



The microviridae: Diversity, assembly, and experimental evolution



Sarah M. Doore, Bentley A. Fane*

School of Plant Sciences and the BIO5 Institute University of Arizona, 1657 E. Helen Street, Tucson, AZ 85721, USA

ARTICLE INFO

Article history:

Received 30 November 2015

Returned to author for revisions

5 January 2016

Accepted 26 January 2016

Keywords:

phiX174

Microviridae

Microvirus

Bacteriophage

Virus

Assembly

Virus evolution

ABSTRACT

The *Microviridae*, comprised of ssDNA, icosahedral bacteriophages, are a model system for studying morphogenesis and the evolution of assembly. Historically limited to the ϕ X174-like viruses, recent results demonstrate that this richly diverse family is broadly divided into two groups. The defining feature appears to be whether one or two scaffolding proteins are required for assembly. The single-scaffolding systems contain an internal scaffolding protein, similar to many dsDNA viruses, and have a more complex coat protein fold. The two-scaffolding protein systems (ϕ X174-like) encode an internal and external species, as well as an additional structural protein: a spike on the icosahedral vertices. Here, we discuss recent *in silico* and *in vivo* evolutionary analyses conducted with chimeric viruses and/or chimeric proteins. The results suggest 1) how double scaffolding systems can evolve into single and triple scaffolding systems; and 2) how assembly is the critical factor governing adaptation and the maintenance of species boundaries.

© 2016 Elsevier Inc. All rights reserved.

Introduction

Like all icosahedral ssDNA viruses, *Microviridae* capsids display $T=1$ icosahedral symmetry (Bennett et al., 2008; Crowther et al., 2003; Tsao et al., 1992; Zhang et al., 2001). First isolated in the 1920s, (Sertic and Bulgakov, 1935), ϕ X174 is still widely studied today. Over 50 *Escherichia coli* microviruses have since been described (Bowes and Dowell, 1974; Godson et al., 1978; Kay and Bradley, 1962; Rokyta et al., 2006; Taketo, 1976). Along with recently described pequeñoviruses, these viruses constitute the candidate *Microvirinae* subfamily (Bryson et al., 2015). Historically, microvirus species were defined by differences in growth properties, e.g. temperature and host range. More recently, sequence similarity has been used to define species. Based on genomic analyses, the microviruses are broadly organized into at least 3 clades, represented by ϕ X174, G4 and α 3 (Fig. 1A). However, using two methods to define species has complicated microvirus taxonomy, as phages with very similar sequences can have disparate temperature or host ranges.

Icosahedral ssDNA phages have also been isolated from *Spiroplasma*, *Chlamydia*, and *Bdellovibrio* (Brentlinger et al., 2002; Garner et al., 2004; Liu et al., 2000; Renaudin et al., 1987; Richmond et al., 1982). Members of this new group, named *Gokushovirinae* (gokusho=Japanese for “very small”), are frequently detected in metagenomic studies (Krupovic and Forterre, 2011; Labonte and Suttle, 2013; Roux et al., 2012; Tucker et al.,

2011). Based on *in silico* methods, two additional sub-families have been proposed: the *Alpavirinae*, which are prophages of *Bacteroidetes*; and the *Pichovirinae*, from fresh and ocean water environments (Krupovic and Forterre, 2011; Roux et al., 2012). Although alpa- and pichovirus particles have yet to be isolated, their prevalence in environmental samples provides a more comprehensive picture of the *Microviridae* as a whole. The linearized maps in Fig. 2 illustrate key similarities and differences between *Microviridae* genomes. For the alpa- and pichoviruses, only genes encoding proteins of known function, based on sequence similarities, and larger ORFs are depicted.

Notably, subfamily members differ in the number and nature of scaffolding proteins used for assembly. Microviruses are unique in utilizing two scaffolding proteins. Most small viruses ($T \leq 3$) do not require scaffolding proteins, while larger viruses typically utilize a single protein (Fane and Prevelige, 2003; Prevelige and Fane, 2012; Zlotnick and Fane, 2010). Unlike microviruses, the gokushoviruses only have an internal scaffolding protein and lack a major spike protein. Scaffolding proteins have not been identified in alpa- or pichoviruses (Brentlinger et al., 2002; Krupovic and Forterre, 2011; Roux et al., 2012). However, characterizations of these genomes may have missed the internal scaffolding protein signature – a highly aromatic C-terminus – that is even weak in gokushoviruses. The gokusho scaffolding protein was identified biochemically by comparing the protein composition of purified Chp2 virions and procapsids (Clarke et al., 2004). The available data suggest that the internal scaffolding protein may be the original microvirus scaffolding protein, as two-scaffolding systems are rare. Phylogenetic studies suggest that the extant external scaffolding protein first arose within the ϕ X174-like clade and subsequently spread to the

* Corresponding author. Tel.: 520-626-6634.

E-mail address: bfane@email.arizona.edu (B.A. Fane).

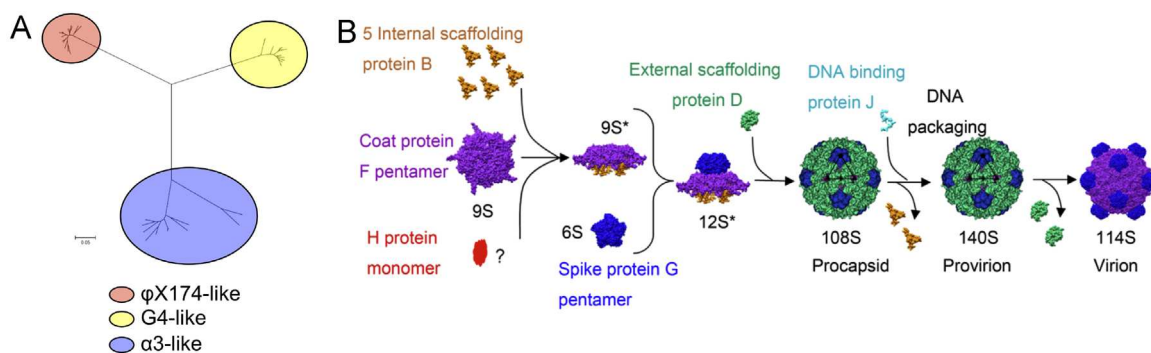


Fig. 1. (A) Microvirus phylogenetic tree based on whole genome sequences, adapted from Rokyta et al. (2006). (B) The microvirus assembly pathway. Assembly begins with the pentamerization of coat protein F to form the 9S particle; and the major spike protein G to form the 6S particle. Five copies of internal scaffolding protein B, along with one DNA pilot protein H, bind to the underside of the 9S particle, inducing a conformational change and producing the 9S* particle. This intermediate can bind to the 6S pentamer, producing the 12S* particle. 240 copies of external scaffolding protein D organize twelve 12S* particles into a procapsid. The DNA packaging complex associates with procapsids to initiate the synthesis of single-stranded genomic DNA. The genome, along with 60 copies of DNA binding protein J, enter the procapsid through one of the pores at the three-fold axes of symmetry. Protein J displaces the internal scaffolding protein B, which undergoes autoproteolysis. Once packaging is complete, the pentamers fall inwards, dissociating from the external scaffolding protein D. This produces the mature, infectious virion.

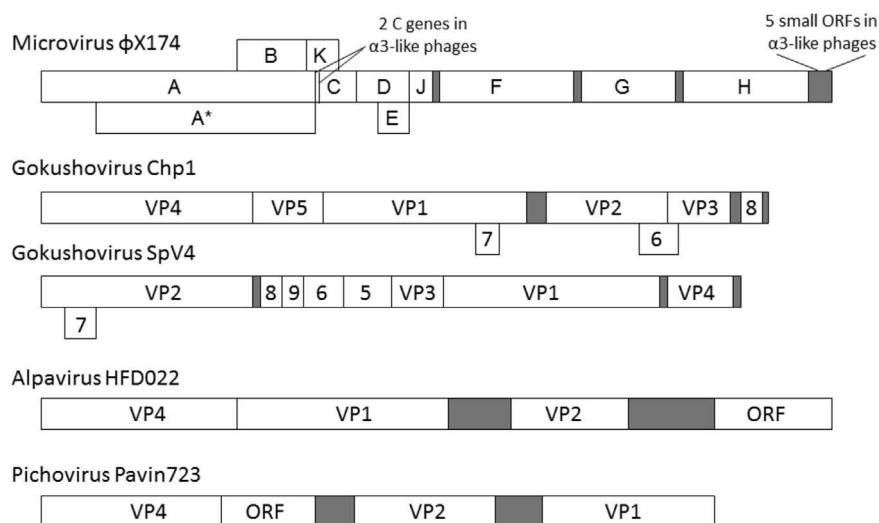


Fig. 2. Linearized maps of *Microviridae* genomes. Abbreviations and functions of gene products: A, DNA replication protein (VP4 in Chp1, alpa- and pichoviruses; VP2 in SpV4); A*, nonessential for laboratory propagation; B, internal scaffolding protein (VP3 in Chp1); K, nonessential for laboratory propagation; C, switch from dsDNA to ssDNA synthesis (VP5 in Chp1); D, external scaffolding protein; E, lysis protein; J, DNA binding protein (VP8 in Chp1, SpV4); F, coat protein (VP1 in Chp1, SpV4, alpa- and pichoviruses); G, major spike protein; H, DNA pilot protein (VP2 in Chp1, alpa- and pichoviruses; VP4 in SpV4).

G4- and α 3-like viruses (Rokyta et al., 2006). Alternatively, divergence of this gene may have been selected in only the G4- and α 3-like viruses (Rokyta et al. 2006). The only other known two-scaffolding systems are parasitic viruses, e.g. the P4 satellite of bacteriophage P2. P4 encodes an external scaffolding protein called Sid (size determining), whereas the P2 helper phage encodes an internal scaffolding protein. The P2 capsid is $T=7$; however, Sid organizes P2 capsomeres into a smaller $T=4$ structure, which can only accommodate the smaller satellite genome (Goldstein et al., 1974). Whether the microviruses acquired their external scaffolding proteins from a parasite virus is not known. It is also possible that the two-scaffolding system was lost in all *Microviridae* except microviruses. Since other members of the family infect intracellular parasites or are capable of lysogeny, the external scaffolding gene could have been lost during the genome streamlining that often occurs from these lifestyles.

The development of a powerful morphogenetic system

The tools for manipulating microvirus genomes were developed alongside their *E. coli* host, which was the model bacterial

system at the onset of molecular biology. Oligonucleotide-mediated mutagenesis, which required ssDNA, was first developed using ϕ X174 (Hutchison et al., 1978). This facilitated the construction of amber mutants that were used to define the role of each gene product and dissect the temporal order of assembly pathway (Hayashi et al., 1988). Microviruses were also among the first bacteriophages studied by x-ray crystallography. High-resolution virion and/or procapsid structures have been determined for representatives of all three clades (Bernal et al., 2004, 2003; Dokland et al., 1999, 1997; McKenna et al., 1994, 1992). Biochemically characterized mutations, as well as their second-site suppressors, can be mapped onto these structures. While structures are static, the location of characterized mutations can fill in temporal details between biochemically defined assembly intermediates and identify transient interactions (Uchiyama et al., 2007; Uchiyama and Fane, 2005).

The biochemistry of microvirus assembly has also been well characterized. *In vitro* DNA synthesis, DNA packaging and procapsid assembly systems have been developed (Aoyama et al., 1981; Aoyama and Hayashi, 1986; Cherwa et al., 2011a; Kornberg, 1980). As depicted in Fig. 1B, the two microvirus scaffolding proteins effectively divide the assembly pathway into an early stage,

mediated by internal scaffolding protein B; and a late stage, orchestrated by external scaffolding protein D (Cherwa et al., 2011a; Gordon et al., 2012). Intermediates throughout the assembly pathway can be isolated by rate-zonal sedimentation, facilitating the characterization of assembly defects.

Experimental evolution is a more recent approach for studying microvirus biology. It has been used to identify mutations resulting in higher fitness under static conditions and fluctuating environments (Brown et al., 2013, 2010; Bull et al., 1997; Caudle et al., 2014; Crill et al., 2000; Holder and Bull, 2001; Lee et al., 2011; McGee et al., 2014; Pepin et al., 2008; Pepin and Wichman, 2007; Poon and Chao, 2006; Rokyta et al., 2011; Wichman et al., 1999, 2005). Using chimeric proteins, inhibitory proteins, and chimeric viruses, experimental evolutionary data has resolved protein–protein interactions on a systems level and defined the boundaries separating species (Burch and Fane, 2000b, 2003; Chen et al., 2007; Cherwa and Fane, 2009, 2011; Cherwa et al., 2009, 2011b; Doore et al., 2014; Doore and Fane, 2015; Rokyta and Wichman, 2009; Sackman and Rokyta, 2013; Sackman et al., 2015; Uchiyama et al., 2007; Uchiyama and Fane, 2005; Uchiyama et al., 2009). From a molecular standpoint, approaches involving chimeras offer several advantages: 1) instead of individual residues, entire protein–protein interfaces are altered; 2) the introduced gene can accumulate mutations, which cannot occur using cloned genes (Burch and Fane, 2000a, b; Burch et al., 1999; Uchiyama and Fane, 2005); 3) the foreign gene is expressed at the same level as the native gene; and 4) the transferred gene produces a functional product, at least in its native system.

Microviridae structures

High-resolution structures have been determined for several microvirus procapsids and virions (Bernal et al., 2004, 2003; Dokland et al., 1999, 1997; McKenna et al., 1994, 1992). In the virions, which are 250 Å in diameter, 60 copies of coat protein F

are icosahedrally arranged as twelve pentamers or vertices. Each vertex is decorated with a pentamer of the major spike protein G, which is involved in host cell recognition and attachment. This protein is not found in the gokushoviruses. Both F and G proteins contain an eight-stranded, antiparallel β -barrel. The F protein, which is considerably larger, has two extensive insertion loops (Fig. 3A), whereas the G protein has only small, 10–11 residue extensions at its termini (Fig. 3B). Inside the capsid, there are 60 copies of the DNA binding protein J closely associated with the viral genome and grooves within the capsid proteins. Virions also contain 10–12 copies of the DNA pilot protein H. Their exact location(s) and structure(s) have not been determined in the capsid. Based on the stoichiometry of assembly intermediates, each capsomere contains one H protein (Cherwa et al., 2011a, 2008), suggesting it is a monomer during assembly. However, *in vitro*, 10 H proteins form an α -helical barrel: a variation of the classical coiled-coil motif (Sun et al., 2014a, 2014b). This structure has been directly visualized by cryo-electron microscopy (cryoEM) tomography resting under the capsid and spanning the cell wall of infected hosts.

The $\alpha 3$ and $\phi X174$ procapsids are approximately 360 Å in diameter, and multiple structures have been determined by both cryoEM and x-ray crystallography (Bernal et al., 2004, 2003; Dokland et al., 1999, 1997; Ilag et al., 1995). Twelve pentameric capsomeres, containing proteins F, G, B, and H in a respective 5:5:5:1 ratio, are held in close proximity by a lattice of 240 external scaffolding D proteins. Four external scaffolding proteins contact each of the 60 coat and spike proteins in an asymmetric arrangement, with individual positions denoted D₁–D₄ (Fig. 3D). Sixty copies of the internal scaffolding B protein reside inside the structure and interact with the coat protein *via* highly conserved C-terminal aromatic amino acids. There are two major differences between the cryoEM and x-ray structures. The cryoEM structures contain 30 Å pores at the 3-fold axes of symmetry and ~ 10 Å divisions separate the coat protein pentamers (Bernal et al., 2003; Ilag et al., 1995). Thus, the external scaffolding protein lattice holds

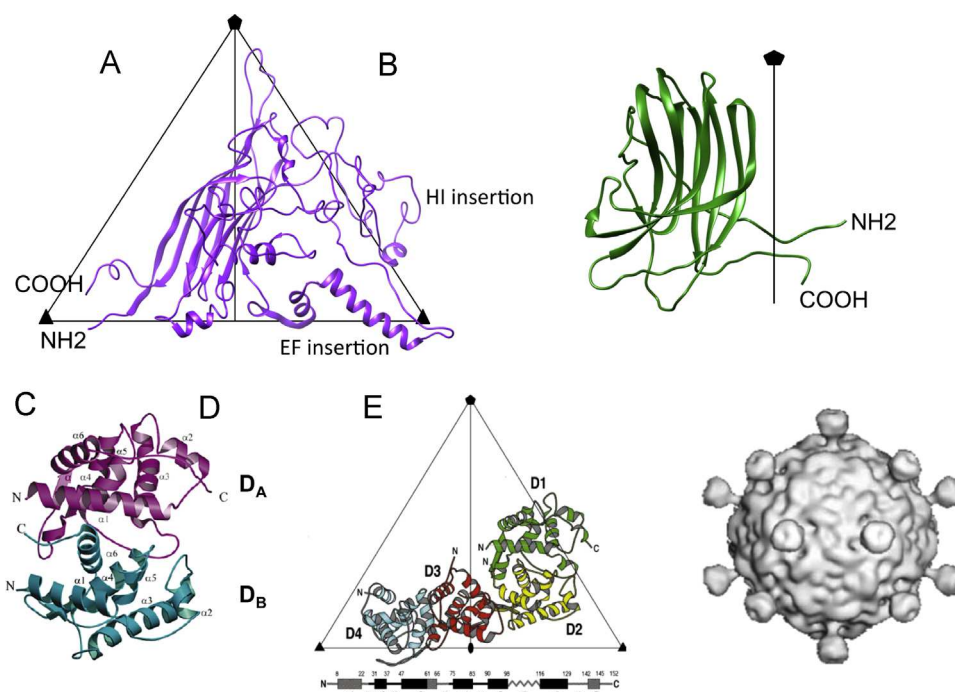


Fig. 3. Atomic structures of microvirus virion and procapsid components. (A) Monomer of the coat protein F. (B) Monomer of the major spike protein G. (C) A dimer of the assembly-naïve external scaffolding proteins, D_AD_B. (D) A tetramer of D proteins as they would be arranged in the procapsid lattice, occupying the quasi-equivalent positions D₁ through D₄. (E) Cryo-EM structure of the gokushovirus Spv4, resolved to 27Å. Panels C and D were adapted from (Morais et al., 2004); panel E was adapted from (Chipman et al., 1998).

procapsids together. In the x-ray structure (Dokland et al., 1999, 1997), the pentamers have collapsed radially inwards, resting in the virion position with closed 3-fold axes. This closed procapsid most likely represents an off-pathway assembly product; consequently, some of the coat-external scaffolding protein interactions were lost. Several genetic analyses have since suggested additional coat-external scaffolding protein contact sites (Burch and Fane, 2003; Uchiyama et al., 2007; Uchiyama and Fane, 2005).

The assembly-naïve external scaffolding protein dimer structure has also been determined to 3.3 Å resolution (Morais et al., 2004). Rather than the D₁–D₄ conformations seen in the closed procapsid, only two conformations were seen. These are denoted D_A and D_B (Fig. 3C). The D_A subunit most closely resembles the D₁ and D₃ structures, whereas the D_B subunit resembles the D₂ and D₄ structures. Thus, the assembly-naïve D_AD_B asymmetric dimer is poised to become either the D₁D₂ or D₃D₄ dimer found in the lattice (Fig. 3D). This asymmetry is largely a function of a 30° kink that allows the subunits to rest in a non-quasi-equivalent icosahedral arrangement. This kink occurs at glycine residue 61 in only the D₁ and D₃ subunits and is critical for assembly (see *Late Assembly*).

Unlike the microviruses, other members of the *Microviridae* are poorly characterized in terms of structure. The only resolved gokushovirus structure is SpV4 (Chipman et al., 1998), determined to 27 Å resolution by cryo-electron microscopy (Fig. 3E). Based on this structure and genetic content, there are clearly major differences between micro- and gokushoviruses. Gokushovirus genomes do not encode a major spike G protein homolog or an external scaffolding protein. Rather than spikes decorating the five-fold axes of symmetry, a region of the coat protein forms “mushroom-like” protrusions at the three-fold axes. These extend radially outwards from the capsid by about 54 Å, exhibiting a globular head and a narrow stalk region. Gokushovirus coat proteins contain ~70-residue insertion loops, which trimerize to form these protrusions (Chipman et al., 1998). Modeling of the protrusions suggests that a hydrophobic depression may form at the top. However, the resolution of the cryoEM reconstruction is far too low to determine whether the depression actually forms. It is assumed that the depression mediates host cell recognition and attachment; however, there is no experimental data to support this claim. As more gokushoviruses are characterized, additional structure and structure–function studies may follow. These will help decipher the function of these curious morphogenetic features.

Microvirus assembly

Morphogenesis begins with homogeneous pentameric sub-assemblies of the coat F and major spike G proteins, designated the 9S and 6S particles, respectively (Fig. 1B). Neither protein's primary sequence is highly conserved across microvirus clades (~60% to 70% identity for F proteins and ~30% to 40% identity for G proteins); yet structurally, these proteins are nearly super-imposable (Bernal et al., 2003; McKenna et al., 1996). In the coat protein, most of the amino acid divergence can be mapped to flexible loops. Conversely, the residues forming the β-barrels cores are more conserved across clades. The loop regions mediate most of the two, three, and five-fold related coat protein contacts, as well as interactions with the substrate specificity domains of the external scaffolding protein (Burch and Fane, 2000b; Uchiyama et al., 2007; Uchiyama and Fane, 2005). This divergence may consequently prevent the formation of mixed coat protein complexes in cells co-infected with phages from different clades. Similarly, the major spike proteins exhibit greatest divergence at the five-fold contact sites, which may prevent the formation of mixed spike pentamers in co-infected cells.

After coat and spike protein pentamerization, assembly and can be divided into early and late stages, respectively mediated by the internal and external scaffolding proteins. Five internal scaffolding B proteins, along with one DNA pilot protein H, bind to the underside of the 9S particle and induce a conformational change to produce the 9S* particle. The 6S particle binds atop the 9S* particle to produce the 12S* intermediate (Cherwa et al., 2008; Gordon et al., 2012). Late assembly is governed by the external scaffolding protein D, where 240 copies of the D protein organize twelve 12S* particles into procapsids (Cherwa et al., 2011a; Dokland et al., 1999, 1997). Four external scaffolding proteins are arranged in two asymmetric dimers per coat protein (D₁D₂ and D₃D₄; Fig. 3D). Each subunit is structurally distinct and forms a unique set of interactions with the underlying five-fold related and two-fold related coat proteins (Burch and Fane, 2003; Dokland et al., 1999, 1997). Once procapsids have formed, genomic DNA is concurrently synthesized and packaged. Packaging triggers the release of the scaffolding proteins and facilitates a radial, inward collapse of the 12S* pentamers. These packaged, sealed particles are the final product of assembly: mature virions.

Using experimental evolution to explore species boundaries

Microvirus genomes may exhibit a lower incidence of horizontal gene transfer (HGT) than double-stranded (ds) DNA bacteriophage genomes, where evidence of HGT is plentiful (Casjens and Thuman-Commike, 2011; Hendrix et al., 1999; Juhala et al., 2000; Lefeuvre et al., 2009). Based on whole genome phylogenies, HGT is not extensive enough to conceal relationships between clades (Rokyta et al., 2006). When examined individually, the genes encoding the coat, major spike, and DNA binding proteins do not form monophyletic groups. However, transfer events are still rare compared to many other phages. Recombination between microviruses should be possible: sequences are similar enough to facilitate homologous recombination but dissimilar enough that a gene transfer event would be detectable. Although temperature and host ranges differ between species, there is still enough overlap that co-infections likely occur. Thus, the dearth of HGT may indicate that some proteins fail to efficiently function after transfer, and natural recombinants do not survive to be sampled.

With easily manipulated genomes, artificial horizontal gene transfer can be engineered between microvirus genes or genomes. The resulting chimeras are useful for examining multiple aspects of the viral lifecycle and for determining why some gene products are unable to function across species lines. Results indicate that sequence and/or structure conservation cannot predict the ability to cross-function. Examining conservation of a single protein ignores the evolutionary history of the whole complex, where interacting proteins have co-evolved, and does not take epistasis into account. In some cases, chimeras can be used to pinpoint important residues or domains of a protein or of its interacting partner. The following sections describe studies that use these tools to investigate microvirus structural, scaffolding, DNA packaging, and DNA binding proteins.

The exchange of structural proteins within a clade

To test the ability of a structural protein to function between species of the same clade, Rokyta et al. exchanged coat F genes between two G4-like viruses, ID2 and ID12 (Rokyta and Wichman, 2009). Despite only 8% amino acid divergence, both chimeric viruses experienced severe fitness losses, with doublings per hour reduced to approximately half the parental values. These strains were experimentally evolved in liquid culture for 350 generations, with samples removed and sequenced throughout the experiment

to document compensatory mutations. By the final passage, the chimeras had recovered ~80% to 90% of parental fitness. Most compensatory mutations mapped to the scaffolding proteins or to the coat protein at coat-scaffolding contact sites, suggesting assembly defects. Thus, while coat protein sequences and structures are conserved within the clade, minor differences had large effects on assembly efficiency.

To test how low these differences could be while still exhibiting an effect on fitness, [Sackman et al. \(2015\)](#) exchanged the coat F gene between extremely similar phages ID12 and NC6. The coat protein sequences differ by only 1% on the amino acid level. Despite this low level of divergence, gene transfer resulted in a significant decrease in the NC6 background. Conversely, there was a very slight, non-significant fitness gain for the ID12 hybrid. This asymmetry indicates that epistasis can still affect the outcome of gene transfer, even when genes are highly similar. The authors adapted the chimeric strains in liquid culture and determined which mutations conferred fitness gains. Similar to the aforementioned study, mutations mapped to coat-external scaffolding protein contact sites. These likely compensating for the few diverged amino acids of the coat protein, which are all located near external scaffolding protein contact sites. These results again indicate that minor differences can have large effects on assembly.

The above studies describe chimeric viruses that are less-fit or as-fit as their ancestors, even after experimental evolution. In contrast, [Sackman and Rokytka \(2013\)](#) demonstrate a case where HGT can ultimately produce a virus with greater fitness than its ancestors. This study used the related phages ID2 and ID204, which seem to share an ID204-like ancestor. ID2 likely arose *via* a recombination event involving the major spike gene, which occurred between the ID204-like ancestor and a more distantly-related phage. By exchanging the major spike G genes between ID2 and ID204, the authors both reversed and re-created this original recombination event. Although the chimeras had significantly lower fitness values compared to their respective ancestors, these losses could be restored after a brief period of adaptation. In the case of the ID204 hybrid, which approximately re-created the generation of ID2, fitness exceeded ID204 values after ten serial passages and reached or exceeded ID2 values by the end of the experiment (80 total passages). Most mutations mapped to the coat or major spike proteins, primarily at the coat-spike interface, and likely optimize interactions between these two proteins. Thus, while often costly in the short-term, horizontal gene transfer can ultimately enable recombinants to reach new, previously inaccessible adaptive peaks.

The exchange of structural proteins between clades

Another recent study examined HGT between clades ([Doore and Fane, 2015](#)). The major spike protein was exchanged between ϕ X174 and G4, which differ by 59% on the amino acid level. This produced two chimeras: G4- ϕ XG and ϕ X-G4G. Again, the phenotypes of the chimeric viruses differed dramatically. The assembly pathway of each microvirus consists of protein-protein interactions that optimize productive assembly and reduce the likelihood that off-pathway reactions will remove and/or sequester proteins and assembly intermediates. When a foreign protein is introduced into an assembly pathway, productive interactions may no longer be optimal. When the ϕ X174 G gene was moved into G4 (G4- ϕ XG), the resulting chimera had a single early assembly defect at low temperature: assembly intermediates simply failed to interact. In this case, a thermodynamic and/or kinetic barrier was modestly elevated: productive interactions were no longer favorable, but off-pathway reactions were still less favorable.

Conversely, when the G4 G gene was moved into ϕ X174 background (ϕ X-G4G), the defects were much more complex. In

this direction, a thermodynamic and/or kinetic barrier was elevated dramatically, making an off-pathway reaction(s) more favorable than productive assembly. The off-pathway reaction most likely involved protein H and/or an H protein-containing intermediate. Since the H protein is produced in relatively small quantities compared to the other structural proteins, these off-pathway interactions would remove all available H protein from the assembly pathway. Thus, the original chimera had to be complemented by two ϕ X174 genes, G and H. Through a series of directed selections using cloned genes, viable (plaque-forming) chimeras were ultimately isolated. In the first selection, the recovered mutations either reduced the likelihood of problematic interactions and/or suppressed the off pathway reactions involving protein H. These mutants required complementation by the ϕ X174 G gene only.

In the second selection, mutants able to utilize the foreign G4 spike protein were isolated. These mutations partially lowered the thermodynamic and/or kinetic barrier to productive assembly. Finally, in order to form plaques without complementation by either spike gene, one of two events occurred. In one scenario, the chimeric spike gene was maintained, and mutations in ribosome binding sites and intercistronic regions presumably elevated protein concentrations. Elevating the concentration of reagents can push inefficient reactions forward: the thermodynamic and/or kinetic barrier is not reduced but is overcome by brute force. Sisyphus must still push the boulder up the mountain, but at least the rock reaches the apex and rolls down the other side. Alternatively, the entirety of the foreign gene could recombine back into the genome with no additional mutations. This foreign yet wild-type protein likely interacts more efficiently than the chimeric protein, reducing the barrier further.

Early assembly – the exchange of internal scaffolding proteins

The internal scaffolding protein B has diverged significantly between microvirus clades, with sequence identities as low as 30% ([Burch and Fane, 2000a](#); [Burch et al., 1999](#)). Burch and colleagues investigated internal scaffolding protein functionality across species lines. For the most part, the internal scaffolding proteins are cross-functional, despite their low level of similarity. This flexibility makes the B protein ideal for an overlapping reading frame: the A gene can co-evolve with the host while experiencing minimal constraints from the B gene sequence. Protein A must physically associate with the host cell Rep helicase to package and synthesize ssDNA genomes. Even single substitutions in the helicase can prevent this interaction from occurring ([Ekechukwu et al., 1995](#); [Tessman and Peterson, 1976](#)).

Divergence is not uniform throughout the B protein sequences. The spacing of C-terminal aromatic amino acids, located between residues 100–120, is highly conserved. When only the C-termini are exchanged between scaffolding proteins, fitness levels are highest when the C-terminus and coat proteins share the same origin ([Burch and Fane, 2000a](#)). The x-ray structure clearly illustrates the evolutionary importance of these aromatic residues, which participate in numerous ring-ring interactions with the coat protein ([Fig. 4](#)). The pi-stacking between the B protein's last residue, F120, and three aromatic coat protein residues is particularly pronounced. Non-aromatic substitutions at this site appear to eliminate all coat protein interactions. Although non-aromatic substitutions at the other sites confer lethal phenotypes, the mutant proteins retain some level of function. After interacting with the coat protein, they kinetically trap one of three early assembly intermediates: the 9S, 9S* or 12S* particles ([Gordon et al., 2012](#)). Although each mutant protein kinetically traps different assembly intermediates, they share a group of common second site suppressors. This suggests that the internal scaffolding protein induces a single conformational

switch, rather than a series of reactions mediated by individual amino acid residues. In this model, no specific function can be ascribed to a single residue, which emphasizes the importance of interacting interfaces. Introducing additional ring–ring interactions can also have detrimental effects. The B-D111 side chain is oriented towards the center of the B protein binding pocket (Fig. 4). Thus, when mutated to tyrosine, it may be able to compete with residue F120. This mutant protein supports the assembly of defective procapsid-like particles that cannot mature into virions (Gordon et al., 2012).

These results indicate that the critical residues of the internal scaffolding protein lie in the C-terminus of the protein, which mediates non-specific binding to coat protein pentamers and induces the necessary conformational change. This is contrasted by the external scaffolding protein, which exhibits specificity for both coat protein and packaging complexes during late assembly.

Late assembly – the exchange of external scaffolding proteins

As was done for the internal scaffolding protein, initial experiments investigated the ability of the external scaffolding protein to function across clades (Burch and Fane, 2000b, 2003; Uchiyama et al., 2007; Uchiyama and Fane, 2005). Despite ~70% amino acid sequence identity, D proteins could not cross-complement (Burch and Fane, 2000b, 2003; Uchiyama et al., 2007; Uchiyama and Fane, 2005). In fact, the expression of the foreign gene mildly inhibited wild-type growth between the ϕ X174 and α 3 clades (Burch and Fane, 2000b, 2003; Uchiyama et al., 2007; Uchiyama and Fane, 2005). Multiple studies have investigated the role of three highly variable regions: α -helix #1, loop #6, and α -helix #7. These are the presumed inhibitory regions in the otherwise strongly conserved protein sequence (Burch and Fane, 2000b, 2003; Uchiyama et al., 2007; Uchiyama and Fane, 2005).

The first of these, α -helix #1, was exchanged between ϕ X174 and α 3, and the resulting chimeric D genes were cloned into plasmids (Burch and Fane, 2000b). Each gene inhibited species-specifically: ϕ X α 3D gene expression inhibited wild-type ϕ X174, whereas α 3 ϕ XD gene expression inhibited wild-type α 3. It was determined that the first α -helix recognizes the coat protein of the identical species, but productive assembly of wild-type phage was hindered by the presence of a foreign α -helix #7 or loop #6 (Burch

and Fane, 2000b, 2003). Biochemical and genetic results suggested that the chimeric proteins prevented genome packaging into procapsids, likely due to an altered external scaffolding protein lattice. Resistance mutations resided in gene A, which encodes the largest component of the packaging complex and presumably interacts with this lattice.

Further analyses were conducted with chimeric genes engineered directly within the ϕ X174 genome (Burch and Fane, 2003), placing transcription and translation under wild-type control. Chimeric α 3 loop #6 mutants were viable but displayed a cold-sensitive phenotype, which was easily suppressed by altering coat-external scaffolding protein contact sites. The chimeric α 3 α -helix #1 mutant had a recessive phenotype in co-infections. By contrast, the chimeric α 3 α -helix #7 gene exhibited a dominant lethal phenotype in co-infections. These results are consistent with the aforementioned hypotheses regarding the function of each region. Attempts to isolate mutants that could utilize either the foreign α -helix #1 or #7 chimeric proteins were not successful.

Although the above studies indicate that the first α -helices mediate the initial, species-specific contact, the interacting coat protein domain was not identified. In the x-ray structure of the ϕ X174 closed procapsid, there are very few interactions between the D protein α -helix #1 and other structural proteins (Dokland et al., 1999, 1997); however, as this structure likely represents an off-pathway product, biologically important interactions were lost. Utilizers of foreign α -helix #1 chimeric proteins may genetically identify the interacting coat domain. Since the first α -helices of the ϕ X174 and α 3 proteins were too dissimilar to isolate utilizer mutations, chimeric genes were generated between ϕ X174 and G4 (Uchiyama and Fane, 2005), which share greater sequence similarity. Genes were cloned as previously described. In this case, no inhibition of wild-type was observed, but the clone was unable to rescue a *nullD* mutant. Both assembly and DNA packaging were affected. Utilizer mutations altered the coat protein near the 3-fold axes of symmetry, which would need to be displaced to form the 3-fold related pores. The chimeric D gene was later engineered into the ϕ X174 genome, and chimeric phage exhibited delayed particle formation (Uchiyama et al., 2007). In addition, a large number of particles were non-infectious. Second-site suppressors partially restored wild-type assembly kinetics but did not correct the packaging defect. Due to its proximity in the x-ray structure, the external scaffolding protein in the D₄ position likely performs these

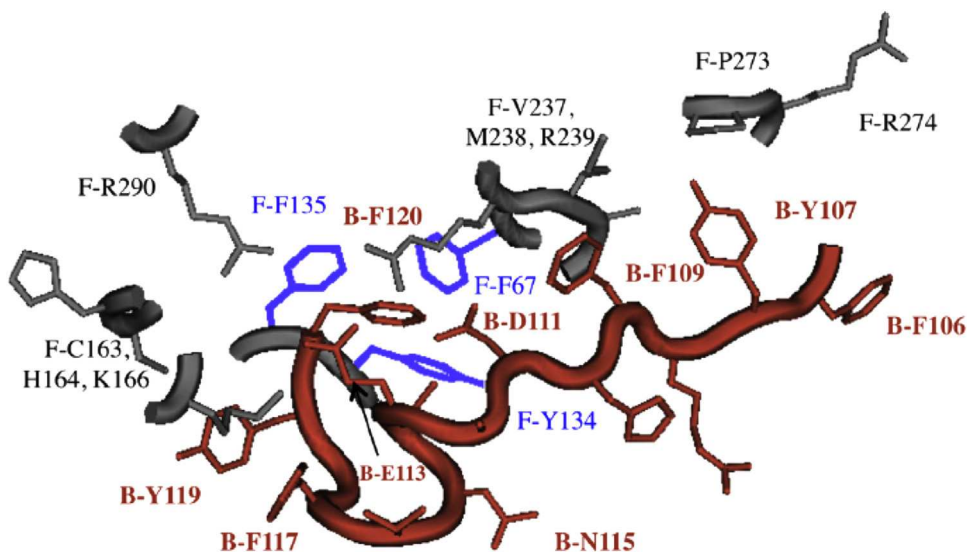


Fig. 4. Atomic structure of coat–internal scaffolding protein interactions. The internal scaffolding protein (red) forms multiple interactions with the coat protein (gray), specifically at coat protein aromatic residues (blue, adapted from Gordon and Fane (2013)).

functions. As this first α -helix interacts species-specifically, it could prevent inefficient cross-species interactions during procapsid formation in co-infected cells. This would contribute to the isolation of species during assembly, preventing the morphogenesis of particles with mixed origins.

Experimental evolution of resistance to an anti-viral protein

To arrange 12 capsomeres into an icosahedron, the 120 asymmetric D protein dimers must form a lattice that does not display $T=4$ quasi-equivalence (Caspar and Klug, 1962). A single amino acid residue facilitates the unique D protein arrangement in which one monomer in each dimer is bent by 30° . This kink is necessary to arrange the subunits within the lattice, where each monomer occupies one of four environmentally distinct positions. The critical residue, glycine 61, is located in α -helix #3. Only a glycine residue permits a 30° kink, as longer side chains would constrain flexibility.

To examine the effects of G $\rightarrow$ X substitutions, mutant external scaffolding genes were cloned and assayed for the ability to inhibit wild-type plaque formation (Cherwa et al., 2008). The induction of all mutant genes assayed (G61S, G61A, G61T, G61D, G61V, G61K, G61P) inhibited wild-type progeny production. The degree of inhibition roughly correlated with the size of the substituted amino acid: the G61A protein was a mild inhibitor, whereas the G61D, V, K, and P were strong inhibitors. By investigating assembly in the presence of the cloned G61D gene, it was determined that the inhibitory species is a heterodimer: the mutant protein requires the wild-type subunit to enter the morphogenetic pathway. When resistance mutants were isolated, most mutations resided at coat-D₃ contact sites (Cherwa et al., 2008). However, two mutations were in the internal scaffolding protein, which could affect the conformation of the coat protein in late assembly (Gordon and Fane, 2013). Both types of mutations likely exclude the inhibitory heterodimers from assembly. Regardless of mechanism, however, the resistance phenotypes of these mutants were weak.

To isolate a strongly resistant strain, wild-type ϕ X174 was experimentally evolved in the presence of increasing amounts of the G61D external scaffolding protein. After approximately 150 generations, one strain had acquired a total of four mutations and was highly resistant. These mutations mapped to the coat and external scaffolding proteins (Cherwa et al., 2009). All coat protein mutations resided under the D₃ subunit, indicating that this region may be altered to accommodate the mutant external scaffolding protein. The last mutation to arise, D-D34G, was characterized separately and did not contribute to the resistance phenotype; rather, it stimulated fitness in the presence of the inhibitory protein, albeit only when expressed in *trans*. (Cherwa and Fane, 2011). This mutation may be increasing flexibility of a neighboring subunit to counteract the lost flexibility in the G61D protein, but on its own, it may lack sufficient rigidity: i.e., it is the combination of D34G and G61D that exhibits the greatest fitness benefit. Furthermore, the D-D34G mutation exhibited allele specificity: it did not confer a significant fitness benefit with any other D protein, mutant or wild-type.

By requiring two alleles of the external scaffolding gene, the evolved strain can be likened to a three-scaffolding protein system: it requires one internal and two different external scaffolding proteins for optimal assembly kinetics. The following sections discuss the cost and benefit for three- or one-scaffolding protein assembly systems compared to a two-scaffolding system, including when and why each type of system may persist.

Increasing complexity – evolving a two scaffolding protein system into a three scaffolding protein system

For the aforementioned scenario to occur in nature, the D gene would have to be duplicated, and each copy of the gene would have to acquire separate mutations. Gene duplication is a common evolutionary mechanism (Zhang, 2003), but duplication of this gene would increase the size of the ϕ X174 genome by approximately 8.5%, surpassing the ideal packaging size (Aoyama and Hayashi, 1985; Russell and Muller, 1984). However, the α 3 genome is approximately 12% larger than the ϕ X174 genome and is stably packaged into a capsid with a very similar internal volume. This suggests that ssDNA genomes undergo selection for both the genetic content (genetic evolution) and the chemical properties of the genome itself (chemical evolution). Alternatively, a conditional three-scaffolding protein system could occur if viruses act cooperatively in a high MOI environment. In this scenario, all individuals of the population would have to be resistant to D-G61D, and individuals in sub-populations would have to harbor complementary D gene alleles. In a shared host cell, both D-G61D and D-D34G versions of the protein would be present without affecting genome size: procapsids could be assembled from this shared protein pool. Here, the volume and/or inner surface of the capsid would not have to be adjusted. However, as maximum benefit would occur in cells co-infected at an optimal ratio, this situation may not succeed or persist long-term.

Reducible complexity – evolving a two scaffolding protein system into a one scaffolding protein system

With its more ancient origin, the internal scaffolding protein may be more intrinsic to the assembly pathway. Conversely, the newer external scaffolding protein may be easier to experimentally evolve out of the system. However, structural and biochemical data strongly suggest the opposite. Structurally, there is little or no contact between neighboring 12S* capsomeres in the procapsid (Bernal et al., 2003): the entire structure is held together by interactions between D proteins. Moreover, purified 12S* particles do not interact *in vitro*, whereas purified D protein dimers readily associate (Cherwa et al., 2011a).

To evolve a one-scaffolding protein system, progressive and targeted genetic selections were employed to lessen and finally eliminate the dependence on the internal scaffolding protein B. *NullB* mutants were selected for the ability to form plaques in cells expressing B genes with progressively larger 5' deletions. After only five selections, a viable "B-free" mutant was able to form plaques without complementation (Chen et al., 2007). This mutant compensated for the loss of the internal scaffolding protein primarily by over-expressing the external scaffolding protein D gene. In addition, point mutations altered D protein-induced conformational switching and interactions with the coat protein or packaging complex. A single mutation in gene H likely affected H protein incorporation, as B protein facilitates the efficiency of this step in wild-type assembly (Novak and Fane, 2004).

When assembly kinetics were investigated in lysis-resistant cells, the B-free mutant exhibited delayed particle formation. B-free progeny were detected by 15 min post infection, whereas wild-type progeny were detectable by 5 min. By 10 min post infection, the wild-type burst size had reached its maximum, approximately 100 phage/cell. Conversely, the B-free mutant did not produce its maximum burst – a meager 10–20 phage/cell – until 40 min post infection. In lysis-sensitive cells, B-free phage would be produced by the time of programmed cell lysis: however, with such low bursts, this mutant could easily be outcompeted in a natural setting. Previous generations, which had fewer mutations, were also assayed. These mutants were not able to form

plaques but did produce particles in lysis-resistant cells. However, the lag phase before progeny detection was much longer: 30–60 min post infection. In lysis sensitive cells, programmed cell lysis would occur before particles could form.

A similar study was also conducted with the external scaffolding D gene. 5' deletions of this gene were constructed directly in the genome, resulting in N-terminal deletions within α -helix #1 (Burch and Fane, 2000b; Uchiyama et al., 2007; Uchiyama and Fane, 2005). Deletions up to seven codons were tolerated (Uchiyama et al., 2009). However, the ϕ X174 Δ D7 mutant formed very small plaques and only within a very narrow temperature range. Similar to the B-free mutant, the kinetics of ϕ X174 Δ D7 assembly was delayed and burst sizes were two orders of magnitude below wild-type levels. Second-site suppressor mutants were isolated, and mutations fell into two classes. One class affected the ribosome binding site of the external scaffolding protein, elevating intracellular protein concentrations. This would make assembly more thermodynamically favorable, as increased component concentrations can force unfavorable reactions forward. The second class of mutations affected coat-external scaffolding protein interactions, which could compensate for altered binding or conformations induced by the deletion. The suppressors shortened lag phases and elevated progeny production one order of magnitude. Larger N-terminal deletions conferred absolute lethal phenotypes. Extensive efforts were made to isolate viable ϕ X174 Δ D9 mutants, but none were recovered above a frequency of 10^{-9} . This indicates that the restoration of viability, if possible, would require multiple mutational events.

Since all Δ D7 suppressors either increased translational efficiency or altered coat-external scaffolding protein interactions, recovering compensatory mutations that would eliminate the need for the D protein seems unlikely. Furthermore, 12S* pentamers do not interact with each other in the procapsid (Bernal et al., 2003; Cherwa et al., 2011a; Ilag et al., 1995); thus, compensatory mutations would have to create novel coat-coat, coat-internal scaffolding or internal scaffolding-internal scaffolding interactions across the two-fold axes of symmetry. Lastly, the packaging complex associates with the procapsid via protein D. Thus, an elimination of the external scaffolding protein would also require novel interactions between the D-less procapsid and the packaging machinery.

The above studies indicate that the benefit of a two-scaffolding system is primarily kinetic. With two scaffolding proteins, microviruses are able to form progeny by 5 min post infection (Chen et al., 2007). This could explain why the external scaffolding gene spread throughout the microviruses (Rokyta et al., 2006). In a mixed environment with larger, dsDNA bacteriophages, microvirus progeny would appear before their competitors began to assemble. The gokushoviruses may not need two scaffolding proteins due to their host ranges. Gokushoviruses infect intracellular bacterial parasites, e.g. *Bdellovibrio* and *Chlamydia*. In this relatively isolated environment, competition from other viruses may be low, reducing the selective pressure to assemble quickly. The approximately 10 known chlamydiae are gokushoviruses, which have a 48 h replication cycle (Salim et al., 2008). The lysogenic phages of *Bacteriodes* have chosen a different lifestyle altogether. In other sub-families where scaffolding proteins have not been identified, other mechanisms of assembly may be employed. Some viruses utilize specialized coat protein regions known as delta domains (Oh et al., 2014). ssRNA viruses build their capsids around their genomes via RNA-protein interactions (Dykeman et al., 2011; Ford et al., 2013). Indeed, interactions between the capsid and the genome do play a scaffolding-like function during the final stages of ϕ X174 maturation, as discussed in the next section.

Genome packaging and specificity

During late assembly, the DNA packaging complex associates with the procapsid to concurrently synthesize and package the viral genome (Aoyama et al., 1981; Ekechukwu et al., 1995; Fujisawa and Hayashi, 1976). This complex contains the host cell Rep protein and viral proteins A and C (Aoyama et al., 1981, 1983a, 1983b; Ekechukwu et al., 1995; Mukai et al., 1979) and likely binds to the procapsid at a two-fold axis of symmetry (Ekechukwu et al., 1995). The A protein nicks the ds replicative form (RF) DNA and initiates DNA synthesis at the origin of replication (Baas et al., 1981; Eisenberg et al., 1977; Eisenberg and Kornberg, 1979; Eisenberg et al., 1976; van Mansfeld et al., 1979). Conserved across all microviruses, the origin is both necessary and sufficient for ssDNA synthesis *in vitro* and *in vivo* (Aoyama and Hayashi, 1985; Hafenstein and Fane, 2002). This ensures that only microvirus genomes are packaged. Moreover, the A protein is *cis*-acting: it preferentially binds to the DNA molecule from which it was produced (Francke and Ray, 1972). Thus, it inherently favors the replication and encapsidation of its encoding genome. At least within the α 3 clade, protein A has evolved more specific interactions with protein C, which may serve to physically isolate α 3 ssDNA synthesis and packaging in co-infected cells (Doore et al., 2014).

Both *in vitro* and *in vivo* results indicate that protein C mediates the switch from ds to ssDNA synthesis (Fig. 5A) by competing with the host cell SSB protein for binding the origin of replication (Aoyama and Hayashi, 1986; Doore et al., 2014). If SSB binds before protein C, another round of dsDNA synthesis ensues. If the converse occurs, the complex is dedicated to ssDNA biosynthesis. To investigate the specificity of A-C interactions, a cross-complementation analysis was performed using cloned C genes and *nullC* mutants. Although ϕ X174 and G4 C genes cross-complemented, α 3 is isolated in this regard (Doore et al., 2014). Thus, protein C ensures that α 3 packaging replication complexes synthesize only α 3 genomes.

The A and C gene sequences are additional examples where phylogeny may predict cross-complementation. These gene phylogenies indicate that the α 3 clade has diverged significantly from the ϕ X174 and G4 clades (Fig. 5B). In addition, the α 3-like viruses possess a more complex control mechanism for DNA synthesis. The ϕ X174- and G4-like viruses encode one C protein, 86 and 84 amino acids in length respectively. The α 3-like viruses, however, encode two C proteins: C_B (Big, 78 amino acids) and C_S (Small, 68 amino acids). The former can efficiently complement *nullC* mutants but is highly toxic to host cells. Like other microviruses, the C_B gene start codon and gene A stop codon overlap, facilitating translational coupling. However, the α 3 C_B protein is translated from a sub-optimal GTG start codon, rather than the canonical ATG. As translation from this alternate start codon is one-third as efficient (Reddy et al., 1985), this would reduce cellular C_B levels. Conversely, the shorter C_S protein complements poorly, but it is less toxic to host cells. It is translated from a standard ATG start codon, but no translational coupling arrangement exists. Instead, there is a relatively weak ribosome-binding site. The ratio and activity of these two proteins likely depends on translational efficiency, competition for ribosome binding, and the possible formation of homo- and heterodimers. The C_B gene likely arose first, as it closely resembles the genome arrangement found in other microviruses. The C_S gene may have subsequently evolved as a way to decrease the intracellular C_B concentration and hence its toxic effects.

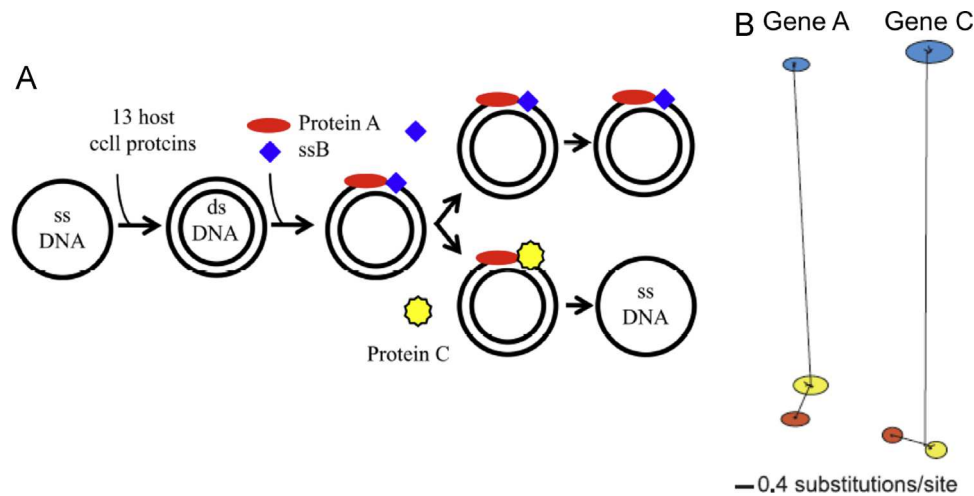


Fig. 5. (A) Model of microvirus DNA biosynthesis. Viral protein C and host cell ssB compete for binding at the origin of replication: if ssB binds first, a round of dsDNA replication ensues, whereas binding by protein C results in ss genomic DNA biosynthesis. The C protein likely gains access to the origin via specific interactions with protein A. (B) Phylogeny of microvirus genes A (left) and C (right).

Genetic and chemical evolution of microvirus genomes

Once single-stranded DNA synthesis commences, 60 copies of the DNA binding protein J associate with nascent ssDNA and enter the procapsid (Fujisawa and Hayashi, 1976). The J protein is divided into two regions: an N-terminal DNA-binding domain (domain 1); and a C-terminal hydrophobic domain (domain 2). The N-terminus binds DNA non-specifically (Jennings and Fane, 1997). The C-terminus is hydrophobic and binds a cleft in the coat protein that is occupied by the internal scaffolding B protein in the procapsid (Dokland et al., 1999, 1997; McKenna et al., 1994, 1992). During packaging, protein J displaces protein B, which undergoes autoproteolysis (Bernal et al., 2003). In the virion, the J protein tethers a portion of the genome to the capsid's inner surface. The ϕ X174 J protein is much longer than the G4 or α 3 J proteins, with an extended, more positively charged N-terminus (domain 0), which is separated from domain 1 by a proline-rich region. ϕ X174 *null*J mutants absolutely require their own J protein, whereas G4 and α 3 can utilize any DNA binding protein (Bernal et al., 2004). The G4 and α 3 genomes are considerably larger than ϕ X174, yet the capsids' inner volumes are almost identical. ϕ X174 likely evolved the larger J protein to guide the genome over a more negatively charged inner surface (Bernal et al., 2004). The biophysical nature of the capsid's inner surface may also influence the evolution of genome size; alternatively, the genome may influence the characteristics of the inner surface. This capsid-genome specificity may explain the lack of horizontal gene transfer into the ϕ X174 clade, unlike the HGT observed in other clades. Divergent members within the ϕ X174-like clade have yet to be isolated; if and when they are, these could drastically change the current phylogenies and challenge this concept.

Jennings and colleagues characterized the first two domains of the ϕ X174 DNA binding protein (Jennings and Fane, 1997). In the basic regions, lysine and arginine residues were mutated to leucine residues. The charges of the resulting proteins ranged from +4 to +9; the wild-type protein has a +12 charge. Although procapsids could be packaged with mutant +9 proteins, the location of the altered residues affected particle infectivity. When N-terminal residues 2, 4, and 5 were altered, the resulting particles could only form small plaques within a constrained temperature range. When residues 21, 23, and 25 were altered, uninfectious large particles were produced. Presumably, in both cases, the packaged DNA would be less condensed and more weakly associated with the capsid's inner surface. A second-site genetic

analysis was conducted, and suppressor mutations were of two types: intragenic mutations, which always restored basic charges; or extragenic mutations, which mapped to the first amino acid in the coat protein. This residue, a serine in the wild-type protein, mutated to either phenylalanine (F-S1F) or proline (F-S1P). This serine participates in several interactions directly atop the 3-fold axes of symmetry. Altering this residue could modify the size, strength, or flexibility of the capsid.

At least three biophysical phenomena can influence the arrangement of the packaged ssDNA genome: 1) its association with the inner capsid, which is mediated by protein J; 2) the neutralization of its phosphate backbone by basic coat and J protein residues; and 3) its secondary structure governed by H-bond base-pairing. To determine whether the relative balance of these three phenomena could influence virion properties, several biophysical characterizations were performed. Particles were filled with N-terminal charge-altered +9 mutant J proteins. Although these particles were viable, plaque size was greatly reduced (Hafenstein and Fane, 2002). The resulting particles were characterized in terms of density, surface charge, and attachment efficiency. In all assays, the mutant particles behaved significantly different from wild-type. Wild-type properties could be partially restored by the aforementioned F-S1F mutation. Similar experiments conducted with mutants that retained a +12 charge, albeit with an altered basic amino acid arrangement, yielded similar results (Hafenstein et al., 2004). In addition, particles were filled with foreign unit length DNAs (Hafenstein and Fane, 2002). Each foreign DNA filled particle exhibited distinctive properties. In all cases, protein stoichiometry was unaltered. Finally, α 3 genomes were packaged with either the ϕ X174 or α 3 J protein into α 3 capsids (Bernal et al., 2004). While these two particles exhibited dramatically different solution properties, they yielded almost identical x-ray structures. A similar phenomenon has been documented with Flock house virus (Bothner et al., 1999). This suggests that modifying the internal organization of the packaged genome can affect particle dynamics in solution, which are expressed on the capsid's external surface. These properties could influence the chemical evolution of virus particles.

Summary

Microviruses are uniquely suited for investigating molecular mechanisms during evolutionary processes. Mutational steps

during experimental evolution can be interpreted with well-defined structural, genetic, and biochemical contexts. Using cloned genes, non-viable variants can be evolved to viability. This addresses both the evolutionary trajectory and the likelihood of such a trajectory occurring in a natural environment. Furthermore, cloned genes and chimeric viruses can be used to determine which proteins in the assembly system are able to cross-complement. This provides information about entire interfaces between interacting proteins and indicates where protein-based barriers exist between species. Finally, these tools can be used to investigate the biophysical and chemical optima of viral particles.

Acknowledgments

The authors acknowledge the support of the National Science Foundation (MCB 0948399 and MCB1408217 to B.A.F.).

References

- Aoyama, A., Hamatake, R.K., Hayashi, M., 1981. Morphogenesis of phi X174: in vitro synthesis of infectious phage from purified viral components. *Proc. Natl. Acad. Sci. USA* 78, 7285–7289.
- Aoyama, A., Hamatake, R.K., Hayashi, M., 1983a. In vitro synthesis of bacteriophage phi X174 by purified components. *Proc. Natl. Acad. Sci. USA* 80, 4195–4199.
- Aoyama, A., Hamatake, R.K., Mukai, R., Hayashi, M., 1983b. Purification of phi X174 gene C protein. *J. Biol. Chem.* 258, 5798–5803.
- Aoyama, A., Hayashi, M., 1985. Effects of genome size on bacteriophage phi X174 DNA packaging in vitro. *J. Biol. Chem.* 260, 11033–11038.
- Aoyama, A., Hayashi, M., 1986. Synthesis of bacteriophage phi X174 in vitro: mechanism of switch from DNA replication to DNA packaging. *Cell* 47, 99–106.
- Baas, P.D., Teertstra, W.R., van Mansfeld, A.D., Jansz, H.S., van der Marel, G.A., Veeneman, G.H., van Boom, J.H., 1981. Construction of viable and lethal mutations in the origin of bacteriophage 'phi' X174 using synthetic oligodeoxyribonucleotides. *J. Mol. Biol.* 152, 615–639.
- Bennett, A.J., McKenna, R., Agbandje-McKenna, M., 2008. A comparative analysis of the structural architecture of ssDNA viruses. *Comput. Math. Methods Med.* 9, 183–196.
- Bernal, R.A., Hafenstein, S., Esmeralda, R., Fane, B.A., Rossmann, M.G., 2004. The phiX174 protein J mediates DNA packaging and viral attachment to host cells. *J. Mol. Biol.* 337, 1109–1122.
- Bernal, R.A., Hafenstein, S., Olson, N.H., Bowman, V.D., Chipman, P.R., Baker, T.S., Fane, B.A., Rossmann, M.G., 2003. Structural studies of bacteriophage alpha3 assembly. *J. Mol. Biol.* 325, 11–24.
- Bothner, B., Schneemann, A., Marshall, D., Reddy, V., Johnson, J.E., Siuzdak, G., 1999. Crystallographically identical virus capsids display different properties in solution. *Nat. Struct. Biol.* 6, 114–116.
- Bowes, J.M., Dowell, C.E., 1974. Purification and some properties of bacteriophage ST-1. *J. Virol.* 13, 53–61.
- Brentlinger, K.L., Hafenstein, S., Novak, C.R., Fane, B.A., Borgon, R., McKenna, R., Agbandje-McKenna, M., 2002. Microviridae, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J. Bacteriol.* 184, 1089–1094.
- Brown, C.J., Millstein, J., Williams, C.J., Wichman, H.A., 2013. Selection affects genes involved in replication during long-term evolution in experimental populations of the bacteriophage phiX174. *PLoS One* 8, e60401.
- Brown, C.J., Zhao, L., Evans, K.J., Ally, D., Stancik, A.D., 2010. Positive selection at high temperature reduces gene transcription in the bacteriophage varphiX174. *BMC Evol. Biol.* 10, 378.
- Bryson, S.J., Thurber, A.R., Correa, A.M.S., Orphan, V.J., Thurber, R.V., 2015. A novel sister clade to the enterobacteria microviruses (Microviridae) identified in methane seep sediments. *Environ. Microbiol.* 17, 3708–3721.
- Bull, J.J., Badgett, M.R., Wichman, H.A., Huelsenbeck, J.P., Hillis, D.M., Gulati, A., Ho, C., Molineux, I.J., 1997. Exceptional convergent evolution in a virus. *Genetics* 147, 1497–1507.
- Burch, A.D., Fane, B.A., 2000a. Efficient complementation by chimeric Microviridae internal scaffolding proteins is a function of the COOH-terminus of the encoded protein. *Virology* 270, 286–290.
- Burch, A.D., Fane, B.A., 2000b. Foreign and chimeric external scaffolding proteins as inhibitors of Microviridae morphogenesis. *J. Virol.* 74, 9347–9352.
- Burch, A.D., Fane, B.A., 2003. Genetic analyses of putative conformation switching and cross-species inhibitory domains in Microviridae external scaffolding proteins. *Virology* 310, 64–71.
- Burch, A.D., Ta, J., Fane, B.A., 1999. Cross-functional analysis of the Microviridae internal scaffolding protein. *J. Mol. Biol.* 286, 95–104.
- Casjens, S.R., Thuman-Commike, P.A., 2011. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology* 411, 393–415.
- Caspar, D.L., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27, 1–24.
- Caudle, S.B., Miller, C.R., Rokyta, D.R., 2014. Environment determines epistatic patterns for a ssDNA virus. *Genetics* 196, 267–279.
- Chen, M., Uchiyama, A., Fane, B.A., 2007. Eliminating the requirement of an essential gene product in an already very small virus: scaffolding protein B-free phiX174, B-free. *J. Mol. Biol.* 373, 308–314.
- Cherwa Jr., J.E., Fane, B.A., 2009. Complete virion assembly with scaffolding proteins altered in the ability to perform a critical conformational switch. *J. Virol.* 83, 7391–7396.
- Cherwa Jr., J.E., Fane, B.A., 2011. From resistance to stimulation: the evolution of a virus in the presence of a dominant lethal inhibitory scaffolding protein. *J. Virol.* 85, 6589–6593.
- Cherwa Jr., J.E., Organtini, L.J., Ashley, R.E., Hafenstein, S.L., Fane, B.A., 2011a. In Vitro Assembly of the phiX174 procapsid from external scaffolding protein oligomers and early pentameric assembly intermediates. *J. Mol. Biol.* 412, 387–396.
- Cherwa Jr., J.E., Sanchez-Soria, P., Wichman, H.A., Fane, B.A., 2009. Viral adaptation to an antiviral protein enhances the fitness level to above that of the uninhibited wild type. *J. Virol.* 83, 11746–11750.
- Cherwa Jr., J.E., Uchiyama, A., Fane, B.A., 2008. Scaffolding proteins altered in the ability to perform a conformational switch confer dominant lethal assembly defects. *J. Virol.* 82, 5774–5780.
- Cherwa Jr., J.E., Young, L.N., Fane, B.A., 2011b. Uncoupling the functions of a multifunctional protein: the isolation of a DNA pilot protein mutant that affects particle morphogenesis. *Virology* 411, 9–14.
- Chipman, P.R., Agbandje-McKenna, M., Renaudin, J., Baker, T.S., McKenna, R., 1998. Structural analysis of the Spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the Microviridae. *Structure* 6, 135–145.
- Clarke, I.N., Cutcliffe, L.T., Everson, J.S., Garner, S.A., Lambden, P.R., Pead, P.J., Pickett, M.A., Brentlinger, K.L., Fane, B.A., 2004. Chlamydia phage Chp2, a skeleton in the phiX174 closet: scaffolding protein and procapsid identification. *J. Bacteriol.* 186, 7571–7574.
- Crill, W.D., Wichman, H.A., Bull, J.J., 2000. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154, 27–37.
- Crowther, R.A., Berriman, J.A., Curran, W.L., Allan, G.M., Todd, D., 2003. Comparison of the structures of three circoviruses: chicken anemia virus, porcine circovirus type 2, and beak and feather disease virus. *J. Virol.* 77, 13036–13041.
- Dokland, T., Bernal, R.A., Burch, A., Pletnev, S., Fane, B.A., Rossmann, M.G., 1999. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus phiX174. *J. Mol. Biol.* 288, 595–608.
- Dokland, T., McKenna, R., Ilag, L.L., Bowman, B.R., Incardona, N.L., Fane, B.A., Rossmann, M.G., 1997. Structure of a viral procapsid with molecular scaffolding. *Nature* 389, 308–313.
- Doore, S.M., Baird, C.D., Roznowski, A.P., Fane, B.A., 2014. The evolution of genes within genes and the control of DNA replication in microviruses. *Mol. Biol. Evol.* 31, 1421–1431.
- Doore, S.M., Fane, B.A., 2015. The kinetic and thermodynamic aftermath of horizontal gene transfer governs evolutionary recovery. *Mol. Biol. Evol.*
- Dykeman, E.C., Grayson, N.E., Toropova, K., Ranson, N.A., Stockley, P.G., Twarock, R., 2011. Simple rules for efficient assembly predict the layout of a packaged viral RNA. *J. Mol. Biol.* 408, 399–407.
- Eisenberg, S., Griffith, J., Kornberg, A., 1977. phiX174 cistron A protein is a multifunctional enzyme in DNA replication. *Proc. Natl. Acad. Sci. USA* 74, 3198–3202.
- Eisenberg, S., Kornberg, A., 1979. Purification and characterization of phiX174 gene A protein. A multifunctional enzyme of duplex DNA replication. *J. Biol. Chem.* 254, 5328–5332.
- Eisenberg, S., Scott, J.F., Kornberg, A., 1976. An enzyme system for replication of duplex circular DNA: the replicative form of phage phi X174. *Proc. Natl. Acad. Sci. USA* 73, 1594–1597.
- Ekechukwu, M.C., Oberste, D.J., Fane, B.A., 1995. Host and phi X 174 mutations affecting the morphogenesis or stabilization of the 50S complex, a single-stranded DNA synthesizing intermediate. *Genetics* 140, 1167–1174.
- Fane, B.A., Prevelige Jr., P.E., 2003. Mechanism of scaffolding-assisted viral assembly. *Adv. Protein Chem.* 64, 259–299.
- Ford, R.J., Barker, A.M., Bakker, S.E., Coutts, R.H., Ranson, N.A., Phillips, S.E., Pearson, A.R., Stockley, P.G., 2013. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J. Mol. Biol.* 425, 1050–1064.
- Francke, B., Ray, D.S., 1972. Cis-limited action of the gene-A product of bacteriophage phiX174 and the essential bacterial site (*E. coli*-electron microscopy-cis-acting protein-specifically-nicked RF). *Proc. Natl. Acad. Sci. USA* 69, 475–479.
- Fujisawa, H., Hayashi, M., 1976. Viral DNA-synthesizing intermediate complex isolated during assembly of bacteriophage phi X174. *J. Virol.* 19, 409–415.
- Garner, S.A., Everson, J.S., Lambden, P.R., Fane, B.A., Clarke, I.N., 2004. Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydomonas reinhardtii*. *Virus Genes* 28, 207–214.
- Godson, G.N., Barrell, B.G., Staden, R., Fiddes, J.C., 1978. Nucleotide sequence of bacteriophage G4 DNA. *Nature* 276, 236–247.
- Goldstein, R., Lengyel, J., Pruss, G., Barrett, K., Calendar, R., Six, E., 1974. Head size determination and the morphogenesis of satellite phage P4. *Curr. Top. Microbiol. Immunol.* 59–75.

- Gordon, E.B., Fane, B.A., 2013. The effects of an early conformational switch defect during ϕ X174 morphogenesis are belatedly manifested late in the assembly pathway. *J. Virol.* 87, 2518–2525.
- Gordon, E.B., Knuff, C.J., Fane, B.A., 2012. Conformational switch-defective ϕ X174 internal scaffolding proteins kinetically trap assembly intermediates before procapsid formation. *J. Virol.* 86, 9911–9918.
- Hafenstein, S., Fane, B.A., 2002. ϕ X174 genome-capsid interactions influence the biophysical properties of the virion: evidence for a scaffolding-like function for the genome during the final stages of morphogenesis. *J. Virol.* 76, 5350–5356.
- Hafenstein, S.L., Chen, M., Fane, B.A., 2004. Genetic and functional analyses of the ϕ X174 DNA binding protein: the effects of substitutions for amino acid residues that spatially organize the two DNA binding domains. *Virology* 318, 204–213.
- Hayashi, M., Aoyama, A., Richardson, D.L., Hayashi, N.M., 1988. Biology of the bacteriophage ϕ X174. In: Calendar, R. (Ed.), *The Bacteriophages*. Plenum Press, New York, pp. 1–71.
- Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., Hatfull, G.F., 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. USA* 96, 2192–2197.
- Holder, K.K., Bull, J.J., 2001. Profiles of adaptation in two similar viruses. *Genetics* 159, 1393–1404.
- Hutchison 3rd, C.A., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P., Smith, M., 1978. Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* 253, 6551–6560.
- Ilag, L.L., Olson, N.H., Dokland, T., Music, C.L., Cheng, R.H., Bowen, Z., McKenna, R., Rossmann, M.G., Baker, T.S., Incardona, N.L., 1995. DNA packaging intermediates of bacteriophage phi X174. *Structure* 3, 353–363.
- Jennings, B., Fane, B.A., 1997. Genetic analysis of the phi X174 DNA binding protein. *Virology* 227, 370–377.
- Juhala, R.J., Ford, M.E., Duda, R.L., Youtton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 299, 27–51.
- Kay, D., Bradley, D.E., 1962. The structure of bacteriophage phi R. *J. Gen. Microbiol.* 27, 195–200.
- Kornberg, A., 1980. *DNA Replication*. Freeman, San Francisco.
- Krupovic, M., Forterre, P., 2011. Microviridae goes temperate: microvirus-related proviruses reside in the genomes of Bacteroidetes. *PLoS One* 6, e19893.
- Labonte, J.M., Suttle, C.A., 2013. Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Front. Microbiol.* 4, 404.
- Lee, K.H., Miller, C.R., Nagel, A.C., Wichman, H.A., Joyce, P., Ytreberg, F.M., 2011. First-step mutations for adaptation at elevated temperature increase capsid stability in a virus. *PLoS One* 6, e25640.
- Lefeuve, P., Lett, J.M., Varsani, A., Martin, D.P., 2009. Widely conserved recombination patterns among single-stranded DNA viruses. *J. Virol.* 83, 2697–2707.
- Liu, B.L., Everson, J.S., Fane, B., Giannikopoulou, P., Vretou, E., Lambden, P.R., Clarke, I.N., 2000. Molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci*. *J. Virol.* 74, 3464–3469.
- McGee, L.W., Aitchison, E.W., Caudle, S.B., Morrison, A.J., Zheng, L., Yang, W., Rokyta, D.R., 2014. Payoffs, not tradeoffs, in the adaptation of a virus to ostensibly conflicting selective pressures. *PLoS Genet* 10, e1004611.
- McKenna, R., Bowman, B.R., Ilag, L.L., Rossmann, M.G., Fane, B.A., 1996. Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications. *J. Mol. Biol.* 256, 736–750.
- McKenna, R., Ilag, L.L., Rossmann, M.G., 1994. Analysis of the single-stranded DNA bacteriophage phi X174, refined at a resolution of 3.0 Å. *J. Mol. Biol.* 237, 517–543.
- McKenna, R., Xia, D., Willingmann, P., Ilag, L.L., Krishnaswamy, S., Rossmann, M.G., Olson, N.H., Baker, T.S., Incardona, N.L., 1992. Atomic structure of single-stranded DNA bacteriophage phi X174 and its functional implications. *Nature* 355, 137–143.
- Morais, M.C., Fisher, M., Kanamaru, S., Przybyla, L., Burgner, J., Fane, B.A., Rossmann, M.G., 2004. Conformational switching by the scaffolding protein D directs the assembly of bacteriophage phiX174. *Mol. Cell.* 15, 991–997.
- Mukai, R., Hamatake, R.K., Hayashi, M., 1979. Isolation and identification of bacteriophage phi X174 prohead. *Proc. Natl. Acad. Sci. USA* 76, 4877–4881.
- Novak, C.R., Fane, B.A., 2004. The functions of the N terminus of the phiX174 internal scaffolding protein, a protein encoded in an overlapping reading frame in a two scaffolding protein system. *J. Mol. Biol.* 335, 383–390.
- Oh, B., Moyer, C.L., Hendrix, R.W., Duda, R.L., 2014. The delta domain of the HK97 major capsid protein is essential for assembly. *Virology* 456–457, 171–178.
- Pepin, K.M., Domsic, J., McKenna, R., 2008. Genomic evolution in a virus under specific selection for host recognition. *Infect. Genet. Evol.* 8, 825–834.
- Pepin, K.M., Wichman, H.A., 2007. Variable epistatic effects between mutations at host recognition sites in phiX174 bacteriophage. *Evolution* 61, 1710–1724.
- Poon, A.F., Chao, L., 2006. Functional origins of fitness effect-sizes of compensatory mutations in the DNA bacteriophage phiX174. *Evolution* 60, 2032–2043.
- Prevelige, P.E., Fane, B.A., 2012. Building the machines: scaffolding protein functions during bacteriophage morphogenesis. *Adv. Exp. Med. Biol.* 726, 325–350.
- Reddy, P., Peterkofsky, A., McKenney, K., 1985. Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. *Proc. Natl. Acad. Sci. USA* 82, 5656–5660.
- Renaudin, J., Pascarel, M.C., Bove, J.M., 1987. Spiroplasma virus 4: nucleotide sequence of the viral DNA, regulatory signals, and proposed genome organization. *J. Bacteriol.* 169, 4950–4961.
- Richmond, S.J., Stirling, P., Ashely, C.R., 1982. Virus infecting the reticulate bodies of an avian strain of *Chlamydia psittaci*. *FEMS Microbiol. Lett.* 14, 31–36.
- Rokyta, D.R., Burch, C.L., Caudle, S.B., Wichman, H.A., 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J. Bacteriol.* 188, 1134–1142.
- Rokyta, D.R., Joyce, P., Caudle, S.B., Miller, C., Beisel, C.J., Wichman, H.A., 2011. Epistasis between beneficial mutations and the phenotype-to-fitness Map for a ssDNA virus. *PLoS Genet.* 7, e1002075.
- Rokyta, D.R., Wichman, H.A., 2009. Genic incompatibilities in two hybrid bacteriophages. *Mol. Biol. Evol.* 26, 2831–2839.
- Roux, S., Krupovic, M., Poulet, A., Debroas, D., Enault, F., 2012. Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS One* 7, e40418.
- Russell, P.W., Muller, U.R., 1984. Construction of bacteriophage luminal diameterX174 mutants with maximum genome sizes. *J. Virol.* 52, 822–827.
- Sackman, A.M., Rokyta, D.R., 2013. The adaptive potential of hybridization demonstrated with bacteriophages. *J. Mol. Evol.* 77, 221–230.
- Sackman, A.M., Reed, D., Rokyta, D.R., 2015. Intergenic incompatibilities reduce fitness in hybrids of extremely closely related bacteriophages. *PeerJ* 3, e1320.
- Salim, O., Skilton, R.J., Lambden, P.R., Fane, B.A., Clarke, I.N., 2008. Behind the chlamydial cloak: the replication cycle of chlamydiaophage Chp2, revealed. *Virology* 377, 440–445.
- Sertic, V., Bulgakov, N., 1935. Classification et identification des typhi-phage. *C. R. Soc. Biol. Paris* 119, 1270–1272.
- Sun, L., Rossmann, M.G., Fane, B.A., 2014a. High-resolution structure of a virally encoded DNA-translocating conduit and the mechanism of DNA penetration. *J. Virol.* 88, 10276–10279.
- Sun, L., Young, L.N., Zhang, X., Boudko, S.P., Fokine, A., Zbornik, E., Roznowski, A.P., Moulineux, I., Rossmann, M.G., Fane, B.A., 2014b. Icosahedral ϕ X174 forms a tail for DNA transport. *Nature* 505, 432–435.
- Taketo, A., 1976. Host genes involved in the replication of single-stranded DNA phage phiK. *Mol. Gen. Genet.* 148, 25–29.
- Tessman, E.S., Peterson, P.K., 1976. Bacterial rep- mutations that block development of small DNA bacteriophages late in infection. *J. Virol.* 20, 400–412.
- Tsao, J., Chapman, M.S., Wu, H., Agbandje, M., Keller, W., Rossmann, M.G., 1992. Structure determination of monoclinic canine parvovirus. *Acta Crystallogr. B* 48, 75–88.
- Tucker, K.P., Parsons, R., Symonds, E.M., Breitbart, M., 2011. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J.* 5, 822–830.
- Uchiyama, A., Chen, M., Fane, B.A., 2007. Characterization and function of putative substrate specificity domain in microvirus external scaffolding proteins. *J. Virol.* 81, 8587–8592.
- Uchiyama, A., Fane, B.A., 2005. Identification of an interacting coat-external scaffolding protein domain required for both the initiation of phiX174 procapsid morphogenesis and the completion of DNA packaging. *J. Virol.* 79, 6751–6756.
- Uchiyama, A., Heiman, P., Fane, B.A., 2009. N-terminal deletions of the phiX174 external scaffolding protein affect the timing and fidelity of assembly. *Virology* 386, 303–309.
- van Mansfeld, A.D., Langeveld, S.A., Weisbeek, P.J., Baas, P.D., van Arkel, G.A., Jansz, H.S., 1979. Cleavage site of phiX174 gene-A protein in phiX and G4 RFI DNA. *Cold Spring Harb. Symp. Quant. Biol.* 43 (Pt 1), 331–334.
- Wichman, H.A., Badgett, M.R., Scott, L.A., Boulianne, C.M., Bull, J.J., 1999. Different trajectories of parallel evolution during viral adaptation. *Science* 285, 422–424.
- Wichman, H.A., Millstein, J., Bull, J.J., 2005. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170, 19–31.
- Zhang, J.Z., 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298.
- Zhang, W., Olson, N.H., Baker, T.S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J.W., McKenna, R., 2001. Structure of the maize streak virus geminate particle. *Virology* 279, 471–477.
- Zlotnick, A., Fane, B.A., 2010. Mechanisms of icosahedral virus assembly. In: Agbandje-McKenna, M., McKenna, R. (Eds.), *Structural Virology*. Royal Society of Chemistry, London, pp. 180–202.