chiu, ed., Oxford University Press, New York, 1997, pp. 39–79.
- [2] R.A. Bernal, S. Hafenstein, N.H. Olsen, V.D. Bowman, P.R. Chipman, T.S. Baker, B.A. Fane, and M.G. Rossmann, *Structural studies of bacteriophage alpha3 assembly*, *J. Mol. Biol.* 325(1) (2003), pp. 11–24.
- [3] S. Bleker, F. Sonntag, and J.A. Kleinschmidt, *Mutational analysis of narrow pores at the five fold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase A2 activity*, *J. Virol.* 79(4) (2005), pp. 2528–2540.
- [4] B. Bottcher, S. Unseld, H. Ceulemans, R.B. Russell, and H. Jeske, *Geminate structures of African cassava mosaic virus*, *J. Virol.* 78(13) (2004), pp. 6758–6765.
- [5] M.I. Boulton, *Functions and interactions of mastrevirus gene products*, *Physiol. Mol. Plant Pathol.* 60 (2002), pp. 243–255.
- [6] K.L. Brentlinger, S. Hafenstein, C.R. Novak, B.A. Fane, R. Borgon, R. McKenna, and M. Agbandje-Mckenna, *Microviridae, a family divided: Isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus*, *J. Bacteriol.* 184(4) (2002), pp. 1089–1094.
- [7] C.G. Casado, G. Javier Ortiz, E. Padron, S.J. Bean, R. McKenna, M. Agbandje-McKenna, and M.I. Boulton, *Isolation and characterization of subgenomic DNAs encapsidated in ‘single’ T = 1 isometric particles of maize streak virus*, *Virology* 323(1) (2004), pp. 164–171.
- [8] D.L. Caspar and A. Klug, *Physical principles in the construction of regular viruses*, *Cold Spring Harb. Sympos. Quant. Biol.* 27 (1962), pp. 1–24.
- [9] M.S. Chapman and M. Agbandje-McKenna, *Atomic structure of viral particles*, in *Parvoviruses*, J.R. Kerr et al., eds., Edward Arnold, Ltd, New York, 2006.
- [10] F.H.C. Crick and J.D. Watson, *Structure of small viruses*, *Nature* 177 (1956), pp. 473–475.
- [11] W.L. DeLano, *The PyMol Molecular Graphics System*, DeLano Scientific, San Carlos, CA, 2002.
- [12] T. Dokland, R. McKenna, L.L. Ilag, B.R. Bowman, N.L. Incardona, B.A. Fane, and M.G. Rossmann, *Structure of a viral procapsid with molecular scaffolding*, *Nature* 389(6648) (1997), pp. 308–313.
- [13] T. Dokland, R. McKenna, D.M. Sherman, B.R. Bowman, W.F. Bean, and M.G. Rossmann, *Structure determination of the phi X174 closed procapsid*, *Acta Crystallograph D Biol. Crystallograph* 54(Pt 5) (1998), pp. 878–890.

- [14] R. Dubielzig, J.A. King, S. Weger, A. Kern, and J.A. Kleinschmidt, *Adeno-associated virus type 2 protein interactions: Formation of pre-encapsidation complexes*, J. Virol. 73(11) (1999), pp. 8989–8998.
- [15] B.A. Fane and P.E. Prevelige, Jr., *Mechanism of scaffolding-assisted viral assembly*, Adv. Protein Chem. 64 (2003), pp. 259–299.
- [16] H. Fujisawa and M. Hayashi, *Viral DNA-synthesizing intermediate complex isolated during assembly of bacteriophage phi X174*, J. Virol. 19(2) (1976), pp. 409–415.
- [17] H. Fujisawa and M. Hayashi, *Gene A product of phi X174 is required for site-specific endonucleolytic cleavage during single-stranded DNA synthesis in vivo*, J. Virol. 19(2) (1976), pp. 416–424.
- [18] A. Girod, C.E. Wobus, Z. Zádori, M. Ried, K. Leike, P. Tijssen, J.A. Kleinschmidt, and M. Hallek, *The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity*, J. Gen. Virol. 83(Pt 5) (2002), pp. 973–978.
- [19] M. Hayashi, A. Aoyama, D.L. Richardson, and M.N. Hayashi, *Biology of the bacteriophage Φ X174*, in *The Bacteriophages*, R. Calendar, ed., Vol. 2, Plenum Publishing Corporation, New York, 1988, pp. 1–71.
- [20] J.E. Johnson, *Functional implications of protein–protein interactions in icosahedral viruses*, Proc. Natl Acad. Sci. USA 93(1) (1996), pp. 27–33.
- [21] J.E. Johnson and J.A. Speir, *Quasi-equivalent viruses: A paradigm for protein assemblies*, J. Mol. Biol. 269(5) (1997), pp. 665–675.
- [22] R. Khayat, L. Tang, E.T. Larson, C.M. Lawrence, M. Young, and J.E. Johnson, *Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses*, Proc. Natl Acad. Sci. USA 102(52) (2005), pp. 18944–18949.
- [23] S. Kronenberg, J.A. Kleinschmidt, and B. Bottcher, *Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids*, EMBO Rep. 2(11) (2001), pp. 997–1002.
- [24] H. Liu, A.P. Lucy, J.W. Davies, and M.I. Boulton, *A single amino acid change in the coat protein of Maize streak virus abolishes systemic infection, but not interaction with viral DNA or movement protein*, Mol. Plant Pathol. 2 (2001), pp. 223–228.
- [25] S. Mc Grath and D. van Sinderen, *Bacteriophage: Genetics and Molecular Biology*, Vol. 1, Academic Press, Caister, 2007.
- [26] R. McKenna, L.L. Iiag, and M.G. Rossmann, *Analysis of the single-stranded DNA bacteriophage phi X174, refined at a resolution of 3.0 Å*, J. Mol. Biol. 237(5) (1994), pp. 517–543.
- [27] R. McKenna, D. Xia, P. Willingman, L.L. Iiag, S. Krishnaswamy, M.C. Rossmann, N.H. Olson, T.S. Baker, and N.L. Incardona, *Atomic structure of single-stranded DNA bacteriophage phi X174 and its functional implications*, Nature 355(6356) (1992), pp. 137–143.
- [28] R. Mukai, R.K. Hamatake, and M. Hayashi, *Isolation and identification of bacteriophage phi X174 prohead*, Proc. Natl Acad. Sci. USA 76(10) (1979), pp. 4877–4881.
- [29] N. Muzyczka and K.I. Berns, *Parvoviridae: The viruses and their replication*, in *Fields Virology*, D.M. Knipe and P.M. Howley, eds., Lippincott Williams and Wilkins, New York, 2001, pp. 2327–2360.
- [30] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera – a visualization system for exploratory research and analysis*, J. Comput. Chem. 25(13) (2004), pp. 1605–1612.
- [31] C.R. Pringle, *Virus taxonomy at the XIth International Congress of Virology, Sydney, Australia, 1999*, Arch. Virol. 144(10) (1999), pp. 2065–2070.
- [32] F. Rohwer and R. Edwards, *The phage Proteomic tree: A genome-based taxonomy for phage*, J. Bacteriol. 184(16) (2002), pp. 4529–4535.
- [33] M.G. Rossmann and J.E. Johnson, *Icosahedral RNA virus structure*, Annu. Rev. Biochem. 58 (1989), pp. 533–573.
- [34] R.L. Sinsheimer, *A single-stranded DNA from bacteriophage phi X174*, Brookhaven Symp. Biol. (12) (1959), pp. 27–34.
- [35] C. Sumida-Yasumoto and J. Hurwitz, *Synthesis of phi X174 viral DNA in vitro depends on phiX replicative form DNA*, Proc. Natl Acad. Sci. USA 74(10) (1977), pp. 4195–4199.
- [36] P. Tattersall, *The evolution of parvovirus taxonomy*, in *Parvoviruses*, J.R. Kerr et al., eds., Edward Arnold, Ltd, New York, 2006, pp. 5–14.
- [37] A. Wistuba, S. Weger, A. Kern, and J.A. Kleinschmidt, *Intermediates of adeno-associated virus type 2 assembly: Identification of soluble complexes containing Rep and Cap proteins*, J. Virol. 69(9) (1995), pp. 5311–5319.

- [38] P. Wu, W. Xiao, T. Conlon, J. Hughes, M. Agbandje-Mckenna, J. Ferkol, T. Flotte, and N. Muzyczka, *Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism*, *J. Virol.* 74(18) (2000), pp. 8635–8647.
- [39] Q. Xie, W. Bu, S. Bhatia, J. Hare, T. Somasundaram, A. Azzi, and M.S. Chapman, *The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy*, *Proc. Natl Acad. Sci. USA* 99(16) (2002), pp. 10405–10410.
- [40] K. Yazaki, *Electron microscopic studies of bacteriophage phi X174 intact and 'eclipsing' particles, and the genome by the staining, and shadowing method*, *J. Virol. Methods* 2(3) (1981), pp. 159–167.
- [41] W. Zhang, N.H. Olson, T.S. Baker, L. Faulkner, M. Agbandje-Mckenna, M.I. Boulton, J.W. Davies, and R. McKenna, *Structure of the Maize streak virus geminate particle*, *Virology* 279(2) (2001), pp. 471–477.
- [42] M.H.V. van Regenmortel, C.M. Fauquet, D.H.L. Bishop, E. Carstens, M.K. Estes, S.M. Lemon, J. Maniloff, M.A. Mayo, D.J. McGeoch, C.R. Pringle, and R.B. Wickner (eds.), *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, Academic Press, London, San Diego, 1999.