DNA PACKAGING AND EJECTION BY P22-LIKE BACTERIOPHAGE

by

Justin Conrad Leavitt

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

The University of Utah

August 2016

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Justin Conrad Leavitt**

has been approved by the following supervisory committee members:

Sherwood R. Casjens	, Chair	**2/29/2016** Date Approved
Constantine P. Georgopoulos	, Member	**2/29/2016** Date Approved
John S. Parkinson	, Member	**2/29/2016** Date Approved
David F. Blair	, Member	**2/29/2016** Date Approved
Michael S. Kay	, Member	**2/29/2016** Date Approved

and by **M. Denise Dearing** , Chair/Dean of

the Department/College/School of **Biology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

This work analyzes the DNA translocating mechanisms by the "P22-like" bacteriophages *Salmonella enterica* phage P22 and *Shigella flexneri* phage Sf6. DNA packaging into virions during assembly and ejection of the DNA into target cells is examined. To explore DNA packaging, interactions of the terminase proteins (TerS and TerL) and their role in recognizing the correct DNA to be packaged are explored experimentally *in vivo*. Evidence was obtained showing that the N-terminal domain of TerS is involved in the recognition of bacteriophage DNA, and the C-terminal domain of TerS likely interacts with TerL. The P22 *pac* site was shown to be necessary and sufficient to induce DNA packaging. Additionally, the Sf6 *pac* site was identified and shown to be similar to the P22 *pac* site. The phenotypes of a *pac*-null P22 mutant and mutations that alter TerS were analyzed. For the first time, the *pac* site was shown to be essential for P22 lytic growth. Mutations altering the TerS protein that affect *pac* site recognition were shown to be located in various locations on the protein. Additionally, a mutation in *terL* was also identified that allows normal progeny phage production in the absence of a functional *pac* site. This is the first evidence that TerL can work in conjunction with TerS to enable DNA recognition. In our studies of DNA ejection from the virion, a number of insights were obtained. The tail needle (gp26 protein) is the plug that keeps the DNA inside the complete virion, while the ejection proteins (gp7, gp20, and gp16) are required to deliver the DNA into the cytoplasm after gp26 release. We found that alterations of the C-terminal protein domain at the tip of the tail needle can have a strong affect on the rate at which DNA is released into the host cytoplasm, indicating that it may serve as a secondary switch controlling DNA delivery. We also examined the numbers and locations of the three different ejection proteins present in the complete P22 virion and found that they occupy space that is also available to the DNA inside the viral head.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I am sincerely appreciative to all of my family and friends who offered encouragement throughout this entire endeavor. I must also thank all the members of the Sherwood Casjens lab who contributed to an invigorating intellectual atmosphere. Specifically, Lasha Gogokhia, Alexandra Heitkamp, Kassandra Wilson, and Eddie Gilcrease each contributed direct practical assistance. Anshul Bhardwaj and Gino Cingolani at Thomas Jefferson University were excellent and productive collaborators. Additionally, my dissertation committee provided very intelligent guidance, and have each taught me the meaning of scientific excellence through word and example. Lastly, I am grateful to my advisor, Dr. Sherwood Casjens, who has been a most patient teacher and has had a profound influence on my development as a scientist.

CHAPTER 1

INTRODUCTION TO DNA PACKAGING AND INJECTION BY TAILED BACTERIOPHAGES

Tailed bacteriophage DNA transactions

It is estimated that there are approximately 10^{31} virus particles on Earth (Suttle, 2005), with the vast majority being tailed, double-stranded (ds) DNA bacteriophages. It is also estimated that global bacteriophage infection occurs at a rate of approximately 10^{24} individual infections per second (Hendrix, 2003). These large numbers suggest that bacteriophages play an important role in our global ecosystem. Despite the staggering abundance and apparent diversity of tailed dsDNA phages, there are common traits widely shared among all such bacteriophages (reviewed by Grose and Casjens, 2014). For example, capsid assembly and DNA packaging strategies are quite well conserved among all the tailed bacteriophages that have been examined so far. Other features, such as regulatory and lysis mechanisms, are more varied among the dsDNA tailed bacteriophages. This conservation enables biologists to generalize the themes discovered through the detailed study of a few "classic" model system bacteriophages. For example, control of gene expression by proteins like the bacteriophage lambda prophage repressor, transduction by bacteriophages like P22, and DNA injection by bacteriophages like T4 are examples of biological processes actually happening in nature on a nearly incomprehensible scale.

All virus particles have an apparently simple function, to protect and deliver their nucleic acid code that is needed to produce more viruses when a new host is infected. Thus, all viruses must have a means of packaging the nucleic acid inside the viral capsid, and safely delivering this cargo to another viable host. These two interconnected functions, DNA packaging and delivery, are among the most poorly understood aspects of the tailed bacteriophage life cycle, in spite of the fact that they are necessarily utilized by all tailed bacteriophages. In this work, these two processes will be studied.

DNA packaging

All tailed bacteriophages and other large dsDNA viruses that have been examined package DNA into a preformed capsid shell (called a procapsid) using an ATP-cleavage driven DNA translocating motor (reviewed by Casjens and Hendrix, 1988; Casjens, 2011; Rao and Feiss,

2008). This motor, often referred to as "terminase", not only packages the DNA into the preformed shell, but also contains a nuclease that creates the ends or termini of the intravirion DNA during the packaging process. The terminase complex typically consists of two types of subunits: TerL (terminase large subunit), and TerS (terminase small subunit). TerL is the ATPase that powers the motor that pumps the DNA into the procapsid and contains the nuclease domain that cuts the DNA to virion length. In contrast, TerS is responsible for recognition of the DNA to be packaged and works in concert with TerL to cut and package the DNA. There are a small number of known variations of this general strategy, one of the most notable being bacteriophage ø29 which has only the TerL packaging protein that serves as the ATPase, a terminal protein that is covalently bound to the ends of ø29's linear DNA and a structural "packaging" RNA that are essential for packaging (Guo et al., 1987b).

All tailed bacteriophages have a "portal protein" that forms a dodecameric ring at one of the icosahedral vertices of the head. DNA enters the capsid through this vertex structure, and tails attach to it after DNA is packaged. Atomic structures are known for several tailed bacteriophage portal proteins, and they have a common fold in spite of the fact they do not always have recognizable amino acid (AA) sequence similarity (reviewed by Casjens, 2011). The terminase is thought to bind to the portal protein ring during packaging, and in that sense, the portal protein is part of the packaging motor, but it is thought not to be involved in power-generation. Figure 1.1 shows an artist's depiction of the motor. In addition to forming the hole in the capsid through which DNA enters and being the anchor point for terminase, genetic evidence implicates portal protein in the decision to stop the motor when the head has been filled with DNA (Casjens et al., 1992b).

Many details of the packaging motor remain poorly understood. The somewhat atypical ø29 motor has worked best *in vitro* with purified components and therefore has been studied in the most detail. In particular, optical tweezer and biochemical experiments have shown that approximately 2 bp are packaged per ATP cleaved (Guo et al., 1987a; Morita et al., 1993), and packaging proceeds in major kinetic steps of about 10 bp (about one helix turn) with four fast 2.5

bp sub-steps within each major step (Moffitt et al., 2009). *In vitro*, the ø29, lambda, and T4 motors package DNA at different maximum rates, of 180, 700, and 1800 bp/sec, respectively (reviewed by Casjens, 2011). Other bacteriophages have not been studied in this regard. In spite of these detailed measurements on the whole motor, the TerL and TerS subunit arrangements and stoichiometry during packaging are not known for any tailed bacteriophage, and the detailed mechanism of force generation for packaging remains mysterious. In addition, herpesviruses appear to possess portal and terminase proteins with functions similar to those of the tailed bacteriophages (Rixon and Schmid, 2014). Finally, in no case is the detailed mechanism of recognition of DNA for packaging by TerS understood.

DNA injection

In order to infect a bacterial cell, the bacteriophage virion must first recognize the correct host and then provide the needed machinery to enable the viral DNA to transverse the cell's normal barriers into the cytoplasm (reviewed by Casjens and Molineux, 2012). Adsorption to a host is typically mediated through a specific "receptor" molecule on the outside surface of the target host cell that positions the virus at the correct location for DNA delivery. When the correct position is achieved, the bacteriophage particle is triggered to release its DNA into the cell. The DNA inside virion is packed very tightly at several hundred times its density in the cell cytoplasm (Casjens, 1997). Since the packaged DNA is devoid of bound proteins that could potentially stabilize its compaction, the repulsive negative phosphate charges, resistance to bending, and partial dehydration all contribute to the compacted DNA being in a very high energy state (Lander et al., 2013; Qiu et al., 2011; Rau et al., 1984). Therefore, the DNA is thought to be present in the virion as a tightly coiled "spring" at a very high "DNA pressure". It might be expected that such tightly packed DNA would be spontaneously released with great force when triggered, and indeed, the DNA molecule exits rapidly when virions are opened *in vitro* (*e.g.*, Van Valen et al., 2012). However, several observations suggest that this is an oversimplified view. First, as the DNA comes out of the capsid during injection, the DNA that remains inside the virion

gets shorter, so the pressure decreases drastically as injection nears completion; furthermore, turgor pressure inside the bacterium should push back against DNA entry. Thus, calculations suggest that at least getting the last half of the virion DNA into the cell may require something beyond its stored internal kinetic energy. As a consequence, it has been argued that DNA injection cannot be accounted for purely by internal virion DNA pressure (Molineux, 2001, 2006; Molineux and Panja, 2013).

The events following the initial release from the virion are not well known either, and have been problematic to study since the process is very rapid. Additionally, the proteins involved are also virion structural proteins so most mutations that affect them block assembly and cannot be used to study their role in DNA release. In Gram negative bacteria, the DNA transverses though the outer membrane, cell wall, and plasma membrane to enter the bacterial cytoplasm. Bacteriophage T4 has been a model for the ejection process (reviewed by Leiman et al., 2004), but it is certainly not a suitable morphological example for all tailed bacteriophages. In the *Myoviridae* and *Siphoviridae* families, the tail serves as a tube or conduit to deliver the DNA to the cell. In the *Myoviridae*, the outer sheath of the tail shaft actually contracts, thereby physically forcing the central tail tube through the outer membrane and cell wall and presumably into the inner membrane so DNA can be directly deposited into the cytoplasm. Injection is much less well understood for the *Siphoviridae* family, whose long tails do not contract, and for the *Podoviridae* such as P22 and T7 that have very short tails. Little is known about how their DNA traverses the membranes and cell wall.

Recent electron cryo-microscopic studies of short-tailed phage T7 (Hu et al., 2013) have shown that after initial adsorption, a long thin structure (a tube?) forms below the tail that projects through the periplasm to the inner membrane. The authors of these studies speculate that proteins released from the virion serve to build a periplasmic structure that acts as a conduit to deliver the virion DNA to the cytoplasm. It has been shown in the case of T7 that multiple molecules of several "minor" virion proteins are ejected into the cell with the DNA (Chang et al., 2010; Moak and Molineux, 2000), but it has not been shown that these proteins (called "ejection

proteins") actually form the observed periplasmic structure. How these proteins are ejected from the virion, how they might be involved in the formation of this complex, as well as how this whole process might be triggered remains unknown. In addition, it is not known what the specific functions of the different ejection proteins are in any podophage. How are ejection proteins recruited to the virion during assembly? Where are they in the virus structure? How are they released from the virion? Do they form a DNA translocating complex and if so how? Are other host factors involved in the process? Clearly, much remains to be learned about tailed bacteriophage DNA injection.

Historical background of phage P22

The focus of this work is on bacteriophage P22 and its close relatives. P22 is a short-tailed bacteriophage belonging to the *Podoviridae* family. It is characterized by a short noncontractile tail. Bacteriophage P22 is a well-studied member of this family group that has historic importance in the development of bacterial genetics. P22 was isolated by Zinder and Lederberg (1952) and was the first bacteriophage found to have the ability to perform generalized transduction, a process by which P22 can package host rather than its own DNA and inject that DNA into another cell where it can be permanently incorporated into the resident genome (Ebel-Tsipis et al., 1972a; Ebel-Tsipis et al., 1972b; Susskind and Botstein, 1978a). This property was very useful in the "early days" of molecular biology, as it allowed the transfer of DNA markers between different bacterial host strains. Although P22 has been used for decades as a tool in molecular biology, the precise mechanisms of virion assembly, DNA packaging, and the delivery of DNA into a new host cell are still not completely understood.

P22 infects strains of *Salmonella enterica* Serovar Typhimurium (also sometimes known as *Salmonella typhimurium*), and it uses the cell surface O-antigen polysaccharide as a receptor (Lindberg et al., 1970). Closely related bacteriophages infect different serovars of *Salmonella* as well as other Gram negative bacteria, including members of the *Escherichia, Shigella, Serratia*, and many others in the family *Enterobacteriaceae* (Casjens and Thuman-Commike, 2011).

The genetic organization of P22

The organization of the P22 genome is typical of other temperate members of the lambda-like or "lambdoid" group of bacteriophages. The genome has been completely sequenced (Pedulla et al., 2003) and is 41724 bp in length (see Appendix for map). The genetics of P22 and its basic life cycle are well understood, and these early studies have been thoroughly reviewed by Susskind and Botstein (1978a) and Hendrix and Casjens (2006). The genes in the divergent early operons (expressed early in the lytic cycle) encode the functions involved in control of phage gene expression, DNA replication and recombination, as well as host cell cycle inhibition. The genes in the late operon encode the proteins involved in cell lysis and progeny virion assembly. In addition to these operons, several other functions relevant to this report are encoded by other transcripts as follows: (i) the prophage repressor that keeps lytic functions turned-off in the lysogen, (ii) the "immunity I" region or antirepressor module (the adjacent *mnt*, *arc,* and *ant* genes) encodes and controls the expression of an antirepressor that binds to and inactivates lambdoid bacteriophage repressors (including that of P22) (Susskind and Botstein, 1975, 1978b), (iii) the two super-infection exclusion genes, *sieA* and *sieB*, encode functions that block infection of a P22 lysogen by other bacteriophages (Susskind et al., 1974a; Susskind et al., 1971, 1974b), and (iv) three genes, *gtrABC*, that modify the host's O-antigen polysaccharide (Vander Byl and Kropinski, 2000).

The single transcriptional promoter of the late operon controls the expression of all the P22 virion assembly genes and thus the timing of progeny bacteriophage appearance in uninfected cells (Roberts et al., 1976). P22 early right operon gene *23*, which is nearly identical to the well-studied bacteriophage lambda gene *Q* (Guo and Roberts, 2004; Strobel and Roberts, 2014), is a transcriptional antiterminator that allows late transcription by specifically blocking the premature termination of the late operon leader RNA transcript, thus enabling the transcription of the downstream genes.

P22 virion assembly

There are thirteen genes in the portion of the late operon that encodes the virion assembly proteins. Of these, twelve have been shown to be essential for functional bacteriophage P22 virion assembly. Botstein, King, and coworkers isolated conditional lethal (nonsense, temperature sensitive, and cold sensitive) mutations to both genetically identify these genes, and to examine the bacteriophage structures that are assembled under nonpermissive conditions (Botstein et al., 1973; King et al., 1973; Poteete and King, 1977). This large body of work (and parallel work on bacteriophages such as T4, lambda, P2, and T7) gave rise to the idea that virion proteins do not assemble in random order, but rather, virions are built by specific "assembly pathways" in which the different proteins assemble onto the growing virion in a very specific order. Figure 1.2 shows the bacteriophage P22 assembly pathway (reviewed in Casjens and Weigele, 2003; Casjens and Thuman-Commike, 2011). The P22 procapsid contains six different proteins, specifically, 415 molecules of the coat protein (gp5 - nomenclature in this field names the proteins after the gene that encodes them, in this case gene product 5) which build the icosahedral shell of the virion, and 12 molecules of portal protein (gp1) which form the channel or portal through which DNA enters during packaging and leaves during injection; about 250 molecules of scaffolding protein (gp8) fill the interior during assembly, but all of them exit the structure (without being degraded) to make room for DNA during the packaging process, and a few molecules of each of three different ejection proteins are present in the virion. The latter three proteins, the products of genes *7, 16* and *20*, are present in small numbers (see Chapter 5), and are required for successful delivery of DNA from the virion into the cytoplasm of the cell that is being infected. The terminase recruits the DNA and associates with the procapsid. Presumably, this association is with the portal protein, but this has not been shown directly. The DNA is then pumped into the capsid. After packaging is completed, the terminase leaves the structure and the short tail is built on the portal vertex of the structure. During the DNA packaging process, the coat protein undergoes a major conformational change that stabilizes the shell and expands its diameter by about 11%. Twelve molecules of gp4, 6 molecules of gp10, 3 molecules of gp26, and 18

molecules of gp9 are added sequentially to build the tail and complete the virion. The tailspikes, which bind to the surface O-antigen polysaccharide to initiate DNA delivery into cells, are present as six gp9 trimers (Lander et al., 2006; Lindberg et al., 1970; Steinbacher et al., 1996).

Structure of the P22 virion

The complete P22 virion has an approximate diameter of 60 nm and is filled with about 43400 bp of dsDNA. P22 virion coat protein shells possess an icosahedral T=7 *levo* structure. The structure of the complete virion has been determined to a resolution of 7.6 Å using cryo-electron microscopic reconstruction (Chang et al., 2006; Lander et al., 2006; Tang et al., 2011). The dodecameric portal protein ring is present at one of the five-fold vertices, occupying a position analogous to one of the coat protein pentamers that are present at the other icosahedral vertices in the virion. The tail is built on the portal protein and does not contact the coat protein portion of the shell. In addition, the coat protein structure has been determined to 3.4 Å by icosahedrally averaged cryo-electron microscopy and nmr (Chen et al., 2011; Parent et al., 2010; Rizzo et al., 2014), and x-ray structures are known for four of the portal vertex and tail proteins (gp1, gp4, gp26, and gp9) (Olia et al., 2007; Olia et al., 2011; Steinbacher et al., 1997; Steinbacher et al., 1994). Thus, the protein part of the P22 virion is among the best understood of the large viruses. It is notable, however, that there is no electron density in the cryo-electron microscopic virion structure that accounts for the locations of the three ejection proteins (see Chapter 5), so their location in the virion is not known.

DNA is packed very densely in the particle, at about the density of crystals of short dsDNA fragments (Earnshaw and Casjens, 1980). Cryo-electron microscopic virion reconstruction indicates that the DNA is present as a solenoid whose central axis is aligned with the bacteriophage tail. The DNA appears to be more ordered near the portal vertex and against the inside of the coat protein shell and becomes less ordered toward the center of the virion (Chang et al., 2006; Lander et al., 2006; Tang et al., 2011).

P22 DNA packaging

During P22 life cycle, DNA is replicated by a rolling circle mechanism that results in the synthesis of long concatemers of head-to-tail repeats of the P22 genome sequence (Tye et al., 1974b and references therein). This newly synthesized concatemer is the substrate for P22 DNA packaging. Two proteins, the products of P22 genes *2* (TerL) and *3* (TerS), are necessary (and thought to be sufficient) for P22 DNA packaging into procapsids. Null mutants in either of these genes cause the infected cell to accumulate concatemeric DNA and procapsids (Botstein et al., 1973; King et al., 1973).

Although the packaging process is undoubtedly dynamic, we do have anecdotal "snap shots" of different structures and intermediates. The DNA-empty P22 capsid has been purified with bound TerL and TerS (Poteete and Botstein, 1979), but this intermediate is unstable, so the exact composition, structure, and subunit stoichiometry of the assembled terminase motor are unknown. The P22 TerS is a 162 AA protein that forms a stable 9mer (Roy et al., 2012). It binds DNA nonspecifically *in vitro* (Nemecek et al., 2008; Roy et al., 2012), yet has been shown to be responsible for *pac* specificity *in vivo* (Casjens et al., 1987; Casjens et al., 1992a; Jackson et al., 1982). The atomic structure of a TerS 9-mer has been solved (Roy et al., 2012). P22 TerL is a monomer in solution (Nemecek et al., 2007), and like other TerLs contains two separate domains, a C-terminal nuclease and an N-terminal ATPase (Roy and Cingolani, 2012). The atomic structure of whole P22 TerL is not known; however, the structure of the moderately closely related TerL from P22-like bacteriophage Sf6 is known (Zhao et al., 2013), as is the structure of the nuclease domain of P22 TerL (Roy and Cingolani, 2012). Ribbon diagrams of these structures are shown in Figure 1.3. P22 TerS and TerL form a complex in solution (Poteete and Botstein, 1979; Roy et al., 2012; H. Brown and S. Casjens, unpublished). Recently, McNulty *et al*. (2015) purified an *in vitro*-formed complex of TerS and TerL and found that one oligomer (most likely a 9-mer) of TerS is bound to two TerL subunits, and they obtained a low-resolution cryo-electron microscopic reconstruction of this complex. Although these structures may provide us with hints as to the detailed mode of action of terminase, the precise mechanisms and physiological

intermediates for the packaging motor remain unknown. In order to understand the DNA packaging process, recent efforts in the "model system" bacteriophages have focused on specific aspects of the process such as the bio-energetics and kinetics of packaging for bacteriophages ø29, T4, and lambda and DNA cleavage by bacteriophage lambda TerL (reviewed by Rao and Feiss, 2008). Recognition of DNA for packaging is very poorly understood in all bacteriophages. It is not known if the TerS protein oligomeric structures obtained are physiologically correct, the numbers of TerS oligomers and TerL monomers that are needed for correct DNA recognition and packaging are not known, and the manner in which TerS and TerL interact also remains unknown. Thus, we are left with only hypothetical models for the motor's structure and action.

P22 is a headful packaging bacteriophage. It is so named because the length of DNA packaged is determined by the available space inside the capsid. Alterations in the length of the DNA sequence do not affect the length of the DNA molecules that are packaged (Casjens and Hayden, 1988; Tye et al., 1974a; Tye et al., 1974b), whereas changing the capsid size (Moore and Prevelige, 2001) or placing additional protein molecules inside the capsid (Weigele et al., 2005) do affect the length of the packaged DNA. Packaging is initiated by a DNA cleavage near a specific site called *pac* (Jackson et al., 1978). After the initiation cleavage at *pac,* one of the DNA ends thus created is threaded into the procapsid through the portal vertex and packaging proceeds unidirectionally (rightward on the standard map) until a signal that the head is full triggers packaging termination, and a "headful" cleavage is made between the packaged DNA and the remaining unpackaged portion of the concatemer. This results in more than a complete genome being packaged into the virion, and this terminal redundancy allows recircularization by homologous recombination when a subsequent infection occurs. Once packaging is initiated on a concatemer, this same DNA molecule can be used for up to ten sequential packaging events, with each subsequent event starting where the previous event ended with a headful cleavage (Adams et al., 1983; Casjens and Hayden, 1988; Jackson et al., 1978). Thus, one *pac* site-specific packaging initiation event programs a series of packaging events on a single substrate DNA molecule. Figure 1.4 diagrams several packaging events on a concatemer. The wild type P22

genome is 41724 base pairs in length (Pedulla et al., 2003), and approximately 43400 base pairs are packaged into one headful, resulting in 4% more DNA packaged than is contained in one genome sequence (Casjens and Hayden, 1988). A headful sensor, most likely a portal protein function (Casjens et al., 1992b), triggers TerL cutting of the DNA when the head is full.

P22 TerS is responsible for recognition of the DNA packaging site. This was deduced from the study of high frequency of transduction (HT) mutants isolated by Schmieger (Raj et al., 1974; Schmieger, 1972). Casjens et al. (1987; 1992a) showed that these mutations alter the TerS protein. Their altered or lowered *pac* specificity strongly indicates that TerS recognizes DNA *in vivo*. The location of the DNA cleavage(s) that occurs near the *pac* site during initiation of P22 packaging *series* was found to be inside the *3* (*terS*) gene (Backhaus, 1985; Casjens et al., 1987), and the initiation cleavage was found to be imprecise in that in different series, cleavage occurs at different sites within an approximately 120 bp region surrounding the *pac* site, with more frequent cutting occurring at several locations within this region that are separated by about 20 bp (Casjens and Huang, 1982; Casjens et al., 1987; Casjens et al., 1992a). The P22 *pac* site itself was genetically identified by placement of an artificial second *pac* site in the P22 genome and examining its functionality. Site-directed modification of this second *pac* site, once its location was known, and restriction enzyme cleavage analysis to determine the ratio of utilization of the wild type and mutant *pac* sites determined which bps in the *pac* site are critical to its function (Wu et al., 2002). The *pac* site is an approximately 21 bp sequence that lies near the center of the 120 bp cleavage region inside gene *3*.

P22 DNA injection

The six trimeric P22 tailspike proteins bind to O-antigen polysaccharide and possess a catalytic activity that cuts the serovar Typhimurium O-antigen (Berget and Poteete, 1980; Iwashita and Kanegasaki, 1976; Muller et al., 2008). Successive cleavages are thought to bring the virus progressively closer to the cell's outer membrane and orient the particle with its tail pointed toward the bacterial surface (Casjens and Molineux, 2012). Gp26 is a 233 AA protein that

forms the trimeric "tail needle" that extends past the distal tips of the tailspikes (Berget and Poteete, 1980; Lander et al., 2006; Olia et al., 2007). As the virion moves toward the surface, the N-terminal distal tip of the tail needle should be the first part of the virion to make contact with the outer membrane. The C-terminal end of the gp26 trimer, also called the tail needle, serves as a plug to keeping the highly pressured nucleic acid inside the virion (Bauer et al., 2015; Berget and Poteete, 1980). Thus, Gp26 minus particles package DNA and procapsids are expanded, but the packaged DNA is unstable and is not held in the virion. It is unknown how the triggering of tail needle release is orchestrated, but it is known that it is released from the virus particle during the infection process (Israel, 1977).

The later events during P22 infection are even more mysterious. It is known that defective mutations of gp7, gp20, and gp16 produce apparently normal-appearing but noninfectious virus particles, while the DNA is packaged normally. Gp7, gp16, and gp20 were identified as P22 virion proteins, but their locations are not known and the numbers of molecules per virion were only crudely estimated to be between 6 and 20 (Casjens and King, 1974). Genes *7, 20,* and *16* are clustered together on the P22 genome, suggesting that they may interact with each other, but nothing is known concerning the biophysical and biochemical properties of the ejection proteins. Particles missing any one of the ejection proteins absorb to the host and release the DNA from the virion, but the DNA does not enter the cytoplasm (Botstein et al., 1973; King et al., 1973). Currently, the most widely favored model is that these ejection proteins form a periplasmic structure (above) that delivers the DNA from the virion into the cytoplasm. In support of this model, P22 gp16 from a wild type bacteriophage has been shown to possess the ability to rescue other *16* minus particles (Hoffman and Levine, 1975a, b). In contrast, gp16 expressed from a plasmid cannot rescue defective particles in the same way (Umlauf and Dreiseikelmann, 1992). This indicates that gp16 can diffuse away from the virion that delivered it to form a structure in the periplasm that can be used by other infecting P22 virions.

The mechanism by which the ejection proteins are recruited to the virion during assembly is unknown, but they are present in the procapsid, and therefore are placed in the virion before

DNA is packaged. All three are still recruited to procapsids in infections by bacteriophages with nonsense mutations in the portal protein gene (Bazinet and King, 1988; Botstein et al., 1973; King et al., 1973), so portal protein, like DNA, is *not* responsible for their recruitment. On the other hand, several mutations in scaffolding protein disrupt the recruitment of gp16. If gp16 molecules are bound to scaffold during their incorporation, they must be released to remain in procapsids after scaffold leaves. All ejection proteins were reported to be recruited independently of one another since nonsense mutations in each one do not block the incorporation of the other two (Botstein et al., 1973; King et al., 1973; Poteete and King, 1977). However, the more recent discovery that N-terminal *amber* fragments of gp20 (Adhikari and Berget, 1993) and gp7 (E. Gilcrease and S. Casjens, unpublished results) are assembled into P22 capsids opens the possibility that incorporation of each of the three ejection proteins may not be independent of the others but is instead dependent on the N-terminal regions of the others.

After their initial recruitment to the assembling virion, the ejection proteins' organization may be altered. One would expect them to potentially interact with each other, portal protein, the DNA, or all of the above in order to exit the virion correctly and enable infection. For example, Israel (1977) reported that particles lacking gp7 or gp20 are able to eject the other two proteins, but particles lacking gp16 do not eject gp7. It is possible to imagine many organizational strategies that could explain all the current data and necessary functions of the ejection proteins; however, further study is required to understand the exact recruitment, assembly, and exit of these proteins. These proteins are likely to be subject to very extreme conformational changes and function (recruitment, being in concentrated DNA, passage through the portal channel, and membrane insertion?). For this reason, they are very interesting not only due to their macromolecular versatility but, also, for this same reason, have been extremely difficult to study.

In vitro studies using purified lipopolysaccharide (lipid conjugated to a core oligosaccharide with the long O-antigen chains attached) to trigger DNA release from P22 virions has been used as a powerful new tool to study DNA release (Andres et al., 2010). This allows the study of ejection without the complexity of an entire cell and subsequent viral production. The ejection

proteins are released from the virion with the DNA in this system, and evidence shows that ejection proteins can be released under conditions (particular polyethyleneglycol concentration and temperature) where the DNA is not released (Jin et al., 2015). The authors suggest that this means that the proteins are likely released before the DNA during normal injection; however, this has not been shown directly.

Interestingly, bacteriophage P22 carries a superinfection exclusion gene *sieA* whose product is expressed from the prophage and appears to block DNA injection by P22 and other P22-like bacteriophages (Susskind et al., 1974a). The SieA protein is reported to be in the membrane fraction after cell disruption (Hofer et al., 1995). We note that among the >500 known P22-like bacteriophage genomes tailspike and the ejection protein genes are much more diverse than the rest of the genes involved in virion assembly and function (Casjens and Thuman-Commike, 2011 S. Casjens, unpublished findings). This observation has been interpreted to suggest that these more diverse proteins interact directly with host, the increased diversity being driven by the ongoing evolutionary battle between the host acquiring changes that make the bacteriophage work less efficiently and the bacteriophage responding with its own changes to overcome these host changes. In support of this idea, the tailspike is known to interact directly with a host component, the O-antigen, and it is easy to imagine that the ejection proteins interact with membrane and/or periplasmic host components while performing their injection function. Thus, although the mechanism of SieA action is currently completely mysterious, we suggest that SieA could disrupt an important interaction between the ejection proteins and a host component(s).

References

Adams, M.B., Hayden, M., Casjens, S., 1983. On the sequential packaging of bacteriophage P22 DNA. J Virol 46, 673-677.

Adhikari, P., Berget, P.B., 1993. Sequence of a DNA injection gene from *Salmonella typhimurium* phage P22. Nucleic Acids Res. 21, 1499.

Andres, D., Baxa, U., Hanke, C., Seckler, R., Barbirz, S., 2010. Carbohydrate binding of Salmonella phage P22 tailspike protein and its role during host cell infection. Biochem Soc Trans 38, 1386-1389.

Backhaus, H., 1985. DNA packaging initiation of *Salmonella* bacteriophage P22: determination of cut sites within the DNA sequence coding for gene 3. J Virol 55, 458-465.

Bauer, D.W., Li, D., Huffman, J., Homa, F.L., Wilson, K., Leavitt, J.C., Casjens, S.R., Baines, J., Evilevitch, A., 2015. Exploring the balance between DNA pressure and capsid stability in herpesviruses and phages. J Virol 89, 9288-9298.

Bazinet, C., King, J., 1988. Initiation of P22 procapsid assembly *in vivo*. J Mol Biol 202, 77-86.

Berget, P.B., Poteete, A.R., 1980. Structure and functions of the bacteriophage P22 tail protein. J Virol 34, 234-243.

Botstein, D., Waddell, C.H., King, J., 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. I. Genes, proteins, structures and DNA maturation. J Mol Biol 80, 669-695.

Casjens, S., 1997. Principles of virion structure, function and assembly, in: Chiu, W., Burnett, R., Garcea, R. (Eds.), Structural Biology of Viruses. Oxford University Press, Oxford, pp. 3-37.

Casjens, S., Hayden, M., 1988. Analysis in vivo of the bacteriophage P22 headful nuclease. J Mol Biol 199, 467-474.

Casjens, S., Hendrix, R., 1988. Control mechanisms in dsDNA bacteriophage assembly., in: Calendar, R. (Ed.), The Bacteriophages. Plenum Press, New York City, pp. 15-91.

Casjens, S., Huang, W.M., 1982. Initiation of sequential packaging of bacteriophage P22 DNA. J Mol Biol 157, 287-298.

Casjens, S., Huang, W.M., Hayden, M., Parr, R., 1987. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. J Mol Biol 194, 411-422.

Casjens, S., King, J., 1974. P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. J Supramol Struct 2, 202-224.

Casjens, S., Sampson, L., Randall, S., Eppler, K., Wu, H., Petri, J.B., Schmieger, H., 1992a. Molecular genetic analysis of bacteriophage P22 gene *3* product, a protein involved in the initiation of headful DNA packaging. J Mol Biol 227, 1086-1099.

Casjens, S., Weigele, P., 2003. Headful DNA packaging by bacteriophage P22, in: Catalano, C. (Ed.), Viral genome packaging machines. Landes Bioscience, Georgetown, TX, pp. 80-88.

Casjens, S., Wyckoff, E., Hayden, M., Sampson, L., Eppler, K., Randall, S., Moreno, E.T., Serwer, P., 1992b. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J Mol Biol 224, 1055-1074.

Casjens, S.R., 2011. The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol 9, 647-657.

Casjens, S.R., Molineux, I.J., 2012. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol 726, 143-179.

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.

Chang, C.Y., Kemp, P., Molineux, I.J., 2010. Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. Virology 398, 176-186.

Chang, J., Weigele, P., King, J., Chiu, W., Jiang, W., 2006. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. Structure 14, 1073-1082.

Chen, D.H., Baker, M.L., Hryc, C.F., DiMaio, F., Jakana, J., Wu, W., Dougherty, M., Haase-Pettingell, C., Schmid, M.F., Jiang, W., Baker, D., King, J.A., Chiu, W., 2011. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. Proc Natl Acad Sci U S A 108, 1355-1360.

Earnshaw, W.C., Casjens, S.R., 1980. DNA packaging by the double-stranded DNA bacteriophages. Cell 21, 319-331.

Ebel-Tsipis, J., Botstein, D., Fox, M.S., 1972a. Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. J Mol Biol 71, 433-448.

Ebel-Tsipis, J., Fox, M.S., Botstein, D., 1972b. Generalized transduction by bacteriophage P22 in *Salmonella typhimurium*. II. Mechanism of integration of transducing DNA. J Mol Biol 71, 449-469.

Grose, J.H., Casjens, S.R., 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468-470, 421-443.

Guo, J., Roberts, J.W., 2004. DNA binding regions of Q proteins of phages lambda and ø80. J Bacteriol 186, 3599-3608.

Guo, P., Peterson, C., Anderson, D., 1987a. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage phi 29. J Mol Biol 197, 229-236.

Guo, P.X., Erickson, S., Anderson, D., 1987b. A small viral RNA is required for in vitro packaging of bacteriophage ø29 DNA. Science 236, 690-694.

Hendrix, R., Casjens, S., 2006. Bacteriophage λ and its genetic neighborhood, in: Calendar, R. (Ed.), The Bacteriophages, 2nd Edition. Oxford Press, New York City, N.Y., pp. 409-447.

Hendrix, R.W., 2003. Bacteriophage genomics. Curr Opin Microbiol 6, 506-511.

Hofer, B., Ruge, M., Dreiseikelmann, B., 1995. The superinfection exclusion gene (*sieA*) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. J Bacteriol 177, 3080-3086.

Hoffman, B., Levine, M., 1975a. Bacteriophage P22 virion protein which performs an essential early function. I. Analysis of *16-ts* mutants. J Virol 16, 1536-1546.

Hoffman, B., Levine, M., 1975b. Bacteriophage P22 virion protein which performs an essential early function. II. Characterization of the gene *16* function. J Virol 16, 1547-1559.

Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339, 576-579.

Israel, V., 1977. E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. J Virol 23, 91-97.

Iwashita, S., Kanegasaki, S., 1976. Enzymic and molecular properties of base-plate parts of bacteriophage P22. Eur J Biochem 65, 87-94.

Jackson, E.N., Jackson, D.A., Deans, R.J., 1978. EcoRI analysis of bacteriophage P22 DNA packaging. J Mol Biol 118, 365-388.

Jackson, E.N., Laski, F., Andres, C., 1982. Bacteriophage P22 mutants that alter the specificity of DNA packaging. J Mol Biol 154, 551-563.

Jin, Y., Sdao, S.M., Dover, J.A., Porcek, N.B., Knobler, C.M., Gelbart, W.M., Parent, K.N., 2015. Bacteriophage P22 ejects all of its internal proteins before its genome. Virology 485, 128-134.

King, J., Lenk, E.V., Botstein, D., 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. J Mol Biol 80, 697-731.

Lander, G.C., Johnson, J.E., Rau, D.C., Potter, C.S., Carragher, B., Evilevitch, A., 2013. DNA bending-induced phase transition of encapsidated genome in phage lambda. Nucleic Acids Res 41, 4518-4524.

Lander, G.C., Tang, L., Casjens, S.R., Gilcrease, E.B., Prevelige, P., Poliakov, A., Potter, C.S., Carragher, B., Johnson, J.E., 2006. The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science 312, 1791-1795.

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., Rossmann, M.G., 2004. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell 118, 419-429.

Lindberg, A.A., Sarvas, M., Makela, P.H., 1970. Bacteriophage attachment to the somatic antigen of *Salmonella*: Effect of O-specific structures in leaky R mutants and S, T1 Hybrids. Infect Immun 1, 88-97.

McNulty, R., Lokareddy, R.K., Roy, A., Yang, Y., Lander, G.C., Heck, A.J., Johnson, J.E., Cingolani, G., 2015. Architecture of the Complex Formed by Large and Small Terminase Subunits from Bacteriophage P22. J Mol Biol 427, 3285-3299.

Moak, M., Molineux, I.J., 2000. Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. Mol Microbiol 37, 345-355.

Moffitt, J.R., Chemla, Y.R., Aathavan, K., Grimes, S., Jardine, P.J., Anderson, D.L., Bustamante, C., 2009. Intersubunit coordination in a homomeric ring ATPase. Nature 457, 446-450.

Molineux, I.J., 2001. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. Mol Microbiol 40, 1-8.

Molineux, I.J., 2006. Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? Virology 344, 221-229.

Molineux, I.J., Panja, D., 2013. Popping the cork: mechanisms of phage genome ejection. Nat Rev Microbiol 11, 194-204.

Moore, S.D., Prevelige, P.E., Jr., 2001. Structural transformations accompanying the assembly of bacteriophage P22 portal protein rings *in vitro*. J Biol Chem 276, 6779-6788.

Morita, M., Tasaka, M., Fujisawa, H., 1993. DNA packaging ATPase of bacteriophage T3. Virology 193, 748-752.

Muller, J.J., Barbirz, S., Heinle, K., Freiberg, A., Seckler, R., Heinemann, U., 2008. An intersubunit active site between supercoiled parallel beta helices in the trimeric tailspike endorhamnosidase of *Shigella flexneri* phage Sf6. Structure 16, 766-775.

Nemecek, D., Gilcrease, E.B., Kang, S., Prevelige, P.E., Jr., Casjens, S., Thomas, G.J., Jr., 2007. Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. J Mol Biol 374, 817-836.

Nemecek, D., Lander, G.C., Johnson, J.E., Casjens, S.R., Thomas, G.J., Jr., 2008. Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. J Mol Biol 383, 494-501.

Olia, A.S., Casjens, S., Cingolani, G., 2007. Structure of phage P22 cell envelope-penetrating needle. Nat Struct Mol Biol 14, 1221-1226.

Olia, A.S., Prevelige, P.E., Jr., Johnson, J.E., Cingolani, G., 2011. Three-dimensional structure of a viral genome-delivery portal vertex. Nat Struct Mol Biol 18, 597-603.

Parent, K.N., Khayat, R., Tu, L.H., Suhanovsky, M.M., Cortines, J.R., Teschke, C.M., Johnson, J.E., Baker, T.S., 2010. P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. Structure 18, 390-401.

Pedulla, M.L., Ford, M.E., Karthikeyan, T., Houtz, J.M., Hendrix, R.W., Hatfull, G.F., Poteete, A.R., Gilcrease, E.B., Winn-Stapley, D.A., Casjens, S.R., 2003. Corrected sequence of the bacteriophage P22 genome. J Bacteriol 185, 1475-1477.

Poteete, A.R., Botstein, D., 1979. Purification and properties of proteins essential to DNA encapsulation by phage P22. Virology 95, 565-573.

Poteete, A.R., King, J., 1977. Functions of two new genes in Salmonella phage P22 assembly. Virology 76, 725-739.

Qiu, X., Rau, D.C., Parsegian, V.A., Fang, L.T., Knobler, C.M., Gelbart, W.M., 2011. Salt-dependent DNA-DNA spacings in intact bacteriophage lambda reflect relative importance of DNA self-repulsion and bending energies. Phys Rev Lett 106, 028102.

Raj, A.S., Raj, A.Y., Schmieger, H., 1974. Phage genes involved in the formation generalized transducing particles in *Salmonella*--Phage P22. Mol Gen Genet 135, 175-184.

Rao, V.B., Feiss, M., 2008. The bacteriophage DNA packaging motor. Annu Rev Genet 42, 647-681.

Rau, D.C., Lee, B., Parsegian, V.A., 1984. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. Proc Natl Acad Sci U S A 81, 2621-2625.

Rixon, F.J., Schmid, M.F., 2014. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. Curr Opin Virol 5, 105-110.

Rizzo, A.A., Suhanovsky, M.M., Baker, M.L., Fraser, L.C., Jones, L.M., Rempel, D.L., Gross, M.L., Chiu, W., Alexandrescu, A.T., Teschke, C.M., 2014. Multiple functional roles of the accessory I-domain of bacteriophage P22 coat protein revealed by NMR structure and CryoEM modeling. Structure 22, 830-841.

Roberts, J.W., Roberts, C.W., Hilliker, S., Botstein, D., 1976. Transcription termination and regulation in bacteriophages P22 and lambda, in: Losick R, Chamberlin M (Eds.), RNA Polymerase. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 707-718.

Roy, A., Bhardwaj, A., Datta, P., Lander, G.C., Cingolani, G., 2012. Small terminase couples viral DNA binding to genome-packaging ATPase activity. Structure 20, 1403-1413.

Roy, A., Cingolani, G., 2012. Structure of p22 headful packaging nuclease. J Biol Chem 287, 28196-28205.

Schmieger, H., 1972. Phage P22-mutants with increased or decreased transduction abilities. Mol Gen Genet 119, 75-88.

Steinbacher, S., Baxa, U., Miller, S., Weintraub, A., Seckler, R., Huber, R., 1996. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc Natl Acad Sci USA 93, 10584-10588.

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., Huber, R., 1997. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. J Mol Biol 267, 865-880.

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., Reinemer, P., 1994. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. Science 265, 383-386.

Strobel, E.J., Roberts, J.W., 2014. Regulation of promoter-proximal transcription elongation: enhanced DNA scrunching drives lambdaQ antiterminator-dependent escape from a sigma70-dependent pause. Nucleic Acids Res 42, 5097-5108.

Susskind, M.M., Botstein, D., 1975. Mechanism of action of *Salmonella* phage P22 antirepressor. J Mol Biol 98, 413-424.

Susskind, M.M., Botstein, D., 1978a. Molecular genetics of bacteriophage P22. Microbiol Rev 42, 385-413.

Susskind, M.M., Botstein, D., 1978b. Repression and immunity in *Salmonella* phages P22 and L: phage L lacks a functional secondary immunity system. Virology 89, 618-622.

Susskind, M.M., Botstein, D., Wright, A., 1974a. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. III. Failure of superinfecting phage DNA to enter sieA+ lysogens. Virology 62, 350-366.

Susskind, M.M., Wright, A., Botstein, D., 1971. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. II. Genetic evidence for two exclusion systems. Virology 45, 638-652.

Susskind, M.M., Wright, A., Botstein, D., 1974b. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. IV. Genetics and physiology of sieB exclusion. Virology 62, 367-384.

Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356-361.

Tang, J., Lander, G.C., Olia, A.S., Li, R., Casjens, S., Prevelige, P., Jr., Cingolani, G., Baker, T.S., Johnson, J.E., 2011. Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in p22. Structure 19, 496-502.

Tye, B.K., Chan, R.K., Botstein, D., 1974a. Packaging of an oversize transducing genome by *Salmonella* phage P22. J Mol Biol 85, 485-500.

Tye, B.K., Huberman, J.A., Botstein, D., 1974b. Non-random circular permutation of phage P22 DNA. J Mol Biol 85, 501-528.

Umlauf, B., Dreiseikelmann, B., 1992. Cloning, sequencing, and overexpression of gene *16* of *Salmonella* bacteriophage P22. Virology 188, 495-501.

Van Valen, D., Wu, D., Chen, Y.J., Tuson, H., Wiggins, P., Phillips, R., 2012. A single-molecule Hershey-Chase experiment. Curr Biol 22, 1339-1343.

Vander Byl, C., Kropinski, A.M., 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J Bacteriol 182, 6472-6481.

Weigele, P.R., Sampson, L., Winn-Stapley, D., Casjens, S.R., 2005. Molecular genetics of bacteriophage P22 scaffolding protein's functional domains. J Mol Biol 348, 831-844.

Wu, H., Sampson, L., Parr, R., Casjens, S., 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol Microbiol 45, 1631-1646.

Zhao, H., Christensen, T.E., Kamau, Y.N., Tang, L., 2013. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. Proc Natl Acad Sci U S A 110, 8075-8080.

Zhao, H., Finch, C.J., Sequeira, R.D., Johnson, B.A., Johnson, J.E., Casjens, S.R., Tang, L., 2010. Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. Proc Natl Acad Sci U S A 107, 1971-1976.

Zhao, H., Kamau, Y.N., Christensen, T.E., Tang, L., 2012. Structural and functional studies of the phage Sf6 terminase small subunit reveal a DNA-spooling device facilitated by structural plasticity. J Mol Biol 423, 413-426.

Zinder, N., Lederberg, J., 1952. Genetic exchange in *Salmonella*. J Bacteriol 64, 679-699.

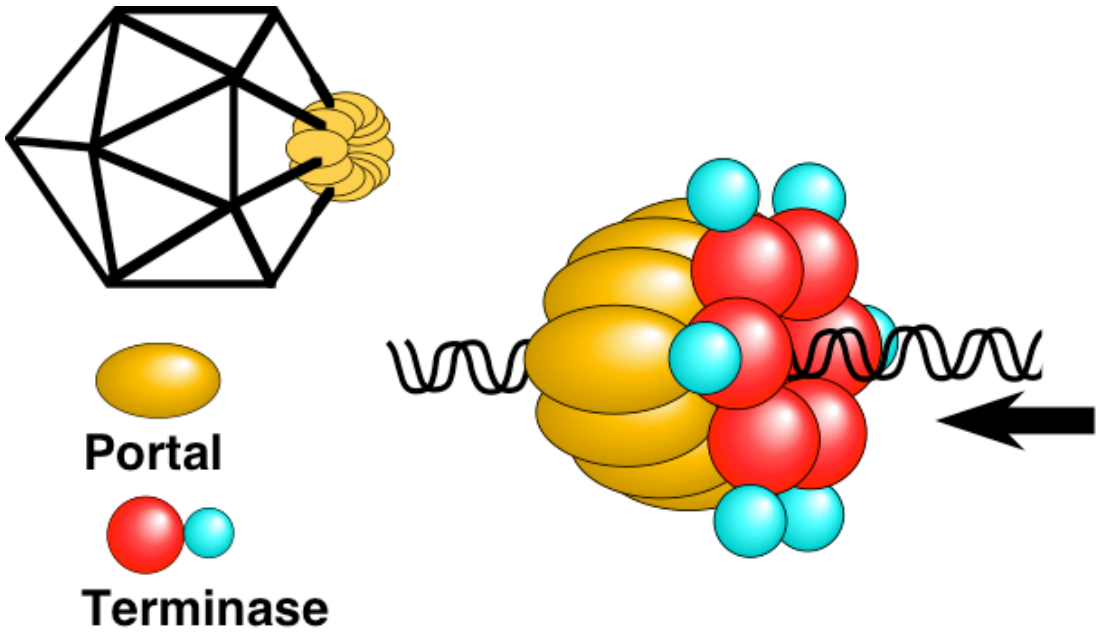


Figure 1.1 Artist's depiction of DNA packaging motor.

A highly simplified depiction of the DNA-packaging motor is shown as follows: upper left, the yellow dodecameric portal ring is present at one vertex of the icosahedral bacteriophage head; lower left and right, the red packaging motor ATPase (large terminase subunit or TerL) and blue DNA recognition protein (small terminase subunit or TerS) are shown as colored balls. The arrangement of the subunits of the three components within the packaging motor is not known, and the physical arrangement shown is only diagrammatic, especially in terms of the position of TerS (see text). The black arrow indicates the location of the dsDNA during motor action and the direction of DNA movement.

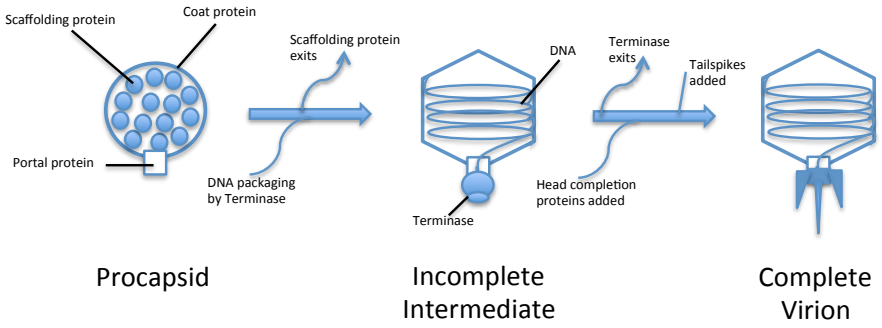


Figure 1.2 Assembly pathway of the bacteriophage P22 virion.

The assembly order of bacteriophage P22. The assembly of coat (gp5) and portal (gp1) proteins are facilitated by the scaffolding protein (gp8). The ejection proteins are recruited at the procapsid stage of assembly, but their location in the virion structure is unknown. The terminase complex, consisting of the large and small terminase (gp2 and gp3), binds to the portal protein and pumps DNA into the procapsid. Scaffolding protein exits the viral head and the procapsid expands to mature dimensions. Subsequently, the head completion proteins (gp4, gp10, and gp26) are added to form a stable particle, and tailspikes are added to form a complete infectious particle.

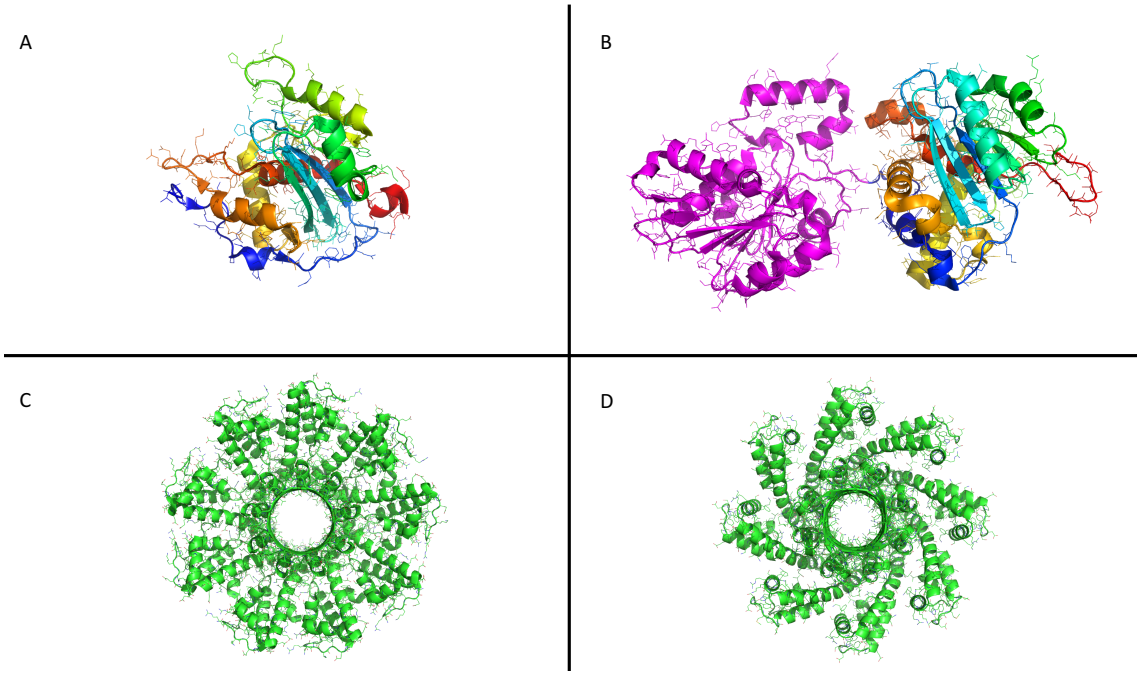


Figure 1.3 Ribbon diagrams of atomic structures of P22-like bacteriophage terminase subunits.

A. The P22 TerL nuclease domain, blue is N-terminus red is C-terminus (PDB 4DKW Roy and Cingolani, 2012).

B. The Sf6 large terminase the N-terminal ATPase domain is colored magenta. The C-terminal nuclease domain is rainbow colored (N-terminus blue, C-terminus red) for ease of comparison to panel A (PDB 4IDH Zhao et al., 2013).

C. The P22 small terminase 9mer (PDB 3P9A Roy et al., 2012).

D. The Sf6 small terminase 8mer (PDB 4DYQ Zhao et al., 2010; Zhao et al., 2012).

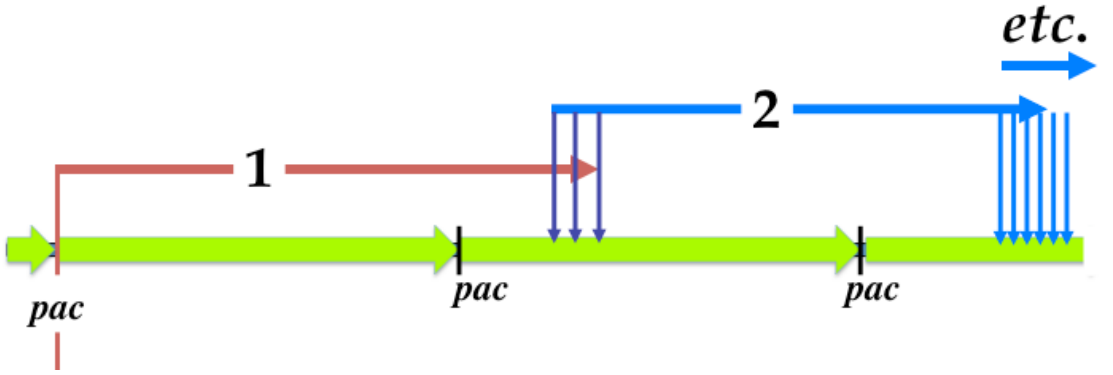


Figure 1.4 Bacteriophage P22 headful DNA packaging series.

The green arrow represents concatemeric P22 genome sequences whose *pac* sequence lies at the left end of the arrow. The first packaging event on this substrate molecule is indicated as a red horizontal arrow; it starts at a *pac* site and packaging proceeds rightward until more than one sequence is packaged. When the head is full of DNA, the headful nuclease (vertical blue arrows) cleaves it. The second packaging event (blue horizontal arrow) in the series then begins at the DNA end created by the first event's headful cleavage and continues until the head is full again, at which point a second headful cleavage occurs. Subsequent events then proceed rightwards analogous to the second event.

CHAPTER 2

FUNCTION AND HORIZONTAL TRANSFER OF THE SMALL TERMINASE SUBUNIT OF THE TAILED BACTERIOPHAGE SF6 DNA PACKAGING NANOMOTOR

Virology 440 (2013) 117–133

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Function and horizontal transfer of the small terminase subunit of the tailed bacteriophage Sf6 DNA packaging nanomotor

Justin C. Leavitt [a], Eddie B. Gilcrease [b], Kassandra Wilson [b], Sherwood R. Casjens [a,b,*]

[a] Biology Department, University of Utah, Salt Lake City, UT 84112, USA
[b] Division of Microbiology and Immunology, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

ARTICLE INFO

Available online 4 April 2013

Keywords:
Bacteriophage Sf6
DNA packaging
Bacteriophage P22
Small terminase subunit
TerS

ABSTRACT

Bacteriophage Sf6 DNA packaging series initiate at many locations across a 2 kbp region. Our *in vivo* studies show that Sf6 small terminase subunit (TerS) protein recognizes a specific packaging (*pac*) site near the center of this region, that this site lies within the portion of the Sf6 gene that encodes the DNA-binding domain of TerS protein, that this domain of the TerS protein is responsible for the imprecision in Sf6 packaging initiation, and that the DNA-binding domain of TerS must be covalently attached to the domain that interacts with the rest of the packaging motor. The TerS DNA-binding domain is self-contained in that it apparently does not interact closely with the rest of the motor and it binds to a recognition site that lies within the DNA that encodes the domain. This arrangement has allowed the horizontal exchange of *terS* genes among phages to be very successful.

Introduction

The virions of tailed bacteriophages and other large dsDNA viruses contain a highly compacted nucleic acid molecule (Casjens, 1997). During the assembly of such virions, an ATP cleavage powered protein nanomotor pumps the DNA into a preformed capsid protein shell called a procapsid (reviewed in Casjens, 2011; Feiss and Rao, 2012). This DNA translocating ATPase is called the large terminase subunit (TerL), and it moves the dsDNA through a dodecameric ring of portal protein subunits that is present at one icosahedral vertex of the procapsid. Terminase protein's name derives from the fact that many of large dsDNA viruses replicate DNA into overlength concatemers of the genome sequence, and in these cases TerL also carries a nuclease activity that cuts this long DNA to virion size, thus creating the *termini* of the mature viral DNA chromosome. Recent single-particle optical tweezer experiments have produced new information that has allowed the building of rather detailed mechanistic models for the mechanism of action of this translocase, and these models are being tested (Duffy and Feiss, 2002; Kondabagil et al., 2006; Oliveira et al., 2006; Tsay et al., 2009, 2010; Casjens, 2011; Feiss and Rao, 2012). In a number of such viruses a second protein, called the small terminase subunit (TerS), has been implicated in the initiation of DNA packaging and in the choice of DNA to be packaged in the virion. In these cases TerS proteins recognize a specific site in the phage DNA, but their detailed mechanism of action is very poorly understood (Jackson et al., 1982; Shinder and Gold, 1988; Casjens et al., 1992a; Chai et al., 1994).

The tailed phage packaging motor usually has three protein components, TerS, TerL and the portal protein which forms the hole through which DNA enters the procapsid. Portal proteins may in some cases also have roles in sensing when the capsid shell is full of DNA (Casjens et al., 1992b; Tavares et al., 1992), in controlling the shape of the coat protein shell (Camacho et al., 1977; Black et al., 1994) and in controlling the conformational change (expansion) that capsid shells undergo during maturation (Ray et al., 2009). TerL interacts with the portal protein ring of the procapsid in the cases where this interaction is understood, and genetic studies with phages λ and T3 suggest that the C-terminal portion of TerL interacts with portal protein (Frackman et al., 1984; Sippy and Feiss, 1992; Morita et al., 1995; Yeo and Feiss, 1995a, 1995b), but other sequences have also been implicated in phage T4 TerL binding (Lin et al., 1999; Gao and Rao, 2011; Hegde et al., 2012). TerL and TerS also often interact in solution (Poteete and Botstein, 1979; Maluf et al., 2005), although perhaps not in all phages (Al-Zahrani et al., 2009). In phage λ the C-terminal region of TerS is thought to bind to TerL (Frackman et al., 1985; Yang et al., 1999b). Although atomic structures have recently been determined for several examples of each of the three motor protein subunits, the structure of the assembled, functioning motor is not yet understood.

Although the DNA packaging motor proteins appear to be evolutionarily conserved in spite of having a huge extant diversity

* Corresponding author at: Room 2200 EEJMRB, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.
Fax: +801 585 2417.
E-mail address: sherwood.casjens@path.utah.edu (S.R. Casjens).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.02.023

in the different tailed phages that have been studied, portal and TerL proteins are universally encoded by tailed phages and are their most highly conserved proteins (Casjens, 2003, 2008). Portal proteins are always found as dodecameric rings that replace five coat protein subunits at the tail vertex of phage heads, and the x-ray structures of portal rings from the very distantly related phages ϕ29, SPP1 and P22 show that they all have the same basic central fold, in spite of the fact that their amino acid sequences are not convincingly similar (Simpson et al., 2000; Lebedev et al., 2007; Olia et al., 2011). They are quite variable in size and different phage portals can have different "accessory" domains (Tang et al., 2011). The TerL proteins are monomeric when not part of the motor, and existing evidence suggests that four or five TerL molecules participate in the assembled motor (reviewed in Casjens, 2011; Nemecek et al., 2007; Feiss and Rao, 2012). TerS sequence diversity is even larger than that seen among the TerL or portal proteins; BLASTp searches of the sequence database identify a number of *apparently* unrelated TerS protein families (Casjens and Thuman-Commike, 2011; S. Casjens, unpublished). Crystal structures of octamers of *Shigella flexneri* phage Sf6 TerS (Zhao et al., 2010), nonamers of *Salmonella enterica* phage P22 TerS (Roy et al., 2012), nonamers and decamers of *Bacillus subtilis* phage SF6 TerS (Buttner et al., 2012), and 11- and 12-mers of a fragment of *Escherichia coli* phage 44RR2 TerS (Sun et al., 2012), as well as the NMR structure of a dimer of a fragment of *E. coli* phage λ TerS (de Beer et al., 2002) have shown the following: (i) The oligomeric state of the purified TerS proteins varies among the tailed phages, but its assembly state in the complete motor is not known in any case. Thus, either the different motors can accommodate a variable number of TerS subunits, or the oligomeric structure of TerS in the functioning motor may be different from that of the purified TerS proteins. (ii) The TerS C-terminus, where its structure has been determined (in Sf6, SF6 and in part in P22), forms a tubular β-barrel that contains one peptide strand from each subunit. (iii) The Sf6 and SF6 TerS proteins have rather similar overall folds; these two phages are only extremely distantly related in spite of their unfortunately similar names. The P22 TerS fold is partly similar but not identical to these two proteins (see below). The relationships between these more complete structures to the fragment structures of λ and phage 44RR2 TerS structures are less clear (but see Gao and Rao, 2011). Nonetheless, in all of the TerS structures the N-terminal domain is largely helical and includes a helix-turn-helix motif that may be the DNA-binding portion of these proteins (Buttner et al., 2012; Roy et al., 2012; Zhao et al., 2012; Sun et al., 2012). Analysis of mutations of Sf6 TerS have indicated that its N-terminal domain is responsible for binding DNA nonspecifically *in vitro* (Zhao et al., 2010, 2012). In apparent contradiction to this view, removal of twenty C-terminal amino acids, which make up most of the C-terminal tubular β-barrel domain of the P22 TerS, does not impair its oligomerization but does block its DNA binding capability (Nemecek et al., 2008; Roy et al., 2012). T4 TerS is dispensable in a T4 *in vitro* DNA packaging system (Al-Zahrani et al., 2009; Zhang et al., 2011), while in P22 and SPP1 TerS protein is required but specific recognition of the packaging target site is not required *in vitro* (Poteete and Botstein, 1979; Schmieger, 1984; Schmieger and Koch, 1987; Oliveira et al., 2005). Thus much remains to be understood regarding small terminase subunit function and its role in DNA packaging motor initiation.

The well-studied phage P22 and its relative Sf6 are both members of the "P22-like" tailed phage group, and twelve different weakly homologous proteins build their very similar virions (Casjens et al., 2004; Casjens and Thuman-Commike, 2011; Parent et al., 2012). A major functional difference between these two phages lies in the initiation of DNA packaging. P22 recognizes a specific 22 bp *pac* site that programs the initiation of processive series of packaging events (Tye et al., 1974; Jackson et al., 1978; Casjens et al., 1992a; Wu et al., 2002), and the DNA cleavage that initiates such a series occurs over a 120 bp region that surrounds the *pac* site (Casjens and Huang, 1982; Casjens et al., 1992a). TerL protein has nonspecific nuclease activity, and it is thought to make these cleavages (Nemecek et al., 2007; Roy and Cingolani, 2012). On the other hand, we found that Sf6 makes its initiation cleavages over a much larger approximately 1600 bp region (Casjens et al., 2004). The reason for this difference was unknown. We report here the localization of the Sf6 *pac* site near the center of the region in which packaging initiation ends are generated and that the Sf6 TerS subunit is responsible for the large spread in initiation cleavage sites. In addition, genetic analysis of the Sf6 TerS protein supports the notion that this TerS protein has substantial flexibility and interacts with DNA through its N-terminal domain and with TerL through is C-terminal domain.

Results

Horizontal transfer of terS genes among the P22-like phages

There were, as of December 1, 2012, 152 available complete or nearly complete genome sequences of P22-like phages and prophages (this group of phages is defined as in Casjens and Thuman-Commike, 2011). The coat proteins of these phages fall into three major sequence types that are typified by *Salmonella* phage P22, *Shigella* phage Sf6 and *E. coli* phage CUS-3 (see Fig. 4 of Casjens and Thuman-Commike, 2011). On the other hand, there are six very different TerS "sequence types" that are mostly not convincingly related to one another in amino acid sequence. Fig. 1 shows a ClustalW (Larkin et al., 2007) amino acid sequence neighbor-joining tree the of representatives of these TerS types. All of the sequences in Fig. 1 except the putative ϕSG1 TerS have amino acid sequences that are weakly related to known TerS proteins. (The putative ϕSG1 TerS has no recognizable homology to any protein in the database, but it was included in this analysis because its gene lies in the position at which *all* other P22-like phages carry their *terS* genes.) Comparisons of the six types of P22-like TerS proteins typically show less than 15% amino acid sequence identity between types, and simple BLASTp (Altschul et al., 1997) searches with one type often do not find matches in the other groups. The correlation between coat types and TerS types in these genomes is poor; for example different phages with the P22 type coat protein encode TerS proteins of the P22, Ugan1 or CUS-3 types, and different phages with CUS-3 type coat protein have CUS-3, P22, Sf6 or Ugan1 type TerS proteins. Clearly there has been extensive horizontal exchange of *terS* genes relative to coat protein genes *within* the P22-like phage group.

Fig. 1 also shows that four of the above TerS types include proteins encoded by phages outside of the P22-like group. For example, the TerS type exemplified by phage CUS-3 (pink box in Fig. 1) includes TerS proteins encoded by prophage Bpert1 in a *Bordetella pertussis* genome (a *Betaproteobacteria*), prophage Zymob1 in a *Zymomonas mobilis* genome (an *Alphaproteobacteria*), prophage Tcarb1 in a *Thermosinus carboxydivorans* genome, and prophage Bact1 in a *Bacteroides* species genome (see Tables S1 and S2 for details of these bacterial hosts and prophage *terS* genes). The first two of these bacterial hosts are in the *Proteobacteria* phylum but reside in different taxonomic classes from the Gamma-Proteobacteria hosts of the P22-like phages, the third is in the *Firmicutes* phylum, and the last is in the *Bacteriodetes* phylum. These four prophages are not P22-like in either the organization of their virion assembly genes or the sequence of the encoded proteins (not shown). Similarly, *bone fide* tailed phages T1 and ES-2 have TerS

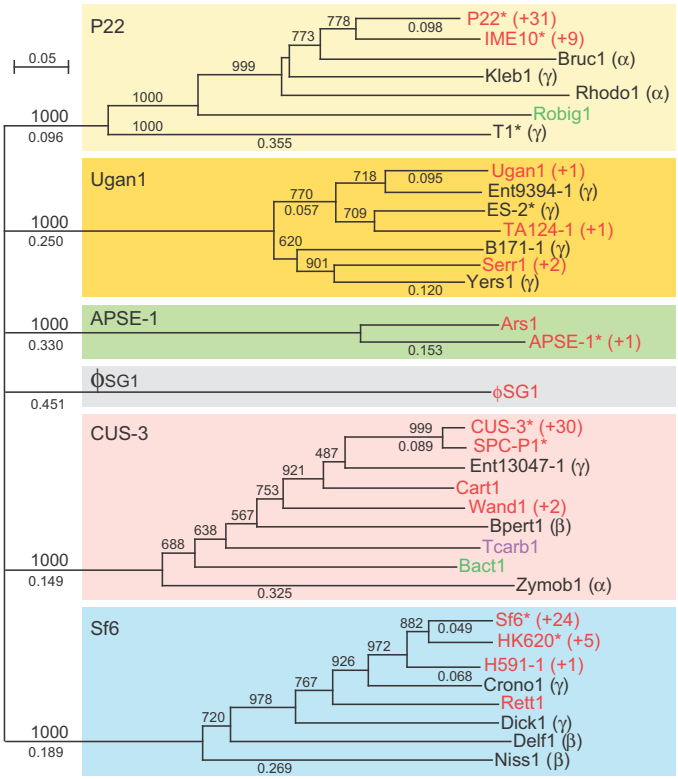


Fig. 1. Neighbor-joining tree of TerS proteins of the P22-like phages. The amino acid sequences of the P22-like TerS proteins were limited to the N-terminal DNA binding domain, and did not contain their short C-terminal domain (such a comparison is dominated by the much larger N-terminal domain and the tree shown is very similar to the tree of the whole proteins; see text). Trees were constructed by Clustal X2 in a Macintosh computer (Larkin et al., 2007), with horizontal branch lengths (numbers between 0 and 1) indicating the fractional amino acid sequence difference, and bootstrap support out of 1000 trials indicated by numbers between 1 and 1000. The very weakly supported deep branching order of the six major TerS types was collapsed, so that these six branches emanate from a single point. The major branches (highlighted with different colored boxes) are depicted on a tree for ease of discussion, in spite of the fact that the six branches may in fact not be homologous; the name of each branch (derived from a typical member) is shown in the upper left corner of the box. The name of the phage or prophage that carries each TerS protein is indicated to the right of each branch tip; functional phages are marked with an asterisk (*). TerS sequences that are nearly identical to those named on the right were collapsed to make the tree more legible, and the number removed is indicated after a plus sign (+) to the right of the phage or prophage name. The colors of the names indicate the following: Red, P22-like phages that infect members of the *Enterobacteriaceae* bacterial family; Black, non-P22-like phages that infect members of the *Proteobacteria* phylum, and the host's class within this phylum is indicated in parentheses (*e.g.*, α for *Alphaproteobacteria*, β for *Betaproteobacteria, etc.*); Green, non-P22-like prophages whose hosts are members of the *Bacteroidetes* phylum; Purple, non-P22-like prophage whose host is a member of the *Firmicutes* phylum. The species of the host and GenBank locus_tag of the TerS proteins in the figure are given in Tables S1 and S2.

proteins that reside in the P22 and Ugan1 families, respectively. T1 infects *E. coli* and is a member of the *Siphoviridae* with a long, noncontractile tail (German et al., 2006), and ES-2 infects *Cronobacter sakazakii* and is a member of the *Myoviridae* with a long contractile tail (Lee et al., 2011). The fact that these proteins encoded by other phage types are present *inside* four of the branches of the P22-like TerS tree means that no matter where the root of the tree actually lies, such "outsiders" are present inside at least three of the P22-like branches.

Not all phage genes are exchanged at this frequency. Sub-nanometer resolution 3-dimensional cryoelectron microscopic reconstructions of virions have been determined for three phages that typify the three types of coat proteins in the P22-like group, P22 (Jiang et al., 2003; Parent et al., 2010; Tang et al., 2011), Sf6 (Parent et al., 2012) and CUS-3 (K. Parent, T. Baker, E. Gilcrease and S. Casjens, unpublished). These structures show that all three coat proteins have a phage HK97 coat protein polypeptide fold (Wikoff et al., 2000) that is embellished with an "extra" telokin-like domain at the same location in the protein in all three cases. This domain is not present in this location in other phage coat proteins whose structures are known, indicating that the P22-like coat proteins have not been subject to horizontal exchange from outside the P22-like group and have diverged within this group from a common ancestor (Parent et al., 2012). In addition, portal and scaffolding proteins are, like coat protein, present as three major types in the P22-like phages, but these types correlate perfectly with the coat protein types; thus, coat, portal and scaffolding protein genes have not been shuffled by evolutionary exchange (Casjens and Thuman-Commike, 2011). We conclude that horizontal exchange of *terS* genes must have occurred among the different P22-like phages, as well as between this group and other tailed phage types, while the procapsid assembly (coat, portal and scaffolding protein) genes have not enjoyed such free exchange.

The tailed phages are well known for having mosaic genomes, and while analyzing a much smaller number of the P22-like phage *terS* genes, we discovered that the locations of the boundaries between "mosaic sections" that were formed during the

exchanges of *terS* genes are correlated with protein domain boundaries. Mosaic boundaries near the C-terminus, but within the *terS* gene could be identified in several P22-like phages (Casjens and Thuman-Commike, 2011). There are now nearly three times as many P22-like phage genome sequences known, and among these we identify eight different situations in which the location of the TerS mosaic boundary can be quite accurately located (other sequence types exist but the mosaic boundary cannot be located because no relatives with shuffled TerS domains have been found to date). Fig. 2 shows that these eight different mosaic boundaries are all present within a small region, and that four different N-terminal TerS domains have been extensively shuffled relative to four different C-terminal domains at these boundaries. The sizes of these small C-terminal domains range from 23 amino acids in Ugan1 to 28 in P22. We also note that in all of the >150 P22-like phage genomes examined to date, the different C-terminal TerS protein section types correlate perfectly with several different TerL N-terminal domain sequence types (right column, Fig. 2), suggesting that evolutionary shuffling events that create phages with different combinations of TerS C-terminal section and TerL protein types have not survived. Since P22 TerS and TerL are known to form a mixed oligomeric protein complex (Poteete and Botstein, 1979; Roy et al., 2012), this failure to survive is most easily explained if the C-terminal domain of TerS is important for the interaction between TerS and TerL (Casjens and Thuman-Commike, 2011).

TerS functional domains

In order to begin to test the idea, derived from the above comparative genomic analysis, that the C-terminal domain of TerS is responsible for its interaction with TerL, we created new phage genome constructs in which the N-terminal TerS domain from phage Sf6 replaces the parallel domain of phage P22 and tested their functionality. Phage Sf6 TerS was chosen because its amino acid sequence is essentially unrelated to that of P22 TerS (the two proteins are only 11.7% identical and the few identities are scattered throughout the protein alignment), and its crystal structure has been solved (Zhao et al., 2010). This combination of N-terminal TerS domain (blue background in Fig. 2) and *terL* (yellow background in Fig. 2) has not been found in nature. We performed these replacements in a P22 prophage, since virion assembly is not required for maintenance of a prophage, and therefore mutations lethal for virion assembly or function can be constructed. The resulting prophages can be induced to lytic growth to test for successful DNA packaging and virion assembly. We previously constructed a P22 *sieA*$^{-}\Delta 1$, *15*$^{-}\Delta$S302::Kan^{R}, *13*$^{-}$*am*H101 in which the three mutations allow efficient tailspike gene expression after induction, kanamycin selection for lysogens, and control of lysis, respectively (Cortines et al., 2011; Padilla-Meier et al., 2012). This prophage, present in sup$^{\circ}$ *S. enterica* serotype Typhimurium LT2 strain UB-1791 (all bacterial and phage strains are listed in Table 1), produces fully-tailed, plaque-forming virions after induction, and it was used in all genetic manipulations of the *terS* gene described here. In short, we first replaced the native bacterial *galK* gene with a tetracycline resistance cassette (TetRA) (Karlinsey, 2007). Then, *galK* recombineering (Warming et al., 2005) was used to replace part of the *terS* gene of the P22 prophage with the *E. coli galK* gene expression cassette from plasmid pGalK (Warming et al., 2005), and this *galK* cassette was in turn replaced by the desired part of the Sf6 *terS* gene (details in Materials and Methods).

Two prophage constructs with hybrid *terS* genes, P22 Sf6-hybA and P22 Sf6-hybB, whose hybrid junctions are shown in Fig. 3, gave yields of plaque-forming phages upon induction to lytic growth with Mitomycin C that were very similar to the parental P22 prophage with its fully P22 *terS* gene. In these two hybrid phages Sf6 *terS* codons 1-114 replace codons 1-128 or 1-134 of P22 *terS*; in both cases the Sf6 *terS* sequence is fused translationally in-frame to the remaining C-terminal P22 *terS* sequences. The observation that these two hybrid phages are functional is perhaps somewhat surprising, since the isolated P22 TerS protein is a nonamer ring (Nemecek et al., 2008; Roy et al., 2012), and the Sf6 TerS forms an octamer ring (Zhao et al., 2010). Fig. 4 shows ribbon diagrams of both oligomers and single subunits of the oligomers; these structures show that the mosaic junction of the

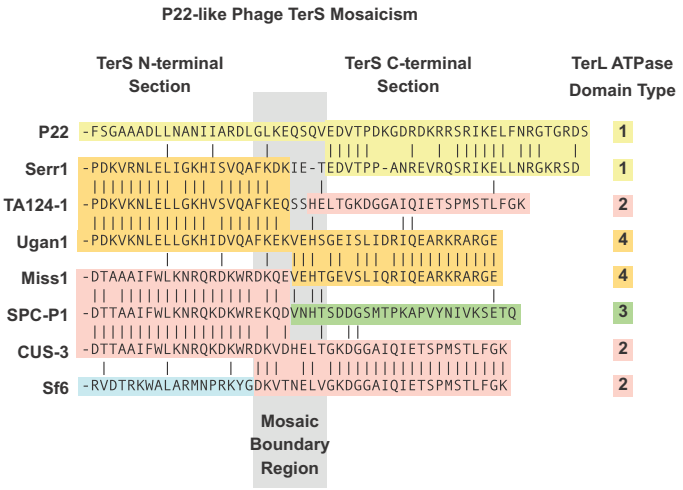


Fig. 2. TerS C-terminal amino acid sequence relationships. The amino acid sequence relationships surrounding the C-terminal and N-terminal mosaic sectional boundaries of the TerS proteins of eight P22-like phages are shown. The same background color denotes similar sequences; each sequence ends on the right at the C-terminus of the protein (the sequence relationships of the remainder of the N-terminal sections are similar to the portions shown). The numbers in the right column represent the sequence type of the TerL N-terminal ATPase domain of each phage (Casjens and Thuman-Commike, 2011; S. Casjens, unpublished). Table S2 gives the Genbank locus_tags for these *terS* genes.

Table 1
Bacteria and bacteriophage strains used in this study.

Name	Genotype[a]	Source
***Salmonella enterica* serovar Typhimurium LT2**		
UB-0002	(DB7004) *leuA am*414, *supE*	Winston et al. (1979)
UB-0020	(MS1868) *leuA am*414, Fels2$^-$, r^-, m^+, sup°; from K. Hughes	Youderian et al. (1983)
UB-0134	*leuA*$^-$*am*414, Fels2$^-$, *cob*$^-$ΔCRR299 (P22 *sieA*$^-$44, *ant*$^-$*am*222, ΔAp68 [tpfr49 *a*1$^-$, 9$^-$, *c*2$^+$, *mnt*$^+$]); from J. Roth	Youderian et al. (1982)
UB-1760	(TT23216) LT2 *ter*Y2::Cam^R; from J. Roth	Kulesus et al. (2008)
UB-1766	(TT25401) LT2 CRR2061(zfa-9223::kan,zfa-9228::*TetRA* Peut) eut-38::MudA; from J. Roth	Pimkin et al. (2009)
UB-1790	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	Padilla-Meier et al. (2012)
UB-1958	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybG, *sieA*$^-$Δ1)	This report
UB-1960	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybJ, *sieA*$^-$Δ1)	This report
UB-1961	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::*galK*-1, *sieA*$^-$Δ1)/pKD46	This report
UB-1982	UB-0020 *galK*$^+$, Cam^R-1, *TetRA*-2	This report
UB-1985	UB-0020 *galK*::P22pacL, Cam^R-1, *TetRA*-2	This report
UB-1988	UB-0020 *galK*::P22pacL, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
UB-1991	UB-0020 *galK*::P22pacR, Cam^R-1, *TetRA*-2	This report
UB-2019	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2021	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybD, *sieA*$^-$Δ1)	This report
UB-2022	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybE, *sieA*$^-$Δ1)	This report
UB-2023	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybF, *sieA*$^-$Δ1)	This report
UB-2024	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybC, *sieA*$^-$Δ1)	This report
UB-2033	UB-0020 galK::*TetRA*-1 (P22 13$^-$*am*H101, 15$^-$Δsc302::Kan^R, 3::Sf6-hybI, *sieA*$^-$Δ1)	This report
UB-2040	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybB, *sieA*$^-$Δ1)	This report
UB-2041	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybK, *3*$^-$ D115A, V116A, T117A, *sieA*$^-$Δ1)	This report
UB-2042	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybL, *3*$^-$ P118A, D119A, K120A, *sieA*$^-$Δ1)	This report
UB-2043	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybM, *3*$^-$ G121A, D122A, R123A, *sieA*$^-$Δ1)	This report
UB-2044	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybN, *3*$^-$ D124A, K125A, R126A, *sieA*$^-$Δ1)	This report
UB-2045	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybO, *3*$^-$ R127A, S128A, R129A, *sieA*$^-$Δ1)	This report
UB-2046	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybP, *3*$^-$ I130A, K131A, E132A, *sieA*$^-$Δ1)	This report
UB-2047	UB-0020 *galK*-Δ1, Cam^R-1, *TetRA*-2	This report
UB-2071	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybQ, *3*$^-$L133A, F134A, N135A, *sieA*$^-$Δ1)	This report
UB-2072	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybR, *3*$^-$D113A, K114A, *sieA*$^-$Δ1)	This report
UB-2073	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybS *3*$^-$ E115A, V116A, *sieA*$^-$Δ1)	This report
UB-2074	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybT SGSG inserted between K114 and E115, *sieA*$^-$Δ1)	This report
UB-2075	UB-0020 *galK*::*TetRA*-1 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybU SGSGSGSG inserted between K114 and E115, *sieA*$^-$Δ1)	This report
UB-2093	UB-0020 *galK*$^+$, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2094	UB-0020 *galK*::P22pacR, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2095	UB-0020 *galK*::Sf6pac3L, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2099	*UB-0020 galK*-Δ1, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2100	UB-0020 *galK*:Sf6pac5R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2101	UB-0020 *galK*::Sf6pac1L, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2102	UB-0020 *galK*::Sf6pac4R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2103	UB-0020 *galK*::Sf6pac3R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2104	UB-0020 *galK*::Sf6pac1R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2105	UB-0020 *galK*::Sf6pac2R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2106	UB-0020 *galK*::Sf6frag7, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybA, *sieA*$^-$Δ1)	This report
UB-2118	UB-0020 *galK*-Δ1, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
UB-2119	UB-0020 *galK*$^+$, Cam^R-1, *TetRA*-2, (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
UB-2120	UB-0020 *galK*::P22pacR, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
UB-2121	UB-0020 *galK*::Sf6pac1L, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
UB-2123	UB-0020 *galK*::Sf6pac2R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
UB-2124	UB-0020 *galK*::Sf6pac1R, Cam^R-1, *TetRA*-2 (P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1)	This report
***Escherichia coli* K-12**		
UB-0049	(NF1829) *araD*$^-$139, Δ7679(*araABOIC*$^-$,*leu*$^-$), *galUK*$^-$, *lac*$^-$ΔX74, *rspL*$^-$, *thi*$^-$/ F' *lac* Iq1, *lac Z*::Tn5(Kan^R), *lacY*$^+$	Shultz et al. (1982)
Shigella flexneri		
UB-1458	PE577; gift of R. Morona	Casjens et al. (2004)
UB-1469	PE577 (Sf6); gift of R. Morona	This report
UB-1564	PE577 (Sf6 *62*::Kan^R)	This report
Bacteriophages		
UC-911	P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1	Padilla-Meier et al. (2012)
UC-920	Sf6 *63*::pac'0 (plasmid pPP311 XhoI-NcoI cloning region with no insert; see text)	This report
UC-921	Sf6 *63*::pac'A (Sf6 bp 1–200)	This report
UC-922	Sf6 *63*::pac'B (Sf6 bp 149–424)	This report
UC-923	Sf6 *63*::pac'C (Sf6 bp 200–424)	This report
UC-924	Sf6 *63*::pac'D (Sf6 bp 140–210)	This report
UC-925	Sf6 *63*::pac'E (Sf6 (bp 149–225)	This report
UC-929	P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybA, *sieA*$^-$Δ1	This report
UC-930	P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *3*::Sf6-hybB, *sieA*$^-$Δ1	This report

[a] Strain names in parentheses are the names used in the laboratory from which the strain was obtained. Strain name UB-0020 in middle column indicates that the strain also carries the UB-0020 alleles. Amino acid numbers refer to the amino acids of the Sf6-hybB TerS protein.

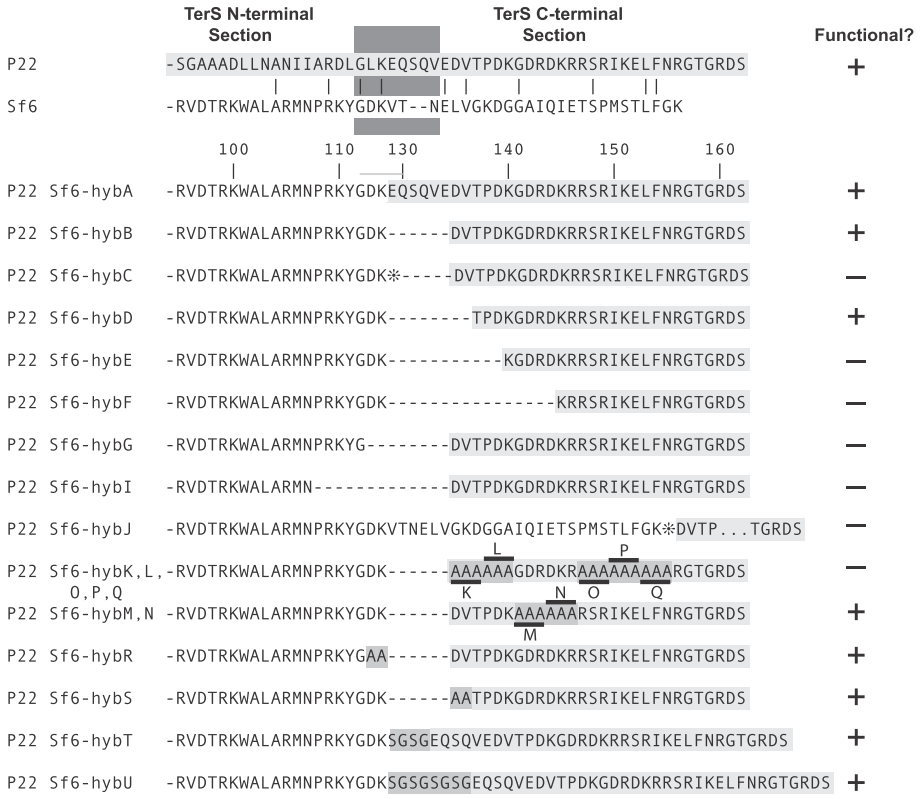


Fig. 3. Phage P22-Sf6 hybrid TerS proteins. The upper two amino acid sequences are the C-termini of the TerS proteins of phages P22 (light gray background) and Sf6 (white background). The dark gray box above marks the approximate division between the two functional domains of the TerS proteins (see text). Below, the junctions are shown for the various P22-Sf6 hybrid TerS proteins discussed in the text; the numbers above indicate the amino acids of the Sf6 protein (left of junction) and P22 protein (right of junction). Medium gray backgrounds indicate amino acid changes or insertions, and asterisks (*) denote UGA stop codons. The rightmost column indicates the functionality of the hybrid TerS proteins as determined by the phage titer on *Salmonella* UB-0002 of the lysate 3 h after induction by the addition of 1.5 μg/ml carbadox of each prophage containing strain at a cell density of 2×10^8/ml; a yield of $1\times10^{10}-2\times10^{11}$ phage/ml indicated a functional protein (+) and yields $<10^5$/ml indicated a nonfunctional protein (–).

Sf6-hybA TerS protein correspond precisely to the boundary between the N-terminal globular domain and the C-terminal β-barrel domain of both proteins. The oligomeric state of the hybrid proteins has not been determined. Nonetheless, the functionality of these two hybrid phages allows two initial conclusions to be drawn. (i) Sf6 TerS amino acids 1-114 are sufficient, when fused to amino acids 137-162 of P22 TerS, to supply TerS function to phage P22. It seems very likely that, like P22 itself, all the P22-like phages including Sf6 are dependent upon a *pac* site that is recognized by TerS. If so, (ii) the Sf6 *pac* site must lie within the first 114 codons of the Sf6 *terS* gene (see below).

In order to test whether the covalent connection between the N-terminal Sf6 domain and the C-terminal P22 domain is essential, a UGA stop codon was engineered between the two TerS domains of prophage P22 *terS* Sf6-hybB to create P22 *terS* Sf6-hybC (Fig. 3). In addition, a prophage carrying the TerS Sf6-hybJ was constructed in which the entire 140-codon Sf6 *terS* gene is present with a stop codon separating it from the P22 C-terminal region. In both of these the progeny yield upon induction is $>10^6$-fold lower than the P22 *terS* Sf6-hybB parent. The nonfunctionality of these two phages indicates that neither the free Sf6 TerS N-terminal domain nor the full-length Sf6 TerS protein can supply TerS function to the P22 packaging apparatus, and that the covalent connection between the N-terminal Sf6 TerS domain and the C-terminal P22 TerS domain is essential. These observations support the idea that the TerS C-terminal domain is required to attach the TerS N-terminal domain to the rest of the DNA packaging apparatus.

The Sf6-hybB *terS* gene was trimmed from the Sf6-P22 junction point to determine how much of the Sf6 N-terminal domain and of the P22 C-terminal domain are required at the fusion point to generate a functional hybrid TerS protein. Deletion of two of the P22 amino acids (P22 D135 and V136 of Sf6-hybD TerS in UB-2021) from the C-terminal side of the hybrid junction resulted in a functional phage, but removal of three (P22 T137, P138 and D139 in Sf6-hybE TerS in UB-2022) or eight (Sf6-hybF TerS in UB-2023) additional P22 amino acids resulted in nonfunctional phages (amino acid numbers are as shown in Fig. 3). Removal of two amino acids, D113 and K114, from the Sf6 (N-terminal) side of the Sf6-hybB junction resulted in the nonfunctional phage P22 *terS* Sf6-hybG (UB-1958), as did removal of seven amino acids (P22 *terS* Sf6-hybI in UB-2033) (Fig. 3). Comparison of Sf6-hybK and -hybS TerS proteins (below) indicates that the third amino acid of the P22 part of Sf6-hybB TerS, T137, is important. Thus, removal or alteration of two or more amino acids from the Sf6 side or three or more amino acids from the P22 side of the

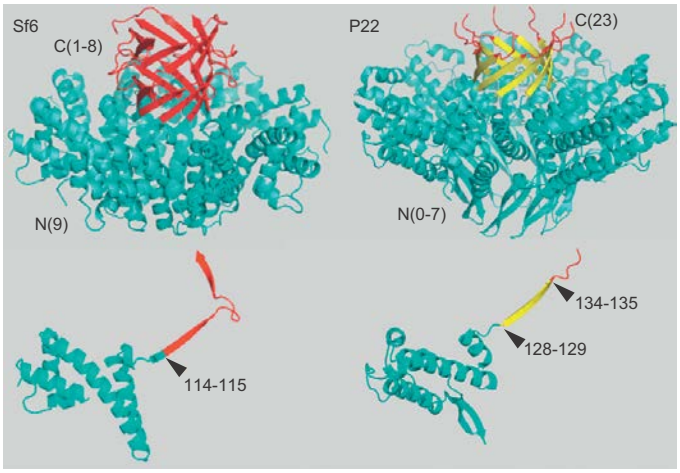


Fig. 4. Structures of the Sf6 and P22 TerS proteins. Phage Sf6 (left) and P22 (right) TerS protein structures are shown as ribbon diagrams (Protein Data Bank ID codes PDB 3HEF and 3P9A, respectively). The native octamer for Sf6 and nonamer for P22 are shown above and single subunits are shown below (Zhao et al., 2010; Roy et al., 2012). The locations of the N- and C-termini in the structures are indicated by "N" and "C", respectively, and the numbers in parentheses indicate the numbers of flexible terminal amino acids that were not seen in the structures even though they were present in the crystals (different subunits have different numbers of such "missing" amino acids at the Sf6 C-terminus and the P22 N-terminus). In the P22 *terS* Sf6-hybA and -hybB hybrid phages, the portion of the Sf6 protein shown in blue replaces the P22 blue or blue plus yellow sections, respectively; in the P22 single subunit, the fusion points for P22 Sf6-hybA and Sf6-hybB are shown by color changes and indications of the P22 amino acids of the hybrid proteins between which the fusion took place.

Sf6-hybB junction drastically reduces the *in vivo* activity of TerS. To determine what parts of the P22 portion of the functional Sf6-hybB protein might be most important for its function, triple alanine substitutions were made that replace sets of three adjacent codons between P22 codons 135 and 155 of the Sf6-hybB *terS* gene (Sf6-hybK through hybQ; similar changes could not be easily made in amino acids 156–162 because their codons overlap the *terL* gene). Of these mutants, changes from P22 TerS amino acids 138 to 143 did not inactivate the TerS protein, while changes in the 135 to 137 or 144 to 155 amino acid intervals did have a strong negative effect on phage yield after induction (Fig. 3). These results suggest that protein-protein (most likely TerS-TerL, above) contacts are not critical in the middle portion of the C-terminal domain of the Sf6-hybB TerS protein.

Since P22 Sf6-hybA and -hybB TerS proteins are both functional, it appeared that the length of the connection between the N-terminal DNA-binding domain and the essential parts of the C-terminal domain is not critical. We tested this by inserting four or eight "linker" amino acids between the Sf6 and P22 sequences of Sf6-hybA TerS protein to create P22 *terS* Sf6-hybT and -hybU (prophages of strains UB-2074 and -2075, respectively) (Fig. 3). Both of these changes result in plaque-forming phages, indicating that these TerS proteins with extended inter-domain linkers are functional. On the other hand the connection between the two TerS domains can apparently be too short. In P22 *terS* Sf6-hybG deletion of Sf6 TerS amino acids D113-K114 inactivates TerS, but in P22 *terS* Sf6-hybR (Fig. 3; strain UB-2072) replacement of D113-K114 with two alanine residues is functional. Thus, it appears that the inter-domain linker length rather than specific D113 and K114 side chain structure is required for TerS function. Comparison of the TerS proteins in P22 *terS* Sf6-hybD (two amino acids shorter than Sf6-hybB) and Sf6-hybU (fourteen amino acids longer than Sf6-hybB) show that the linker between the two domains can vary in length by at least 16 amino acids without inactivating the protein. We conclude that the precise spatial juxtaposition between the two TerS domains is much less important than the presence of a physical peptide backbone connection between them.

The Sf6 pac site

The hybrid phage experiments above suggest that the Sf6 *pac* site should lie within the first 112 codons of the Sf6 *terS* gene (bp 1–336 of the Sf6 genome; Accession no. AF547987), since this is the only Sf6 sequence in the smallest functional hybrid P22 *terS* Sf6-hybR. To test this idea directly, we engineered inverted, nonadjacent duplications of sections of this region into a non-essential location of the Sf6 genome and tested their ability to program the initiation of packaging series. Fig. 5 describes this strategy, which is patterned after our previous genetic analysis of the phage P22 *pac* site (Wu et al., 2002). Packaging series initiation DNA cleavage events can be visualized in agarose gels as the restriction fragment from the initiating end of the DNA of the first packaging event in a series; this fragment is called the "pac fragment," and a packaging initiation cleavage near the *pac* site forms one end of the fragment and restriction endonuclease cleavage forms the other end (see Jackson et al., 1978; Casjens et al., 1992a; Wu et al., 2002). Thus, by analyzing whether or not a pac fragment is generated by a potential *pac* site, the packaging series initiating activity of such a site can be determined.

To generate Sf6 phages with duplicated sequences for such a test, we first constructed a defective Sf6 prophage in which a Kan^R cassette replaces the essential endolysin-encoding gene *62* (Fig. 5B and Materials and Methods). Second, a plasmid, pPP311, was constructed which carries a complete 62 gene, a multicloning site transcriptionally downstream of gene *62* into which a sequence to be tested for *pac* activity can be inserted, and additional Sf6 homology downstream of gene *62* (Fig. 5C). When this plasmid carrying a potential *pac* site in its multicloning site is transformed into a *S. flexneri* cell carrying the above Sf6 *62*::Kan^R defective prophage, and the prophage is induced to lytic growth with mitomycin C, the only plaque-forming phages produced

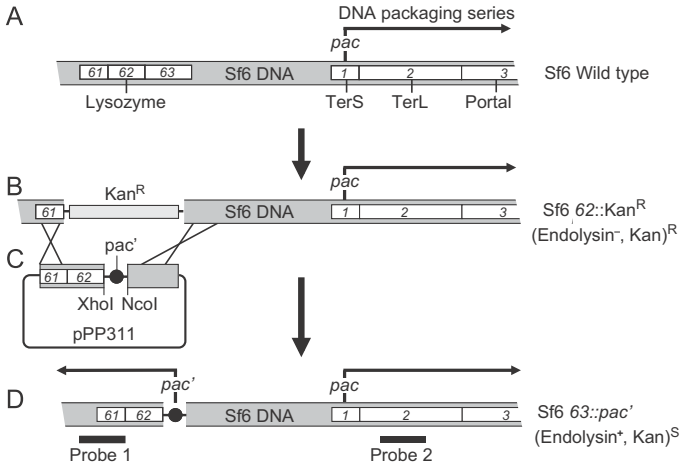


Fig. 5. Building phage Sf6 genomes with two *pac* sites. (A) The section of the Sf6 genome containing genes *61* (holin)-*62* (endolysin)-*63* (lambda *Rz* homolog) and *1* (small terminase)-*2* (large terminase)-*3* (portal protein). (B) The same Sf6 genome section as in part A with a kanamycin resistance cassette replacing genes *62* and *63* (prophage of *Shigella* strain UB-1564). (**C**) Plasmid pPP311 carrying a sequence (represented by a solid black circle) to be tested for Sf6 *pac* site activity; see text for its construction. (D) Sf6 *63*::pac' in which *pac* tester sequences from pPP311 derivatives have replaced the Kan^R cassette of the phage in part B. The black arrows indicates the direction of DNA packaging from an active *pac*' site replacing Sf6 gene *63* and from the native *pac* site in gene *1* (see text), and Southern probes 1 and 2 are defined in the text.

must have their Kan^R gene insertion replaced by the intact *62* gene and the adjacent potential *pac* site by homologous recombination with the plasmid; note that integration of the whole plasmid into the Sf6 genome makes it too large to be completely packaged. The orientation of such a potential *pac* site in the Sf6 genome is determined by its orientation in plasmid pPP311, and in the constructs studied here this orientation was designed to program any packaging series leftward on Sf6 DNA (Fig. 5D).

Sequences spanning various parts of the Sf6 *terS* gene were amplified using primers with XhoI and NcoI site-containing 3′-tails, and the resulting DNAs were inserted between the XhoI and NcoI sites of plasmid pPP311 so that their orientation was opposite to their native orientation in the *terS* gene (Fig. 5). These sequences were then moved into the Sf6 gene *63* region as described above. The resulting phages (strains UC-920 through UC-925; Table 1) form plaques, so the inserted second *pac* site (called *pac*' hereafter) does not interfere with lytic growth (as is also the case with phage P22; Wu et al., 2002). DNA was isolated from the virions thus produced, cut with restriction endonuclease PmeI or BsrGI, and the fragments displayed by 0.8% agarose gel electrophoresis. The gel was stained with ethidium bromide and Southern probed with probe 1 (Fig. 6; probe 1 DNA was PCR amplified from the bp 35981–36507 region of the Sf6 genome, so that it hybridizes to any pac' fragments extending leftward from the XhoI-NcoI cloning site). The PmeI and BsrGI enzymes were chosen because they produce pac' fragments in the 1–4 kbp size range, where analysis of the width of the diffuse pac' fragment band is most accurately performed, and because their overlapping true restriction fragments are much larger than their pac' fragments and so will not interfere with the analysis. Fig. 6A shows that, as expected, if nothing is inserted into the pPP311 cloning site, Southern analysis with probe 1 of the DNA from the resulting phage (UC-920) shows no leftward-extending pac' fragment, indicating that packaging series are not initiated on the DNA present between the XhoI and NcoI sites of pPP311 (or at other fortuitous sites in this region). On the other hand when Sf6 sequence A (Sf6 pac'A bp 1–200; Fig. 6A) was inserted between the XhoI and NcoI sites (phage UC-921), a diffuse band in the Southern autoradiograph is centered at about 1.7 kbp for the BsrGI digest and 3.5 kbp for the PmeI digest (Fig. 6A). The BsrGI-PmeI double digest diffuse band is identical to that in the BsrGI digest, showing that the variable end of this family of fragments is the right end and thus is where packaging initiated. The diffuse PmeI and BsrGI pac' bands extend from about 2.5 to 4.4 kbp and 0.9 to 3.0 kbp in the electrophoresis gel, respectively, and so both have measured widths of 2.0 ± 0.1 kbp (Fig. 6). Probing BsrGI cut DNA with probe 2 (amplified from bp 761–1408 of Sf6 chromosome) showed only the diffuse pac fragment band (at ~6000 bp) initiating from the native *pac* site in the *terS* gene region (data not shown; see also Casjens et al., 2004). Insertion of Sf6 DNA fragment B (Sf6 pac'B bp 200–424) into the XhoI-NcoI insertion site did not show any pac' fragment (data not shown). This finding indicates that packaging series initiation is sequence-specific in that any Sf6 sequence at this position is not sufficient to cause initiation.

The directionality of packaging from the pac'A sequence above is the same as we previously reported for the native Sf6 packaging, and its pac' fragment band width is similar to what was previously reported for normal packaging series initiation by wild type phage Sf6 (Casjens et al., 2004). The Sf6 *pac* site was located more precisely through the analysis of the ability of smaller Sf6 DNA regions to generate pac' fragments. These DNA fragments are shown in Fig. 7. Fragment C (Sf6 bp 149–424) gave rise to a pac' fragment of similar intensity to that from fragment A (data not shown). Since these findings suggested a region with *pac* site activity in the overlap between fragments A and C, shorter fragments D (bp 140–210) and E (bp 149–225) from this overlap region (Fig. 7) were also inserted at the XhoI-NcoI insertion site as described above to create phages UC-924 and UC-925, respectively. Both cause the generation of diffuse pac' fragments (data not shown). These observations strongly support the idea that these DNA sequences at the *pac*' site are programming packaging series initiation in a fashion that accurately reflects wild type initiation. We conclude from these experiments that the information for the Sf6 *pac* site lies in the bp 149–210 interval within the Sf6 *terS* gene.

Because of technical difficulties in accurately analyzing such highly diffuse DNA "bands" in electrophoresis gels, we devised

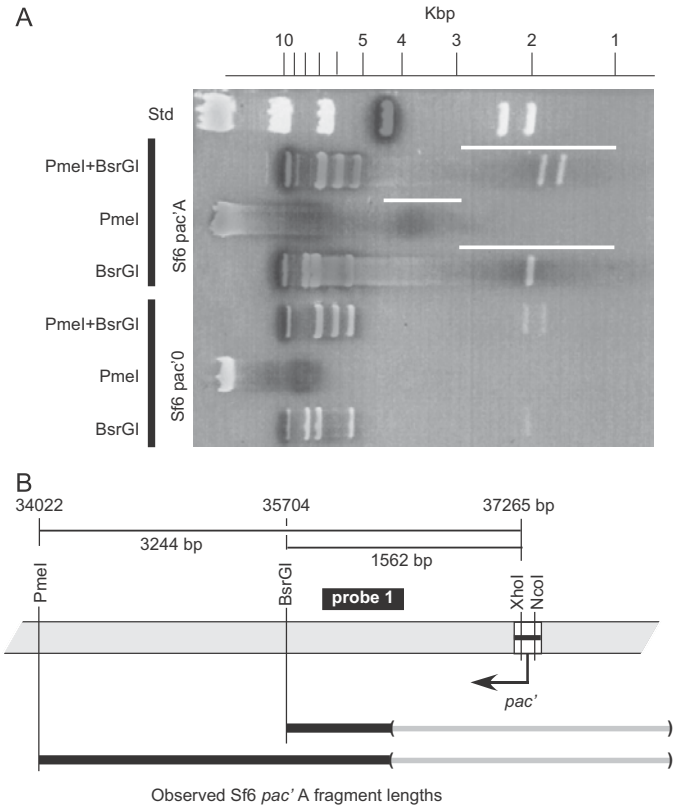


Fig. 6. Pac fragments generated by a second *pac* site in phage Sf6. (A) A 0.8% agarose electrophoresis gel of Sf6 pac'0 (strain UC-920) and Sf6 pac'A (UC-921) DNA cleaved with the indicated restriction enzymes (phage λ DNA cleaved with HindIII serve as size standards in the rightmost lane). White bands are the DNA fragments stained with ethidium bromide. Superimposed on the stained gel is an autoradiogram of the same gel probed with probe 1, a PCR amplified DNA fragment that extends from Sf6 bp 35981 to 36520 (hybridizing DNA shown in black; the Sf6 probe cross-reacts with one lambda DNA band as expected since the Sf6 and lambda holin genes are nearly identical). The white vertical lines to the right of the Sf6 pac'A lanes mark the width of the pac' fragment bands. (B) Above, a map of the gene *62*-pac' region (see Fig. 5) of the Sf6 *63*::pac0 (UC-920) genome is shown with distances between relevant restriction sites. The location of Southern probe 1 is indicated by a black bar. Below, horizontal lines represent the observed pac fragments in part A that were generated by packaging initiation in the pac'A sequence. The left ends of these lines are anchored to the locations of the restriction sites that were cleaved to generate the fragments, and the parentheses at the right ends enclose the regions in which the right ends of these fragments occur.

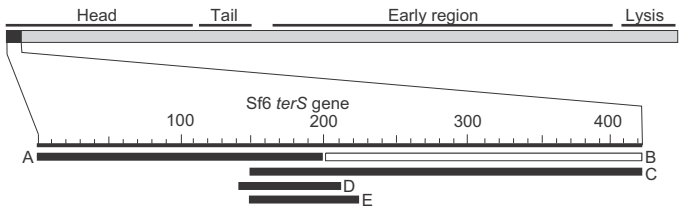
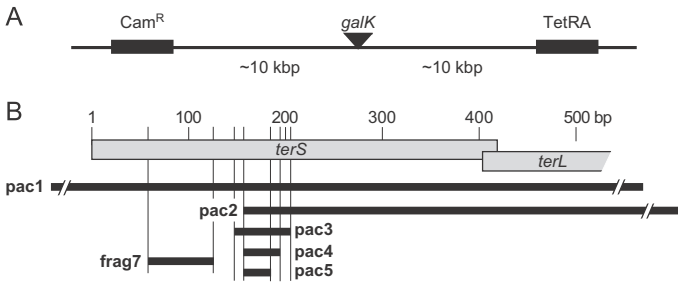


Fig. 7. DNA fragments tested for Sf6 *pac* activity. A map of the Sf6 genome is shown with its functional regions indicated above and the *terS* gene (gene *1*) shown in black. Below, the DNA fragments (A–E, see text) that were test for Sf6 *pac* activity are indicated as horizontal bars for which black and white indicate activity and no activity, respectively (see text).

a second method for genetic identification of the Sf6 *pac* site. P22 that is growing lytically can initiate packaging on induced, defective prophages in the *Salmonella* chromosome (Weaver and Levine, 1978; Youderian et al., 1988); however, the sequences required for such initiation were not analyzed in detail. It therefore seemed possible that a *pac* site inserted into the host bacterium's chromosome would program a high frequency of transduction of markers near that site. As a proof of principle we first inserted a short sequence containing the previously characterized P22 *pac* site (Wu et al., 2002) into the *Salmonella* chromosome and tested its ability to cause an increase in transduction frequency of nearby markers by phage P22. To more easily monitor transduction, we used recombineering methods to construct *Salmonella* strain UB-1982 in which tetracycline (TetRA)

and chloramphenicol (Cam^R) resistance genes are inserted 10 kbp on either side of the native *galK* gene in the chromosome (Fig. 8A); 10 kbp is long enough that the effects of any exonucleolytic nibbling at DNA ends that might occur during transduction packaging events that initiate at *pac* sites at the *galK* location (see below) should be minimized, and the drug resistance genes would be packaged in the first headful of any packaging series that might initiate from such a *pac* site (details of strain constructions are in Materials and Methods). The *galK* gene of UB-1982 was then replaced by a 40 bp P22 sequence containing the *pac* site in either of the two possible orientations, and the resulting strains were lysogenized by P22 UC-911 (Table 1). The prophages of these strains (UB-1988 and UB-2120), as well as control strains in which the *galK* gene was present (UB-2119) or neatly deleted (UB-2118) were induced, and the frequencies of transduction of Cam^R and TetRA by the resulting lysates were measured as described in Materials and Methods. Fig. 8C line 3 shows that the inserted sequence P22 pacL stimulates Cam^R transduction about 100-fold over the control strains UB-2118 and UB-2119 (lines 1 and 2) that have no inserted *pac* site. This stimulation depends on the orientation of the *pac* site, and the orientation of P22 pacL should program packaging only in the direction of the Cam^R marker (Jackson et al., 1978; Casjens et al., 1992a; Wu et al., 2002). The P22 pacR site oriented in the opposite direction does not confer a stimulation of Cam^R transduction. Inserted *pac* site stimulation of transduction of the TetRA cassette by P22 is less striking because of a rather high background transduction when no inserted *pac* site is present ("none" or "galK" in Fig. 8C lines 1 and 2). This is likely due to a natural *pac*-like site in *Salmonella* DNA that programs this transduction; nonetheless, at least a 3-fold increase of TetRA transduction was consistently observed when the P22 *pac* site was oriented so that it directs packaging towards the TetRA cassette (P22 pacR). We conclude that this 40 bp sequence is sufficient to program P22 packaging series initiation on the *Salmonella* chromosome. We also note that the previous experimental characterization of the P22 *pac* site had only shown that this region is *necessary* for packaging initiation (Wu et al., 2002), and this is the



C

DNA	Transduction frequency Cam^R	Transduction frequency TetRA	Transducing phage	Donor strain
None	1.1	20.3	P22	UB-2118
galK	0.5	9.8	P22	UB-2119
P22 pacL	74.3	2.8	P22	UB-1988
P22 pacR	0.3	63.9	P22	UB-2120
Sf6 pac1L	0.9	11.4	P22	UB-2121
Sf6 pac1R	0.1	8.6	P22	UB-2124
Sf6 pac2R	0.1	8.6	P22	UB-2123
None	0.6	2.2	Sf6-hybA	UB-2099
galK	0.1	2.5	Sf6-hybA	UB-2093
P22 pacR	0.3	1.8	Sf6-hybA	UB-2094
Sf6 pac1L	11.2	6.5	Sf6-hybA	UB-2101
Sf6 pac1R	0.3	33.0	Sf6-hybA	UB-2104
Sf6 pac2R	0.3	15.2	Sf6-hybA	UB-2105
Sf6 pac3L	4.9	0.8	Sf6-hybA	UB-2095
Sf6 pac3R	0.3	32.1	Sf6-hybA	UB-2103
Sf6 pac4R	0.5	23.6	Sf6-hybA	UB-2102
Sf6 pac5R	0.1	16.9	Sf6-hybA	UB-2100
Sf6 frag7	0.4	2.1	Sf6-hybA	UB-2106

Fig. 8. Transductional measurement of Sf6 *pac* activity. (A) Chloramphenicol and tetracycline resistance cassettes in the *Salmonella* strain UB-1982 chromosome. (B) The locations of the Sf6 DNA fragments that were used to replace the *galK* gene in the *Salmonella* chromosome are indicated by horizontal black bars (exact coordinates given in Materials and Methods). (C) Transduction frequencies are presented transduced colonies per plaque-forming phage particle ($\times 10^{-6}$), measured as described in Materials and Methods. The leftmost column indicates the DNA that was present in the host donor chromosome, where L and R indicate that the *pac* sites were oriented to program DNA packaging towards Cam^R or TetRA from the insertion site at *galK*. Three or more replicates were performed for each transduction that gave similar results, and a representative set of results is shown.

first demonstration that it is also *sufficient* for initiation. Thus, there are likely no other sequences in the P22 genome that are required in *cis* for packaging in addition to the known *pac* site.

Since this *in vivo* assay for *pac* site activity behaves as predicted for phage P22, we replaced the *Salmonella* UC-1982 *galK* gene (Fig. 8A) with several sections of the phage Sf6 genome and determined their ability to support initiation of packaging (*i.e.*, increased transduction frequency) during lytic growth of P22 *terS* Sf6-hybA (UC-929). These inserted sequences are diagramed in Fig. 8B (see Materials and Methods for precise endpoints). They were chosen from the Sf6 *terS* gene and environs, since the experiments above indicated that this region harbors the Sf6 *pac* site. Each of these strains was lysogenized with P22 *terS* Sf6-hybA, and the transduction of the Cam^R and TetRA markers was measured after induction of the prophage to lytic growth. Six different overlapping Sf6 DNA sequences (Sf6 pac1-pac5 and frag7; Fig. 8B) were tested for P22 *terS* Sf6-hybA *pac* activity, several of them in both orientations. The frag7 sequence (UB-2106) did not show an increase in transduction over strains that carried the native *galK* gene (UB-2093) or a neat deletion of the *galK* gene (UB-2099) indicating that it does not contain an Sf6 *pac* site. The five remaining inserted Sf6 sequences all mediate unidirectional transduction increases (Fig. 8C lines 11–17). For example, in one orientation the Sf6 pac3R fragment (Sf6 bp 145–203; Fig. 8C line 15) stimulates transduction of TetRA about 15-fold over strains with no *pac* site, but does not stimulate transduction of Cam^R; in the other orientation the same sequence (Sf6 pac3L; line 14) stimulates Cam^R transduction about 10-fold but has no effect on TetRA transduction. The direction of these transduction increases agrees perfectly in all cases with our previous determination of the direction of native Sf6 DNA packaging (Casjens et al., 2004) and with the results from our Southern analysis of phages with duplicated *pac* sites above. The difference in background transduction frequency of Cam^R and TetRA without inserted *pac* sites could be due to several factors, but (as with P22 above) the presence of an Sf6 *pac*-like site in the native *Salmonella* chromosome sequence that programs some TetRA transduction by P22 Sf6-hybA is a likely explanation.

Replacement of the *galK* gene by Sf6 pac1R, pac2R, pac3R, pac4R or pac5R sequences all gave rather similar 10- to 20-fold increases in TetRA transduction (Fig. 8C). The smallest Sf6 sequence tested, pac5R, is sufficient for the bulk of this effect, and thus at least most of the Sf6 *pac* site lies within Sf6 bp 154–183. We also found that P22 TerS does not utilize the Sf6 *pac* site (compare UB-2121, -2123 and -2124 with UB-2118 and UB-2119 in Fig. 8C), and the Sf6-hybA TerS does not utilize the P22 *pac* site (UB-2094 vs. UB-2093 and UB-2099 in Fig. 8C). This data strongly supports the idea that the N-terminal globular TerS domain of Sf6 is responsible for *pac* site recognition *in vivo*. All of the above experiments support the idea that the phage Sf6 *pac* site lies inside the Sf6 *terS* gene between bp 154 and 183, and that packaging proceeds rightwards on the Sf6 genome from recognition events at that site. A more detailed analysis of the Sf6 *pac* site will be the subject of future studies.

TerS controls the location of packaging initiation DNA cleavages

As was described above, Sf6 generates packaging series initiation DNA ends over a large approximately 2000 bp region. This distribution can be visualized in agarose electrophoresis gels as the pac DNA fragment(s) (the restriction fragment from the initiation end of the DNA of the first member of a packaging series, above). In order to determine whether the TerS source affects the distribution of packaging initiation DNA cleavages, we analyzed the pac fragments of the functional TerS hybrid phages P22 *terS* Sf6-hybA (UC-921) and Sf6-hybB (UC-922) in comparison to that of P22 (UC-911). Fig. 9 shows that after NdeI cleavage P22 DNA has a rather sharp ~3 kbp long pac fragment band as was previously known (Jackson et al., 1978; Casjens et al., 1992a); however, P22 *terS* Sf6-hybA DNA has a broad and diffuse NdeI pac fragment band that is about 2 kbp wide. Southern analysis was used to visualize these bands to avoid confusion regarding the source of the DNA bands. As above, restriction enzymes were chosen to display the pac fragments in the 2–5 kbp size range and to ensure that the overlapping true restriction fragments do not migrate at the same position as the pac fragments in the electrophoresis gel. Restriction enzyme analysis again showed that the packaging initiation (non-restriction enzyme)-generated ends of the diffuse pac fragment band were centered on the terS gene and that packaging proceeds rightward from the *pac* site (Fig. 9B). Parallel analysis of P22 *terS* Sf6-hybB DNA gave identical results (data not shown). The

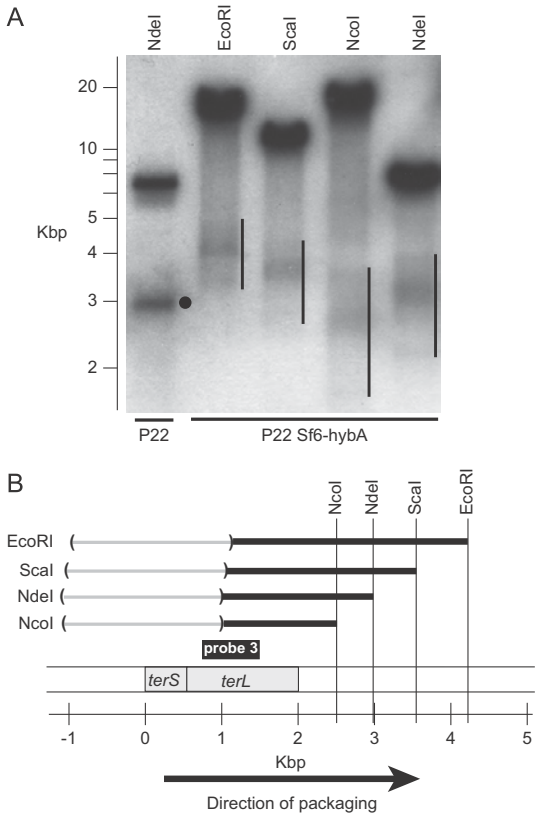


Fig. 9. Pac fragments of P22 Sf6-hybA. (A) Southern analysis of P22 Sf6-hybA pac fragments. DNA from P22 *3*::Sf6-hybA, *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1 DNA was cleaved with the restriction endonuclease indicated above and separated by electrophoresis in 1.0% agarose. The gel was probed with ^{32}P-labeled probe 3 (PCR amplified from bp 732 to 1497 of P22 DNA) and exposed to x-ray film. The black circle marks the P22 NdeI pac fragment location, and vertical black lines mark the widths of the diffuse pac fragments of P22 Sf6-hybA DNA; the bands higher in the gel are the true restriction fragments (restriction cuts at both ends) that are generated from packaging events other than the first one in a series. (B) A map of the packaging series initiation region of P22 Sf6-hybA is shown with a scale in kbp below it; the locations of the *terS* and *terL* genes are noted on the map. Horizontal black lines represent the pac fragments in part A that were generated by initiation of packaging series on P22 Sf6-hybA DNA. The right ends of these lines are anchored to the locations of the restriction sites that were cleaved to generate the pac fragments, and the parentheses enclose the regions in which the left (packaging initiation) ends of these fragments occur.

approximately 2 kbp band width is similar to the approximately 1.6 kbp distribution measured for phage Sf6 itself (Casjens et al., 2004). The difference between the 2 kbp measured here and the 1.6 kbp reported previously is almost certainly due to the rather arbitrary nature of determining the positions of the outer edges of such very diffuse gel "bands"; re-analysis of our previous data (Casjens et al., 2004) indicates that the pac fragment band width in those experiments was slightly underestimated due to the use of more conservative criteria for determining the band's outer edges and is in fact indistinguishable from the band widths measured here. Since the N-terminal domain of the Sf6 *terS* gene is the only Sf6 genetic information present in the hybrid phages analyzed in this way, we conclude that this domain of TerS is responsible for the difference between the P22 packaging initiation cleavage-containing region of 120 bp and the much larger Sf6 region (see Discussion for possible mechanisms).

Discussion

The role of phage Sf6 TerS in DNA packaging

Analysis of the genome sequences of extant natural variants among the P22-like phages led us to the predictions that the C-terminal 23–28 amino acids of their TerS proteins are responsible for the interaction of TerS with the rest of the DNA packaging machinery and that their N-terminal domains recognize specific *pac* sites to choose the DNA molecules to be packaged. To test these ideas we constructed a fusion of the N-terminal domain of the phage Sf6 TerS with the C-terminal portion of P22 TerS in an otherwise completely phage P22 context, a combination that has not yet been found in nature. Several conclusions can be drawn from the fact that such hybrid phages are functional and from our genetic analysis of this hybrid TerS protein.

(1) The N-terminal domain of TerS is responsible for the sequence specificity of *pac* site recognition *in vivo*, since the hybrid TerS with an Sf6 N-terminal domain and P22 C-terminal domain utilizes the Sf6 *pac* site and not the P22 *pac* site. Similarly, as expected from our previous analysis of its *pac* site (Wu et al., 2002), we found that phage P22 TerS cannot utilize the Sf6 *pac* site. However, we (Nemecek et al., 2008) and Roy et al. (2012) have found that removal of twenty or more C-terminal residues from P22 TerS results in a protein that no longer binds DNA nonspecifically *in vitro*. This observation is in apparent conflict with the idea that the N-terminal domain of TerS is solely responsible for DNA binding and could indicate that the C-terminus might also participate in some way.

(2) Since the *pac* recognition specificities Sf6 and P22 TerS proteins are different (neither utilizes the other's site), the functional P22 *terS* Sf6-hybA and -hybB phages should be utilizing an Sf6 *pac* site that resides within the DNA that encodes the N-terminal domain of Sf6 TerS. Our experimental analysis showed that phage Sf6 does indeed carry its *pac* site near the center of this region.

(3) Purified P22 TerS protein forms a complex with TerL (Poteete and Botstein, 1979; Roy et al., 2012; our unpublished results). The genetic findings presented here strongly suggest that the C-terminal domain of TerS is responsible for this binding, and while this work was underway Roy et al. (2012) showed that removal of the C-terminal 22 amino acids from P22 TerS abrogates its ability to bind TerL but not its ability to form nonamers. The evolutionary co-segregation of the C-terminal domain of TerS with the N-terminal half of TerL strongly suggests an interaction between these domains. Genetic studies with phages lambda and T4 TerS proteins indicate that their C-terminal regions also participate in binding to their cognate TerL's (Frackman et al., 1985; Yang et al., 1999a; Gao and Rao, 2011). Since these three terminases are extremely different, having essentially no recognizable sequence similarity beyond an ATPase motif in TerL, this appears to be a very ancient interaction strategy that has been preserved in spite of the long evolutionary divergence between these terminases.

(4) The N-terminal domains of P22 and Sf6 TerS have essentially no amino acid similarity and yet the Sf6 domain can function in an otherwise completely P22 context. It thus seems very unlikely that this domain has important intimate interactions with the rest of the DNA packaging motor. If such interactions existed, the many differences between the Sf6 and P22 TerS proteins would make the interaction very unlikely in the hybrid phages studied here.

(5) The fact that the connecting "linker" region between the essential parts of the N- and C-terminal domains of the TerS protein can vary in length by at least 16 amino acids and still be functional further suggests a very flexible connection between the N-terminal domain of TerS and the rest of the packaging machinery. The apparent flexibility of this connection is somewhat surprising, since it might have been expected that the different components of a DNA translocating molecular motor would have to occupy very specific spatial positions in the motor. However, there are other indications of TerS flexibility which include a mutant of P22 whose isolated TerS is present as decamers instead of nonamers that is functional *in vivo* (Nemecek et al., 2008), P22 TerS has 23 C-terminal amino acids that are not ordered in crystals (Roy et al., 2012), and analyses of alternate crystal structures of each of the SF6, T4 and Sf6 TerS proteins indicate that they have considerable structural flexibility (Buttner et al., 2012; Sun et al., 2012; Zhao et al., 2012).

(6) We also found that the N-terminal domain of TerS is responsible for the distribution of packaging series initiation DNA cleavages, since a P22 that carries only the DNA-binding domain of Sf6 TerS has a packaging initiation end distribution similar to that of phage Sf6 and not to that of phage P22. This could be explained by terminase moving along the DNA between *pac* recognition and DNA cleavage (see below), and if so this putative sliding/rolling would be due to movement of TerS or the TerS-TerL complex on the DNA and not, for example, after the DNA might be handed off from TerS to TerL.

Unlike the cohesive end generating terminases, the terminases of the headful packaging phages that have been examined do not generate precise DNA ends at the start of packaging series (summarized in Wu et al., 2002 and Casjens et al., 2005). We previously showed that two such tailed phages, Sf6 and ES18, generate packaging initiation ends over large regions of about 2000 and 500 bp, respectively (Casjens et al., 2004, 2005). The frequency of such ends across the Sf6 DNA indicates that the frequency of cleavage decreases with increasing distance from near the center of the region (Casjens et al., 2004). We report here the localization of the phage Sf6 *pac* site to a short sequence near the center of this region, which is consistent with models that have a decreasing probability of cleavage with increased distance from the *pac* site.

The mechanism by which such a multiplicity of packaging initiation ends might be generated from targeted *pac* site recognition events is not known, but several models can potentially account for this observation. (i) The region that contains the packaging series initiation ends contains a number of recognition (*pac*) sites, any of which can be used to initiate packaging series. Alternatively, a single terminase complex could bind *pac* and

either (ii) cut nearby DNA that happens to bend or loop into contact with it, (iii) recruit additional adjacent terminase complexes to bind nonspecifically and create a patch of terminase-covered DNA in which any terminase could generate the initiation end for a given packaging series, or (iv) move in either direction along the DNA away from the *pac* site before cleaving the DNA, perhaps by "sliding" or "rolling" of the multimer? Finally, (v) ends created at *pac* by a terminase cleavage could be attacked by an exonuclease before actual packaging starts. The localization of the Sf6 *pac* site to a small region rules out models that depend on the presence of many *pac* sites. Furthermore, the "patch of bound terminases" model seems rather unlikely since the terminase subunits, especially TerL, are not made in large amounts (Casjens and King, 1974; Poteete and Botstein, 1979), and because in our unpublished experiments substantially lowering the amount of P22 TerS during infection did not cause the initiation cleavage events to become less frequent farther from the *pac* site in that phage. If terminase bound to the *pac* site cleaves nearby DNA that has looped into contact with it, the nuclease active site should be separate from the DNA binding site and DNA cleavage should not occur within the *pac* site itself; however, in P22 cleavages are made inside the *pac* site (Casjens et al., 1992a; Wu et al., 2002) which would make this model less tenable (but multimeric TerL structures in which only one subunit is bound to *pac* could obviate this argument). Finally, although terminases have endonuclease activity, there is no evidence for exonuclease participation in packaging in any phage system, and such a model demands initial cleavage far upstream from the *pac* site (in the packaging directionality sense), and it is difficult to imagine how the ends could then be most frequent near the *pac* site. Thus, by default, current evidence suggests that the sliding/rolling model is most likely, although it has not been demonstrated directly. Understanding the mechanism of any such movement must await further experimentation.

Whatever the true mechanism, the four headful phages whose packaging initiation DNA cleavages and *pac* site have been previously precisely characterized, P22, Mu, P1 and SPP1, generate a number of alternate packaging initiation ends within approximately 120, 150, 12 and 7 bp regions, respectively, and all have a *pac* recognition site near or within the region within which the ends are generated (Deichelbohrer et al., 1982; Groenen and van de Putte, 1985; Sternberg and Coulby, 1987; Harel et al., 1990; Casjens et al., 1992a; Chai et al., 1995). Phage T4 likely uses this strategy as well, but it is less well characterized in this regard (Wu and Black, 1995; Wu et al., 1995). In phage P22 the recognition site, *pac*, is a 22 bp sequence that lies approximately in the center of the 120 bp region where the DNA ends are generated (Wu et al., 2002), so this relationship is similar in P22 and Sf6 although the cleavage sites extend much further away from the *pac* site in Sf6 than they do in P22, suggesting a possible underlying mechanistic similarity between them. On the other hand we also note that tailed phages are extremely diverse and variations on any theme are not unexpected. It is known, for example, that phage Mu terminase binds its *pac* site near the end of the phage genome (which is integrated into the host chromosome) and only makes DNA cleavages in one direction from the *pac* site, in the host DNA that is adjacent to the integrated phage DNA. Nonetheless, the lack of precise DNA cleavage during initiation of packaging appears to be a common feature among headful packaging phages.

TerS and pac site evolution and horizontal exchange

The shortest Sf6 sequence that we tested that has *pac* site activity was the 30 bp fragment pac5R (Fig. 8), and its sequence is related to the P22 *pac* site. Nine of the 13 bp that are known to be important in P22 *pac* site recognition (Wu et al., 2002) are present in this Sf6 sequence if a two bp deletion in P22 is allowed (7 of 13 if not; Fig. 10A). Although the Sf6 and P22 TerS proteins do not have convincing sequence similarity, they do have partly similar folds. They both contain a cluster of five α-helices that have the same connectivity and similar but not identical spatial arrangements (Fig. 10B and Fig. 2 of Roy et al., 2012). A major difference between these two protein structures is the replacement of the short connection between α-helices 2 and 3 in Sf6 by a rather long β-hairpin in P22 (Fig. 10B). Although the minimal regions currently known to contain the P22 and Sf6 *pac* sites encode amino acids 89–96 and 51–61 of the two TerS proteins, respectively, it is striking that these two regions actually lie in precisely the same spatial location in the two proteins. Fig. 10B shows that both reside at the N-terminus of α-helix 5. It is also interesting to note that this region is at the outer rim of the TerS multimer and that this region of the Sf6 octamer likely contacts DNA in its *in vitro* non-sequence-specific DNA-binding activity; in particular, Sf6 TerS Lys59, which is encoded within the *pac* site region identified here, is likely in contact with bound DNA (Zhao et al., 2010, 2012). Of the seven amino acids encoded by the P22 *pac* site, there are two identities and three similarities in the parallel Sf6 sequence (Fig. 10A). If, as seems likely, the site on TerS that binds DNA nonspecifically *in vitro* is also involved in specific *pac* site recognition *in vivo,* then the *pac* site DNA itself encodes at least some of the amino acids that participate

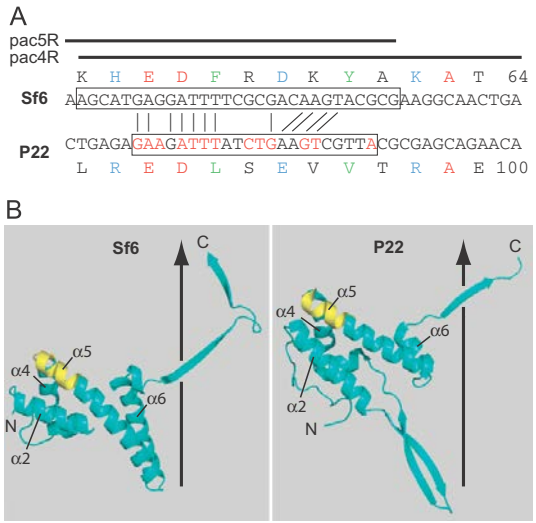


Fig. 10. P22 and Sf6 *pac* sites. (A) The nucleotide sequences surrounding the Sf6 and P22 *pac* site regions are compared. In both cases packaging proceeds rightward from the *pac* site as shown here; red nucleotides indicate those bp in which mutations are known that largely abolish P22 *pac* site activity (Wu et al., 2002). The nucleotides shown are Sf6 bp 154–194 (Accession no. AF547987) and P22 nucleotides 41722–41724 and 1–38 (Accession no. BK00583; this sequence crosses the position at which the circular sequence is opened in the GenBank annotation). The minimal regions that have *pac* activity are boxed. The amino acids encoded by these sequences are shown above (Sf6) and below (P22) the nucleotide sequences with the number of the last amino acid shown on the right. Identical amino acids in the two phages are shown in red, similar amino acids are shown in blue (same charge) and green (hydrophobic). Above, the horizontal black bars indicate the bps present in the pac4R and pac5R fragments tested for *pac* activity in the transduction assay (see text and Fig. 8B). (B) Phage Sf6 and P22 TerS single subunit protein ribbon diagrams are shown (Protein Data Bank ID codes 3P9A and 3HEF, respectively). The N- and C-termini are indicated, as are α–helices 2, 4, 5 and 6 (Zhao et al., 2010; Roy et al., 2012). The yellow portions mark amino acids 89-97 encoded by the *pac* site of P22 and amino acids 51–61 encoded by the shortest Sf6 *pac*-containing fragment tested (pac5R) in this report. The locations of the central channels of the TerS multimer rings (see text) are indicated by vertical black arrows.

directly in recognition of the *pac* site DNA. This would represent a very striking example of molecular efficiency and evolutionary ingenuity. Such an arrangement would ensure that the *pac* site cannot be separated from the protein region that recognizes it during horizontal exchange of genetic material among different phages. It thus seems likely that the P22 and Sf6 TerS proteins are ancient homologs that have diverged a great deal, but have retained common DNA binding features.

We showed here that the phage Sf6 *pac* site lies within its *terS* gene; however, not all tailed phage packaging initiation sites lie within the gene that encodes the TerS protein. For example in phages like lambda whose intravirion (mature) DNAs have sequence-specific, single-strand protruding cohesive ends, the recognition site (called *cos* in these phages) is typically immediately upstream of the adjacent *terS* and *terL* genes (e.g. lambda, N15 and HK97 and many other phages, but there are exceptions, for example phage, *E. coli* phage P2 (Ziermann and Calendar, 1990; Linderoth et al., 1991) and mycophage Giles (Hatfull, 2012) where the *cos* sites are far from the *terS* gene). These *cos* sites are typically more complex than the *pac* sites of headful packaging phages and contain dyad symmetry (Feiss and Rao, 2012), so it may be difficult to integrate them into the *terS* gene in these cases. The *pac* sites of five headful packaging phages have been studied (see above). The *pac* sites of P22 (Wu et al., 2002), P1 (Sternberg and Coulby, 1987; Lobocka et al., 2004), SPP1 (Chai et al., 1995) and Sf6 (shown here) lie inside the *terS* gene, and less direct evidence has indicated a possible site within the T4 *terS* gene (Wu and Black, 1995; Wu et al., 1995). However in one less well-studied headful packaging phage, T1, the *pac* site appears to be in the early region and not within the *terS* gene (Liebeschuetz and Ritchie, 1986; Roberts et al., 2004). Similarly, the phage Mu (a different sort of headful packager, above) *pac* site is nearly half the genome away from the TerS gene (Groenen and van de Putte, 1985; Harel et al., 1990; Morgan et al., 2002). Thus, although it is not universal, it is at least very common for the *pac* sites of headful packaging phages to lie with the *terS* gene.

Our findings thus support a picture in which the DNA that encodes the N-terminal domain of TerS is a self-contained, exchangeable unit that contains the *pac* site. It encodes the protein domain than binds the *pac* site and can function to recognize DNA for phage packaging as long as it can bind to TerL in the packaging apparatus through its C-terminal domain. This arrangement has allowed the horizontal exchange of *terS* genes among phages to be very successful, but the nature of the evolutionary advantage that might be gained by such exchanges remains mysterious.

Materials and methods

Phage and bacterial strains

All bacterial and phage strains used in this study are listed in Table 1. *S. enterica* UB-0002 was used to propagate phage P22 strains and *S. flexneri* PE577 (UB-1458) was used to propagate phage Sf6 strains. *E. coli* NF1829 was used to carry plasmids and as transformation recipient during plasmid construction. All plasmid and phage constructs were confirmed by determination of the sequence of the modified region.

Construction of Sf6 phages with two pac sites

In order to genetically move sequences of choice into the phage Sf6 genome, a plasmid into which such sequences can be inserted and an Sf6 phage that accepts these sequences from the plasmid were constructed. The DNA accepting prophage, Sf6 *62*::Kan^R, was constructed as follows: Plasmid pPP309 was constructed by ligating the following three polymerase chain reaction (PCR) amplified DNA fragments into plasmid pUC18 (Yanisch-Perron et al., 1985) that was opened with restriction enzymes HindIII and EcoRI, transforming into *E. coli*, and selecting for ampicillin resistance: (i) Sf6 bp 35805–36769 (Casjens et al., 2004; Accession no. AF547987) with HindIII and BamHI site containing primer tails at the two ends, respectively, cut with these two enzymes; (ii) kanamycin resistance cassette amplified from bp 1823 to 2741 of plasmid pACYC177 (Rose, 1988) with BamHI and NdeI site containing primer tails, respectively, cut with these two enzymes; and (iii) Sf6 bp 37247–38249 with NdeI and EcoRI site containing tails, respectively, cut with these two enzymes. The resulting plasmid insert contains Sf6 bp 35805–38249 except the Kan^R cassette replaces bp 36770 and 37246. Plasmid pPP309 was electroporated into *S. flexneri* strain PE577 that carried a wild type Sf6 prophage. This strain (UB-1469) was induced to lytic growth with 1 μg/ml mitomycin C, phage particles from the resulting lysate were used to infect *S. flexneri* strain PE577, and the surviving cells were selected for growth in the presence of 50 μg/ml kanamycin and screened for cells that did not carry ampicillin resistance of the pPP309 plasmid. The resulting Kan^R Amp^S bacterial strain (UB-1564) carries a prophage, Sf6 *62*::Kan^R, in which the Kan^R cassette replaces gene the essential gene *62* (phage encoded endolysin). It does not release plaque-forming phages upon induction with mitomycin C.

A plasmid designed to carry the sequences to be placed in the Sf6 genome was constructed as follows: (i) Sf6 bp 36472–37247 were PCR amplified with HindIII and XhoI site containing primer tails, respectively, cut with HindIII and XhoI, and ligated to HindII+SalI cut pUC18 plasmid (New England Biosciences, Ipswich, MA); and (ii) Sf6 bp 37606–38628 were PCR amplified with BglII–XhoI–KpnI–NcoI and EcoRI site containing primer tails, respectively, and ligated into the plasmid from step (i) that was cut with BamHI and EcoRI. The resulting plasmid, pPP311, contains Sf6 bp 36472–38628 in which 36 bps (5′–GTCGACTCTAGAGGATCTCTCGAGGGTACCCCATGG) that contain unique XhoI and NcoI sites (underlined) replace Sf6 bp 37247–37606 (Fig. 5). The latter replacement inactivates gene *63*; this spanin-encoding Sf6 gene is a homolog of phage λ gene *Rz* and phage P22 gene 15 and is not essential in those phages under low divalent cation growth conditions (Casjens et al., 1989; Summer et al., 2007). This work showed that Sf6 gene *63* is also required for efficient plaque formation in the presence of divalent cations, so all gene *63* defective phages were grown in the presence of 10 mM NaCitrate (Casjens et al., 1989).

DNA sequences to be tested for *pac* activity were inserted between the XhoI-NcoI sites of pPP311 and then moved into the phage Sf6 genome as follows: Oligonucleotide primers with 3′-tails containing XhoI and NcoI cleavage sites were used to PCR amplify sequences from the Sf6 genome, or double-stranded oligonucleotides with near terminal XhoI and NcoI sites were synthesized. These DNA fragments were cleaved with XhoI and NcoI and ligated into similarly cleaved plasmid pPP311. The ligation reaction mixture was transformed into *E. coli* NF1829 (UB-0049), and the resulting plasmid structures were confirmed by restriction mapping and sequence determination. These plasmids were moved into *S. flexneri* PE577 (Sf6 62::*Kan*^{R}) UB-1564 by electroporation. These ampicillin and kanamycin resistant cells were grown to 2×10^8/ml in LB broth and induced with 1 μg/ml mitomycin C (Sigma, St. Louis, MO). After about 4 h of shaking at 37 °C, cell lysis was completed by shaking with a few drops of chloroform, and cell debris was removed by centrifugation. The resulting phages, UC-920 through UC-925 (Table 1), all of which form plaques on *S. flexneri* PE577 (UB-1458), must have acquired a functional endolysin gene (gene *62*) by homologous recombination with the plasmid. Integration of the whole plasmid into the

Sf6 *62*::Kan^R prophage makes the genome too large to be packaged, so the plaque-forming phages must have also enjoyed a second homologous recombination event in the bp 37606–38623 region that causes a neat replacement of the Kan^R cassette of Sf6 *62*::Kan^R by the intact endolysin gene and the sequences cloned into the XhoI–NcoI sites of pPP311 (Fig. 5).

P22 prophage recombineering

P22 prophages with Sf6-P22 hybrid *terS* genes were constructed by recombineering as follows: Primer oligonucleotides A and B (Table S3) for PCR amplification were used that amplify the *E. coli galK* gene expression cassette from plasmid pGalK as described by Warming et al. (2005). These primers had 50 nt 3′-tails that correspond to P22 sequence immediately 5′ of the gene 3 (*terS*) start codon and 3′ of its codon 134. This amplified DNA was electroporated into UB-1790 cells (this strain is *galK*$^-$ and carries a P22 *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, *sieA*$^-$Δ1 (UC-911) prophage (Padilla-Meier et al., 2012); Table 1), and colonies were selected that utilize galactose as the sole carbon source. The resulting strain (UB-1961) carries a *galK* expression cassette that replaces gene 3 codons 1-133 of the resident prophage. Here and below, in each recombineering step the phage lambda Red expression plasmid pKD46 (Datsenko and Wanner, 2000; Karlinsey, 2007) was present to stimulate recombinational replacement and was removed by growth overnight at 41 °C before the strain was used further. DNA containing the first 114 codons of phage Sf6 gene 1 (*terS*) was then PCR amplified from plasmid pET28a-gp1 (Zhao et al., 2010) with primers C and D (Table S3). Primer C contains the 50 nucleotides of sequence immediately upstream of P22 gene 3 and 20 nucleotides that correspond to the first 20 nucleotides of the Sf6 gene 1 (*terS*) at its 3′-end; primer D contains 50 nucleotides that correspond to P22 gene 3 codons 129-146 and 18 nucleotides that correspond to Sf6 gene 1 codons 109-114 at its 3′-end. This amplified DNA was electroporated into strain UB-1961, and *galK*$^-$ cells resistant to 2-deoxygalactose (Sigma, St. Louis, MO) were selected as described by Warming et al. (2005). This recombinational replacement of the *galK* gene of UB-1961 resulted in a strain (UB-2019) whose P22 prophage contains a hybrid *terS* gene (called *terS* Sf6-hybA) in which Sf6 gene 1 codons 1-114 neatly replace codons 1-134 of P22 gene 3. Oligonucleotide C and variants of D were used in the same manner to construct prophages with *terS* genes that have different Sf6-P22 fusion points and junction sequences (hybA-hybJ and hybU-hybT; Table 1 and Fig. 3). *Salmonella* strains carrying P22 prophages with amino acid changes to alanine in the Sf6-P22 hybB *terS* gene were made as follows: The *terS* Sf6-hybB gene along with 224 upstream and 104 downstream bp was amplified from strain UB-2040 DNA using primers E and F (Table S3), and the resulting fragment was cleaved by restriction enzymes XbaI and HindIII and cloned into similarly cleaved plasmid pBLSK (Stratagene, La Jolla, *CA*). This plasmid, pPP465, was modified by QUICKCHANGE methodology (Qiagen, Valencia, *CA*) to make the bp changes in *terS* hybrids hybK-hybS (Table 1 and Fig. 3). The modified plasmid phage DNA inserts were PCR amplified and used to recombinationally replace the *galK* gene of UB-1961 as described above.

Insertion of pac sites into the Salmonella chromosome

Strain UB-1982 was constructed by sequentially inserting chloramphenicol and tetracycline resistance cassettes immediately 5′ of bp 827343 and bp 849069, respectively, into the *S. enterica* LT2 strain UB-0020 genome (LT2 Accession no. AE006468; recombineering details described above). Oligonucleotide pairs G/H and I/J (Table S3) were used with template DNA from strains UB-1760 and UB-1766, respectively, to PCR amplify fragments that contain the Cam^R and TetRA resistance cassettes with 40 bp of identity to the *S. enterica* LT2 chromosome at both ends. This placed the Cam^R-1 and TetRA-2 resistance cassettes about 10 kbp on either side of the *galK* gene which lies at bp 838373–839521. The Cam^R-1 insert lies between *Salmonella* open reading frames STM764 and STM765, and the TetRA-2 insert lies between STM784 and STM785; none of these reading frames has a known function, and we did not observe a growth phenotype for either insertion. In strain UB-2047 the *galK* gene was neatly deleted from UB-1982 by replacement it with synthetic oligonucleotide K (Table S3) annealed to its complementary oligonucleotide. The *galK* gene was replaced by the phage P22 *pac* site (P22 pacL and pacR in Fig. 8C) with recombineering replacement of *galK* in UB-1982 by annealed oligonucleotides L and M or N and O, respectively (Table S3). The resulting strains, UB-1985 and UB-1991, contain P22 nucleotides bp 41716–41724/1-31 in opposite orientations (this sequence crosses the opening of the circular P22 sequence in its GenBank annotation; Accession no. BK000583). Strains in which the UB-1982 *S. enterica* LT2 bp *galK* gene is replaced by phage Sf6 sequences (UB-2093-2095, 2099 and 2100–2106) to be tested for *pac* site activity were created in a manner analogous to the construction of strain UB-1991. DNA fragments containing Sf6 bp 61–120 (called "fragN"), 388891-690 (Sf6 pac1), 155–821 (Sf6 pac2), 145–203 (Sf6 pac3), 155–194 (Sf6 pac4), and 154–183 (Sf6 pac5), with 40 bp tails for replacing *galK* were prepared by PCR templated by Sf6 DNA or were made synthetically. Note that the Sf6 pac1 sequence crosses the opening of the circular Sf6 sequence in its GenBank annotation (Accession No. AF547987). Several of these were made with the Sf6 sequences in different orientations; L and R (as in Sf6 pac1L or pac2R) indicate an orientation for which the final construct packaging should proceed towards the Cam^R or Tet^R cassette, respectively. For use in transduction experiments, the strains in this paragraph were lysogenized by phage P22 *13*$^-$*am*H101, *15*$^-$Δsc302::kan^R, *sieA*$^-$Δ1 (UC-911) or P22 *13*$^-$*am*H101, *15*$^-$Δsc302::kan^R, 3::Sf6-hybA, *sieA*$^-$Δ1 (UC-929) by infecting and selecting for kanamycin resistance. Fig. 8C and Table 1 give the strain names and genotypes of these constructs.

Generalized transduction measurements

Transduction frequency measurements were performed as follows: The strains indicated in Fig. 8C were grown in L broth to 2×10^8/ml at 37 °C, induced by the addition of 1.5 μg/ml carbadox (Sigma, St. Louis, MO), and lysed with chloroform after 3 h of shaking at 37 °C (the strains each carry a P22 *sieA*$^-$Δ1, *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R or a P22, *sieA*$^-$Δ1, *13*$^-$*am*H101, *15*$^-$Δsc302::Kan^R, 3::Sf6-hybA prophage). The resulting lysate was titered on UB-0002 to determine the phage yield. Cam^R or Tet^R carrying transducing particles were measured by infecting strain UB-0134 (Youderian et al., 1982) or UB-0020 freshly grown to 2×10^8 cells/ml in LB broth at a multiplicity of infection of 0.5 with phage from the above lysates for 90 min at 37 °C, and the infected cells were plated on selective medium containing tetracycline or chloramphenicol to determine the number of drug resistant transduced colonies. Transduction frequencies are shown as transduced colonies/plaque-forming phage particle ($\times 10^{-6}$) (Schmieger, 1972).

Southern analysis

Probes were labeled with ^{32}P and Southern analysis was carried out as previously described (Casjens and Huang, 1993; Casjens et al., 1995).

Acknowledgments

This work was supported by NIH Grant RO1 AI074825 to SRC. We thank Gino Cingolani for access to the P22 TerS protein structure before its publication, Miriam Susskind, John Roth, Eric Kofoid and Anthony Poteete for *Salmonella* strains, Renato Morona for *Shigella* strains, Surabhi Kasera for help with hybrid phage construction and Liang Tang for plasmid pET28a-gp1.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.02.023.

References

Al-Zahrani, A.S., Kondabagil, K., Gao, S., Kelly, N., Ghosh-Kumar, M., Rao, V.B., 2009. The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. J. Biol. Chem. 284, 24490–24500.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Black, L.W., Showe, M., Steven, A.C., 1994. Morphogenesis of the T4 head. In: Karam, J. (Ed.), Molecular Biology of Bacteriophage T4. ASM Press, Washington, DC, pp. 218–258.

Buttner, C.R., Chechik, M., Ortiz-Lombardia, M., Smits, C., Ebong, I.O., Chechik, V., Jeschke, G., Dykeman, E., Benini, S., Robinson, C.V., Alonso, J.C., Antson, A.A., 2012. Structural basis for DNA recognition and loading into a viral packaging motor. Proc. Natl. Acad. Sci. USA 109, 811–816.

Camacho, A., Jimenez, F., De La Torre, J., Carrascosa, J.L., Mellado, R.P., Vasquez, C., Vinuela, E., Salas, M., 1977. Assembly of *Bacillus subtilis* phage ϕ29. 1. Mutants in the cistrons coding for the structural proteins. Eur. J. Biochem **73**, 39–55.

Casjens, S., 1997. Principles of virion structure, function and assembly. In: Chiu, W., Burnett, R., Garcea, R. (Eds.), Structural Biology of Viruses. Oxford University Press, Oxford, pp. 3–37.

Casjens, S., 2003. Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. 49, 277–300.

Casjens, S., 2008. Diversity among the tailed-bacteriophages that infect the *Enterobacteriaceae*. Res. Microbiol. 159, 340–348.

Casjens, S., Delange, M., Ley, H.L., Rosa, P., Huang, W.M., 1995. Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order. J. Bacteriol. 177, 2769–2780.

Casjens, S., Eppler, K., Parr, R., Poteete, A.R., 1989. Nucleotide sequence of the bacteriophage P22 gene 19 to 3 region: identification of a new gene required for lysis. Virology 171, 588–598.

Casjens, S., Huang, W.M., 1982. Initiation of sequential packaging of bacteriophage P22 DNA. J. Mol. Biol. 157, 287–298.

Casjens, S., Huang, W.M., 1993. Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. Mol. Microbiol. 8, 967–980.

Casjens, S., King, J., 1974. P22 morphogenesis. I: catalytic scaffolding protein in capsid assembly. J. Supramol. Struct 2, 202–224.

Casjens, S., Sampson, L., Randall, S., Eppler, K., Wu, H., Petri, J.B., Schmieger, H., 1992a. Molecular genetic analysis of bacteriophage P22 gene 3 product, a protein involved in the initiation of headful DNA packaging. J. Mol. Biol. 227, 1086–1099.

Casjens, S., Winn-Stapley, D., Gilcrease, E., Moreno, R., Kühlewein, C., Chua, J.E., Manning, P.A., Inwood, W., Clark, A.J., 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. J. Mol. Biol. 339, 379–394.

Casjens, S., Wyckoff, E., Hayden, M., Sampson, L., Eppler, K., Randall, S., Moreno, E., Serwer, P., 1992b. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J. Mol. Biol. 224, 1055–1074.

Casjens, S.R., 2011. The DNA-packaging nanomotor of tailed bacteriophages. Nat. Rev. Microbiol. 9, 647–657.

Casjens, S.R., Gilcrease, E.B., Winn-Stapley, D.A., Schicklmaier, P., Schmieger, H., Pedulla, M.L., Ford, M.E., Houtz, J.M., Hatfull, G.F., Hendrix, R.W., 2005. The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. J. Bacteriol. 187, 1091–1104.

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.

Chai, S., Kruft, V., Alonso, J.C., 1994. Analysis of the *Bacillus subtilis* bacteriophages SPP1 and SF6 gene 1 product: a protein involved in the initiation of headful packaging. Virology 202, 930–939.

Chai, S., Lurz, R., Alonso, J.C., 1995. The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. J. Mol. Biol. 252, 386–398.

Cortines, J.R., Weigele, P.R., Gilcrease, E.B., Casjens, S.R., Teschke, C.M., 2011. Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. Virology 421, 1–11.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97, 6640–6645.

de Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M., Catalano, C.E., 2002. Insights into specific DNA recognition during the assembly of a viral genome packaging machine. Mol. Cell 9, 981–991.

Deichelbohrer, I., Messer, W., Trautner, T.A., 1982. Genome of *Bacillus subtilis* Bacteriophage SPP1: Structure and Nucleotide Sequence of *pac*, the Origin of DNA Packaging. J. Virol. 42, 83–90.

Duffy, C., Feiss, M., 2002. The large subunit of bacteriophage lambda's terminase plays a role in DNA translocation and packaging termination. J. Mol. Biol 316, 547–561.

Feiss, M., Rao, V.B., 2012. The bacteriophage DNA packaging machine. Adv Exp. Med. Biol. 726, 489–509.

Frackman, S., Siegele, D.A., Feiss, M., 1984. A functional domain of bacteriophage lambda terminase for prohead binding. J. Mol. Biol. 180, 283–300.

Frackman, S., Siegele, D.A., Feiss, M., 1985. The terminase of bacteriophage lambda. Functional domains for *cosB* binding and multimer assembly. J. Mol. Biol. 183, 225–238.

Gao, S., Rao, V., 2011. Specificity of interactions among the DNA-packaging machine components of T4-related bacteriophages. J. Biol. Chem. 286, 3944–3956.

German, G., Misra, R., Kropinski, A., 2006. The T1-like bacteriophages. In: Calendar, R. (Ed.), The Bacteriophages, Second Edition Oxford University Press, Oxford, pp. 211–224.

Groenen, M.A., van de Putte, P., 1985. Mapping of a site for packaging of bacteriophage Mu DNA. Virology 144, 520–522.

Harel, J., Duplessis, L., Kahn, J.S., DuBow, M.S., 1990. The *cis*-acting DNA sequences required *in vivo* for bacteriophage Mu helper-mediated transposition and packaging. Arch. Microbiol. 154, 67–72.

Hatfull, G.F., 2012. The secret lives of mycobacteriophages. Adv. Virus Res. 82, 179–288.

Hegde, S., Padilla-Sanchez, V., Draper, B., Rao, V.B., 2012. Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. J. Virol 86, 4046–4057.

Jackson, E.N., Jackson, D.A., Deans, R.J., 1978. EcoRI analysis of bacteriophage P22 DNA packaging. J. Mol. Biol. 118, 365–388.

Jackson, E.N., Laski, F., Andres, C., 1982. Bacteriophage P22 mutants that alter the specificity of DNA packaging. J. Mol. Biol. 154, 551–563.

Jiang, W., Li, Z., Zhang, Z., Baker, M.L., Prevelige Jr., P.E., Chiu, W., 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. Nat. Struct. Biol 10, 131–135.

Karlinsey, J.E., 2007. Lambda-Red genetic engineering in *Salmonella enterica* serovar Typhimurium. Methods Enzymol. 421, 199–209.

Kondabagil, K.R., Zhang, Z., Rao, V.B., 2006. The DNA translocating ATPase of bacteriophage T4 packaging motor. J. Mol. Biol. 363, 786–799.

Kulesus, R.R., Diaz-Perez, K., Slechta, E.S., Eto, D.S., Mulvey, M.A., 2008. Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. Infect. Immun. 76, 3019–3026.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Lebedev, A.A., Krause, M.H., Isidro, A.L., Vagin, A.A., Orlova, E.V., Turner, J., Dodson, E.J., Tavares, P., Antson, A.A., 2007. Structural framework for DNA translocation via the viral portal protein. EMBO J. 26, 1984–1994.

Lee, Y.D., Park, J.H., Chang, H.I., 2011. Genomic sequence analysis of virulent *Cronobacter sakazakii* bacteriophage ES2. Arch. Virol. 156, 2105–2108.

Liebeschuetz, J., Ritchie, D.A., 1986. Phage T1-mediated transduction of a plasmid containing the T1 *pac* site. J. Mol. Biol. 192, 681–692.

Lin, H., Rao, V.B., Black, L.W., 1999. Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. J. Mol. Biol. 289, 249–260.

Linderoth, N.A., Ziermann, R., Haggard-Ljungquist, E., Christie, G.E., Calendar, R., 1991. Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2. Nucleic Acids Res. 19, 7207–7214.

Lobocka, M.B., Rose, D.J., Plunkett 3rd, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M.B., Blattner, F.R., 2004. Genome of bacteriophage P1. J. Bacteriol. 186, 7032–7068.

Maluf, N., Yang, Q., Catalano, C., 2005. Self-association properties of the bacteriophage lambda terminase holoenzyme: implications for the DNA packaging motor. J. Mol. Biol. 347, 523–542.

Morgan, G.J., Hatfull, G.F., Casjens, S., Hendrix, R.W., 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus, Neisseria* and *Deinococcus*. J. Mol. Biol. 317, 337–359.

Morita, M., Tasaka, M., Fujisawa, H., 1995. Structural and functional domains of the large subunit of the bacteriophage T3 DNA packaging enzyme: importance of the C-terminal region in prohead binding. J. Mol. Biol 245, 635–644.

Nemecek, D., Gilcrease, E.B., Kang, S., Prevelige Jr., P.E., Casjens, S., Thomas Jr., G.J., 2007. Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. J. Mol. Biol. 374, 817–836.

Nemecek, D., Lander, G.C., Johnson, J.E., Casjens, S.R., Thomas Jr., G.J., 2008. Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. J. Mol. Biol. 383, 494–501.

Olia, A.S., Prevelige Jr., P.E., Johnson, J.E., Cingolani, G., 2011. Three-dimensional structure of a viral genome-delivery portal vertex. Nat. Struct. Mol. Biol. 18, 597–603.

Oliveira, L., Alonso, J.C., Tavares, P., 2005. A defined *in vitro* system for DNA packaging by the bacteriophage SPP1: insights into the headful packaging mechanism. J. Mol. Biol 353, 529–539.

Oliveira, L., Henriques, A.O., Tavares, P., 2006. Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency. J. Biol. Chem. 281, 21914–21923.

Padilla-Meier, G.P., Gilcrease, E.B., Weigele, P.R., Cortines, J.R., Siegel, M., Leavitt, J.C., Teschke, C.M., Casjens, S.R., 2012. Unraveling the role of the C-terminal helix turn helix of the coat-binding domain of bacteriophage P22 scaffolding protein. J. Biol. Chem. 287, 33766–33780.

Parent, K.N., Gilcrease, E.B., Casjens, S.R., Baker, T.S., 2012. Structural evolution of the P22-like phages: comparison of Sf6 and P22 procapsid and virion architectures. Virology 427, 177–188.

Parent, K.N., Khayat, R., Tu, L.H., Suhanovsky, M.M., Cortines, J.R., Teschke, C.M., Johnson, J.E., Baker, T.S., 2010. P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. Structure 18, 390–401.

Pimkin, M., Pimkina, J., Markham, G.D., 2009. A regulatory role of the Bateman domain of IMP dehydrogenase in adenylate nucleotide biosynthesis. J. Biol. Chem. 284, 7960–7969.

Poteete, A.R., Botstein, D., 1979. Purification and properties of proteins essential to DNA encapsulation by phage P22. Virology 95, 565–573.

Ray, K., Oram, M., Ma, J., Black, L.W., 2009. Portal control of viral prohead expansion and DNA packaging. Virology 391, 44–50.

Roberts, M.D., Martin, N.L., Kropinski, A.M., 2004. The genome and proteome of coliphage T1. Virology 318, 245–266.

Rose, R.E., 1988. The nucleotide sequence of pACYC177. Nucleic Acids Res. 16, 356.

Roy, A., Bhardwaj, A., Datta, P., Lander, G., Cingolani, G., 2012. Small terminase couples viral DNA-binding to genome-packaging ATPase activity. Structure 20, 1403–1413.

Roy, A., Cingolani, G., 2012. Structure of P22 headful packaging nuclease. J. Biol. Chem. 287, 28196–28205.

Schmieger, H., 1972. Phage P22-mutants with increased or decreased transduction abilities. Mol. Gen. Genet. 119, 75–88.

Schmieger, H., 1984. *pac* sites are indispensable for *in vivo* packaging of DNA by phage P22. Mol. Gen. Genet. 195, 252–255.

Schmieger, H., Koch, E., 1987. *In vitro* assay of packaging protein gp3 of *Salmonella* phage P22. Intervirology 28, 157–162.

Shinder, G., Gold, M., 1988. The Nul subunit of bacteriophage lambda terminase binds to specific sites in *cos* DNA. J. Virol. 62, 387–392.

Shultz, J., Silhavy, T.J., Berman, M.L., Fiil, N., Emr, S.D., 1982. A previously unidentified gene in the *spc* operon of *Escherichia coli* K12 specifies a component of the protein export machinery. Cell 31, 227–235.

Simpson, A.A., Tao, Y., Leiman, P.G., Badasso, M.O., He, Y., Jardine, P.J., Olson, N.H., Morais, M.C., Grimes, S., Anderson, D.L., Baker, T.S., Rossmann, M.G., 2000. Structure of the bacteriophage φ29 DNA packaging motor. Nature 408, 745–750.

Sippy, J., Feiss, M., 1992. Analysis of a mutation affecting the specificity domain for prohead binding of the bacteriophage lambda terminase. J. Bacteriol 174, 850–856.

Sternberg, N., Coulby, J., 1987. Recognition and cleavage of the bacteriophage P1 packaging site (*pac*). II. Functional limits of *pac* and location of *pac* cleavage termini. J. Mol. Biol 194, 469–479.

Summer, E., Berry, J., Tran, T., Niu, L., Struck, D., Young, R., 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. J. Mol. Biol. 373, 1098–1112.

Sun, S., Gao, S., Kondabagil, K., Xiang, Y., Rossmann, M.G., Rao, V.B., 2012. Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. Proc. Natl. Acad. Sci. USA 109, 817–822.

Tang, J., Lander, G.C., Olia, A.S., Li, R., Casjens, S., Prevelige Jr., P., Cingolani, G., Baker, T.S., Johnson, J.E., 2011. Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in P22. Structure 19, 496–502.

Tavares, P., Santos, M.A., Lurz, R., Morelli, G., de Lencastre, H., Trautner, T.A., 1992. Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. J. Mol. Biol. 225, 81–92.

Tsay, J.M., Sippy, J., DelToro, D., Andrews, B.T., Draper, B., Rao, V., Catalano, C.E., Feiss, M., Smith, D.E., 2010. Mutations altering a structurally conserved loop-helix-loop region of a viral packaging motor change DNA translocation velocity and processivity. J. Biol. Chem. 285, 24282–24289.

Tsay, J.M., Sippy, J., Feiss, M., Smith, D.E., 2009. The Q motif of a viral packaging motor governs its force generation and communicates ATP recognition to DNA interaction. Proc. Natl. Acad. Sci. USA 106, 14355–14360.

Tye, B.K., Huberman, J.A., Botstein, D., 1974. Non-random circular permutation of phage P22 DNA. J. Mol. Biol. 85, 501–528.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., Copeland, N.G., 2005. Simple and highly efficient BAC recombineering using *galK* selection. Nucleic Acids Res. 33, e36.

Weaver, S., Levine, M., 1978. Replication *in situ* and DNA encapsulation following induction of an excision-defective lysogen of *Salmonella bacteriophage* P22. J. Mol. Biol 118, 389–411.

Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., Johnson, J.E., 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289, 2129–2133.

Winston, F., Botstein, D., Miller, J.H., 1979. Characterization of *amber* and *ochre* suppressors in *Salmonella typhimurium*. J. Bacteriol. 137, 433–439.

Wu, C.H., Black, L.W., 1995. Mutational analysis of the sequence-specific recombination box for amplification of gene 17 of bacteriophage T4. J. Mol. Biol 247, 604–617.

Wu, C.H., Lin, H., Black, L.W., 1995. Bacteriophage T4 gene 17 amplification mutants: evidence for initiation by the T4 terminase subunit gp16. J. Mol. Biol. 247, 523–528.

Wu, H., Sampson, L., Parr, R., Casjens, S., 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol. Microbiol. 45, 1631–1646.

Yang, Q., Berton, N., Manning, M.C., Catalano, C.E., 1999a. Domain structure of gpNu1, a phage lambda DNA packaging protein. Biochemistry 38, 14238–14247.

Yang, Q., Berton, N., Manning, M.C., Catalano, C.E., 1999b. Domain structure of gpNu1, a phage lambda DNA packaging protein. Biochemistry 38, 14238–14247.

Yanisch-Perron, C., Vieira, J., Messing, J., 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103–119.

Yeo, A., Feiss, M., 1995a. Mutational analysis of the prohead binding domain of the large subunit of terminase, the bacteriophage lambda DNA packaging enzyme. J. Mol. Biol. 245, 126–140.

Yeo, A., Feiss, M., 1995b. Specific interaction of terminase, the DNA packaging enzyme of bacteriophage lambda, with the portal protein of the prohead. J. Mol. Biol. 245, 141–150.

Youderian, P., Chadwick, S.J., Susskind, M.M., 1982. Autogenous regulation by the bacteriophage P22 *arc* gene product. J. Mol. Biol. 154, 449–464.

Youderian, P., Sugiono, P., Brewer, K.L., Higgins, N.P., Elliott, T., 1988. Packaging specific segments of the *Salmonella* chromosome with locked-in Mud-P22 prophages. Genetics 118, 581–592.

Youderian, P., Vershon, A., Bouvier, S., Sauer, R.T., Susskind, M.M., 1983. Changing the DNA-binding specificity of a repressor. Cell 35, 777–783.

Zhang, Z., Kottadiel, V.I., Vafabakhsh, R., Dai, L., Chemla, Y.R., Ha, T., Rao, V.B., 2011. A promiscuous DNA packaging machine from bacteriophage T4. PLoS Biol. 9, e1000592.

Zhao, H., Finch, C.J., Sequeira, R.D., Johnson, B.A., Johnson, J.E., Casjens, S.R., Tang, L., 2010. Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. Proc. Natl. Acad. Sci. USA 107, 1971–1976.

Zhao, H., Kamau, Y.N., Christensen, T.E., Tang, L., 2012. Structural and functional studies of the phage Sf6 terminase small subunit reveal a DNA-spooling device facilitated by structural plasticity. J. Mol. Biol 423, 413–426.

Ziermann, R., Calendar, R., 1990. Characterization of the *cos* sites of bacteriophages P2 and P4. Gene 96, 9–15.

Table S1

Non-P22-like phages and prophages in figure 1

Prophage[1] or Phage	Host strain	Locus_tag of *terS* gene	Phylum; Class of host
B171-1	*Escherichia coli* B171	EcB171_0867	Proteobacteria; Gammaproteobacteria
Bact1	*Bacteroides sp.* D2	BSGG_4618	Bacteroidetes; Bacteroidia
Bpert1	*Bordetella pertussis* Tohamal	BP3385	Proteobacteria; Betaproteobacteria
Bruc1	*Brucella sp.* NF2653	BROD_2255	Proteobacteria; Alphaproteobacteria
Crono1	*Cronobacter sakazakii* ATCCBAA-894	ESA_02744	Proteobacteria; Gammaproteobacteria
Delf1	*Delftia acidovora*ns SPH-1	Daci_1941	Proteobacteria; Betaproteobacteria
Dick1	*Dickeya zeae* Ech1591	Dd1591_4010	Proteobacteria; Gammaproteobacteria
Ent13047-1	*Enterobacter cloacae* ATCC13047	ECL_03566	Proteobacteria; Gammaproteobacteria
Ent9394-1	*Enterobacter cloacae* NCTC9394	ENC_27700	Proteobacteria; Gammaproteobacteria
ES-2*	*Cronobacter sakazakii*	teminase small subunit	Proteobacteria; Gammaproteobacteria
Kleb1	*Klebsiella pneumoniae* ATCC13884	HMPREF0484_3793	Proteobacteria; Gammaproteobacteria
Niss1	*Neisseria gonorrhoeae* FA1090	NGO0494	Proteobacteria; Betaproteobacteria
Rhodo1	*Rhodopseudomonas palustris* TIE-1	Rpa1_3055	Proteobacteria; Alphaproteobacteria
Robig1	*Robiginitalea biformata* HTCC2501	RB2501_01380	Bacteroidetes; Flavobacteria
T1*	*Escherichia coli*	T1p54	Proteobacteria; Gammaproteobacteria
Tcarb1	*Thermosinus carboxydivorans* Nor1	TcarDRAFT_1282	Firmicutes; Negativicutes
Yers1	*Yersinia pseudotuberculosis* YPIII	YPK_1220	Proteobacteria; Gammaproteobacteria
Zymob1	*Zymomonas mobilis* ZM4	ZMO0378	Proteobacteria; Alphaproteobacteria

* Fully functional phages (those without asterisks are putative prophages in bacterial genome sequences)

1. These are prophages that we have identified in bacterial genome sequenced in the extant GenBank database. Prophages are not generally named by the researchers who sequenced these bacterial genomes. These are provisional names we gave to these prophages during our studies (some prophage names are from our previous publication Casjens& Thuman-Commike (2011) Virology 411, 393-415).

Table S2

P22-like TerS proteins used in figures 1 and 2

Prophage[1] or Phage	Host species and strain	Genbank Locus_tag or *terS* gene name	Class; Family of host
Ars1	*Arsophonas nasoniae*	ARN_26050	*Gammaproteobacteria; Enterobacteriaceae*
APSE-1	*Hamiltonella defensa*	gene *17*	*Gammaproteobacteria; Enterobacteriaceae*
Cart1	*Pectobacterium carotovorum* PBR1692	PcarbP_010200010662	Gammaproteobacteria; Enterobacteriaceae
CUS-3*	*Escherichia coli* serotype K1	gene *22*	*Gammaproteobacteria; Enterobacteriaceae*
H591-1	*Escherichia coli* H591	ECPG_00988	*Gammaproteobacteria; Enterobacteriaceae*
HK620	*Escherichia coli* serotype H	gene *hkbO*	*Gammaproteobacteria; Enterobacteriaceae*
IME10*	*Enterobacteria* (host species not published)	gene *3*	*Gammaproteobacteria; Enterobacteriaceae*
Miss1	*Salmonella enterica* serotype Mississippi A4-633	LTSEMIS_1103	*Gammaproteobacteria; Enterobacteriaceae*
ϕSG1	*Sodalis glossinidius*	gene *01*	*Gammaproteobacteria; Enterobacteriaceae*
P22*	*Salmonella enterica* serotype Typhimurium	gene *3*	*Gammaproteobacteria; Enterobacteriaceae*
Rett1	*Providencia rettgeri* DSM 1131	PROVRETT_06006	Gammaproteobacteria; Enterobacteriaceae
Serr1	*Serratia plymuthica* AS9	SerAS9_2692	*Gammaproteobacteria; Enterobacteriaceae*
Sf6*	*Shigella flexneri* serotype Y	gene *1*	*Gammaproteobacteria; Enterobacteriaceae*
SPC-P1*	*Salmonella enterica* serotype Paratyphi B	gene *1*	*Gammaproteobacteria; Enterobacteriaceae*
TA124-1	*Escherichia coli* TA124	ESRG_00795	*Gammaproteobacteria; Enterobacteriaceae*
Ugan1	*Salmonella enterica* serotype Uganda R8-3404	LTSEUGA_0536	*Gammaproteobacteria; Enterobacteriaceae*
Wand1	*Salmonella enterica* serotype Wandsworth A4-580	LTSEWAN_0986	*Gammaproteobacteria; Enterobacteriaceae*

* Fully functional phages (those without asterisks are putative prophages in bacterial genome sequences)

1. These are prophages that we have identified in bacterial genome sequenced in the extant GenBank database. Prophages are not generally named by the researchers who sequenced these bacterial genomes. These are provisional names we gave to these prophages during our studies (some prophage names are from our previous publication Casjens& Thuman-Commike (2011) Virology 411, 393-415).

Table S3

Oligonucleotides used in this study

A
5'-GAGGGCGCGAAAGAGTCCTCAGCCTGAAATAACAACTAAGTGAGATGAATCCTGTTGACAATTAATCATCCGCA

B
5'-GATACGAGAGCGTCGCTTATCGCGATCTCCCTTATCAGGTGTCACGTCTTTCAGCACTGTCCTGCTCCTT

C
5'-GAGGGCGCGAAAGAGTCCTCAGCCTGAAATAACAACTAAGTGAGATGAATATGGCGACTGAACCAAAAGC

D
5'-GATACGAGAGCGTCGCTTATCGCGATCTCCCTTATCAGGTGTCACGTCCTTGTCGCCATACTTACGGG

E
5'-ACGATTCTAGAGACTTACCAAGCTGGTTACC

F
5'-CAATCAAGCTTCTGTTCACTACCGGAGCATG

G
5'-TCATTGAGGATTATATTAAAGAACATTTTACAATTAAATAACACCAAACACCCCCCAAAACC

H
5'-CGAAGAGCTATGGGAAATTTTTGTTGAAGAACAAATATAACACACAACCACACCACACCAC

I
5'-ATTAAGAGTAATAACTCCGTCTGTCGGCATGGTACACCGGTTAAGACCCACTTTCACATT

J
5'-AATGCTGAATTCAGACGCAGAATTTTGTCTGCTTACCGTCCTAAGCACTTGTCTCCTG

K
5'-CGGTCAGCGACATCCATTTTCGCGAATCCGGCCGCACTGGCCCCCGATGGTCAACCGTACCG

L
5'-CGGTCAGCGACATCCATTTTCGCGAATCCGGAGTATAAATATCGCCTGAGAGAAGATTTATCTGAAGTCGTTACGC
GAG

M
5'-ACCGGTACGGTTGACCATCGGGGGCCAGTGCGGGCGTTTCATCTCGCGTAACGACTTCAGATAAATCTTCTCTCAGGC
GATA

CHAPTER 3

THE TIP OF THE TAIL NEEDLE AFFECTS THE RATE OF DNA DELIVERY BY BACTERIOPHAGE P22

Reprinted from PLoS ONE

Leavitt JC, Gogokhia L, Gilcrease EB, Bhardwaj A, Cingolani G, et al. (2013)

The Tip of the Tail Needle Affects the Rate of DNA Delivery by Bacteriophage P22.

PLoS ONE 8(8): e70936. doi:10.1371/journal.pone.0070936

OPEN ACCESS Freely available online

PLOS ONE

The Tip of the Tail Needle Affects the Rate of DNA Delivery by Bacteriophage P22

Justin C. Leavitt[1], Lasha Gogokhia[2], Eddie B. Gilcrease[2], Anshul Bhardwaj[3], Gino Cingolani[3], Sherwood R. Casjens[1,2]*

1 Biology Department, University of Utah, Salt Lake City, Utah, United States of America, **2** Division of Microbiology and Immunology, Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, **3** Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, Pennsylvania, United States of America

Abstract

The P22-like bacteriophages have short tails. Their virions bind to their polysaccharide receptors through six trimeric tailspike proteins that surround the tail tip. These short tails also have a trimeric needle protein that extends beyond the tailspikes from the center of the tail tip, in a position that suggests that it should make first contact with the host's outer membrane during the infection process. The base of the needle serves as a plug that keeps the DNA in the virion, but role of the needle during adsorption and DNA injection is not well understood. Among the P22-like phages are needle types with two completely different C-terminal distal tip domains. In the phage Sf6-type needle, unlike the other P22-type needle, the distal tip folds into a "knob" with a TNF-like fold, similar to the fiber knobs of bacteriophage PRD1 and Adenovirus. The phage HS1 knob is very similar to that of Sf6, and we report here its crystal structure which, like the Sf6 knob, contains three bound L-glutamate molecules. A chimeric P22 phage with a tail needle that contains the HS1 terminal knob efficiently infects the P22 host, *Salmonella enterica*, suggesting the knob does not confer host specificity. Likewise, mutations that should abrogate the binding of L-glutamate to the needle do not appear to affect virion function, but several different other genetic changes to the tip of the needle slow down potassium release from the host during infection. These findings suggest that the needle plays a role in phage P22 DNA delivery by controlling the kinetics of DNA ejection into the host.

Citation: Leavitt JC, Gogokhia L, Gilcrease EB, Bhardwaj A, Cingolani G, et al. (2013) The Tip of the Tail Needle Affects the Rate of DNA Delivery by Bacteriophage P22. PLoS ONE 8(8): e70936. doi:10.1371/journal.pone.0070936

Editor: Yaakov Koby Levy, Weizmann Institute of Science, Israel

Received April 25, 2013; **Accepted** June 25, 2013; **Published** August 12, 2013

Funding: This work was supported by National Institutes of Health grant RO1 AI074825 to SRC and GM100888 to GC. Research in this publication includes work carried out at the Kimmel Cancer Center X-ray Crystallography and Molecular Interaction Facility, which is supported in part by NCI Cancer Center Support Grant P30 CA56036. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sherwood.casjens@path.utah.edu

Introduction

Tailed bacteriophage virions deliver DNA to susceptible cells after adsorbing to specific receptors on the surface of bacteria. In the Gram negative bacteria these receptors are surface proteins or polysaccharides. The phage virion proteins that bind to these receptors reside at the tip of the tail and usually have fibrous or elongated shapes. Considerable information is known about phage virion proteins that bind various bacterial receptors, but much less is known of the detailed mechanism of DNA release from the virion into the cell. The best understood systems are *Myoviridae* phages such as T4 and P1 with the well-known contraction of their long tails during DNA delivery [1,2], although questions remain about exactly how this process is controlled and carried out, as well as the energetics of the process [3]. The *Myoviridae* virions insert a preassembled tube through the host membranes that functions as a conduit for DNA transit into the cytoplasm [1]. On the other hand, DNA delivery by the *Siphoviridae* (long non-contractile tails) and *Podoviridae* (short tails) is much less well understood [4,5]. The *Podoviridae* in particular have been shown to release some virion protein molecules, called ejection proteins, along with the DNA. These proteins are required for successful DNA injection, but they have no pre-existing structure that could deliver the DNA through the membranes and periplasm [4,6,7]. The short-tailed phage T7 has recently been shown to rearrange its virion proteins during injection to build a structure that could serve as such a conduit, but the details of this structure and exactly how it might function remain mysterious [8,9]. In addition, little is known about the trigger mechanism that signals the virion to release its DNA for any phage.

Podoviridae in the P22-like group bind their O-antigen polysaccharide primary receptors through a virion protein called the tailspike [10,11]; six trimers of the tailspike polypeptide protrude from the sides of the base of the short tail [12–15]. In addition to giving these phages specificity for initial target cell recognition (*e.g.*, *Salmonella enterica* serotype Typhimurium specificity for phage P22), their tailspike proteins have a catalytic activity that hydrolyzes the polysaccharide receptor backbone. This cleavage may be important in bringing the virion close to the surface, but simple polysaccharide binding cannot be sufficient for rapid and spatially controlled DNA release from the virion [4]. Purified *S. enterica* serotype Typhimurium lipopolysaccharide (LPS, which contains the O-antigen polysaccharide) causes a rather slow P22 DNA release, suggesting that other factors could play a role in this process [16]. The short tails of the virions of P22 and its relatives are assembled from only four phage-encoded proteins. Twelve and six molecules of gp4 and gp10 proteins, respectively, form the

body of its stubby tail. The six trimeric tailspikes (above) are bound to the lower sides of the tail, and a single trimer of the product of P22 gene *26* (gp26) forms a long "needle" that extends outward from the center of the base of the tail [12,17–21]. This needle extends well beyond the outer radius of the tailspikes, suggesting that it should make first contact with the surface of the outer membrane during adsorption. In addition, a few copies of ejection proteins gp7, gp16 and gp20 are released from the virion during injection [13,22–24], but their detailed roles remain unknown.

The N-terminal, virion-proximal end of the gp26 needle forms the exit channel plug that traps packaged DNA in the virion. In its absence DNA is packaged normally, but the particles without plugs are very unstable and DNA falls back out of the capsid, a process that starts even before cell lysis [19,25]. In addition, gp26 is missing from wild type virions that have injected their DNA into *Salmonella* cells [24]. Thus an attractive model has been that gp26 serves as the sensor that determines when DNA should be released from the virion and then triggers its release. We have previously reported the structures of the tail needle proteins from phage P22 and its close relative phage Sf6. The trimeric needles of various P22-like phages range from about 220–320 Å long and have shafts that are only 20–30 Å wide. The long shaft domains have a three-strand coiled-coil structure with 11–16 repeats of heptad amino acid sequences, and the N-terminal 27 residues fold back on the surface of the coiled-coil in the P22 structure to form what may be the site through which the needle binds in the tail channel to form the plug. Although they are homologous in their N-terminal virion-binding domain, P22 and Sf6 have completely different domains at their C-termini; the biological reason for this is not known [26,27,28]. We report here the X-ray structure of the C-terminal domain of the phage HS1 needle (HS1 is an *E. coli* P22-like phage with an Sf6-like needle) which confirms the very unusual feature that this protein domain binds three L-glutamate molecules. In addition, experiments are presented that examine the function of the C-terminal needle domain.

Results and Discussion

Recent Evolutionary Arrival of the Sf6 Type Tail Needle Knob Domain into the P22-like Phage Group

There are currently 164 genome sequences available for P22-like phages and prophages that infect 12 different species of *Enterobacteriaceae* bacteria (we make the reasonable assumption that prophages arrived at their present locations by the normal infection route). We also note that there are over 100 *E. coli* genomes that carry a small, several gene remnant of P22 that includes a homologue of the Sf6 tail needle gene; these latter genes are not exceptional and were only included in this analysis by the inclusion of the gene from *E. coli* strain EC4113 which carries such a phage genome fragment (locus_tag ECH7EC4113_3989 in defective prophage EC4113-1; our provisional prophage name). Among these 164 phage genomes, 69 of the needle genes encode Sf6 type C-terminal knob domains (*i.e.*, have sequences similar to Sf6 amino acids 140 to 282) and 95 have C-terminal domains that are related to the P22 needle tip. All of these Sf6 type knob domains are very closely related, with the most distantly related knob domains being only 4.9% different in amino acid sequence (see figure S1), and all the P22-like phages with this type of needle domain infect only bacteria in the *Escherichia* or *Shigella* genera. On the other hand, the P22-like C-terminal domains are much more diverse, with the most distant pairs being about 50% different in amino acid sequence (figures 1 and S1). Neighbor-joining tree analysis shows that there are currently eleven major branches (*i.e.*, sequence types) of the P22 type needle tip domain (figure 1). Three of these have many known members and are encoded by phages that infect *Escherichia* or *Salmonella* bacterial genera - subtype A, whose members all infect *E. coli* (typified by phage HK620 [29]), B, whose members infect *E. coli, S. enterica* or *Cronobacter sakazakii* (typified by phage epsilon 34 [30]), and C, whose members all infect *S. enterica* (typified by phage P22); in addition, still more diverse homologs are found in phages that infect other *Enterobacteriaceae* species. A possible simple explanation for lack of Sf6 tip domain diversity is that the P22 type needle tip domain is ancestral in the P22-like phage group, and it has been there long enough to allow considerable divergence to occur, while DNA encoding the Sf6 type needle knob has entered to P22-like phage population quite recently and has not had time for much divergence. This direction of transfer is supported further by the observation that there are no database matches to the P22 type needle tip domain outside the P22-like phages, but there are a number of such matches to the Sf6 type tip domain (see below). This scenario seems quite plausible because the P22-like phages are known to have undergone considerable horizontal transfer of genetic information, both within the P22-like group and with other phage types, in the genes that encode their virion assembly proteins [31].

The evolutionary source of the Sf6 type tail needle knob domain is not known; however, there are currently 21 convincing database matches of the Sf6 knob to proteins that fall outside the P22-like phage knob branch by neighbor-joining analysis (figure S2). These "outside" matches range from about 37% to 63% identical to the knobs of needle proteins present in the P22-like phages, and all are encoded by phages that are outside the P22-like group. One of these phages is an *Aeromonas* prophage, nine are somewhat phage Mu-like *Vibrio* prophages (our unpublished analysis), and the remaining eleven reside in *bona fide* phage genomes. The latter phages are all large virulent T4-like phages that infect *Enterobacteriaceae* species *E. coli*, *S. enterica* and *Dickeya solani* (formerly *Erwinia chrysanthami)* or *Aeromonas salmonicida*. In all of these "outside match" proteins, none of whose function is known, the homology to the Sf6 knob lies at or near the C-terminus of the protein. The phage 31, 44RR2.8 t and RB43 homologues have an N-terminal extension of up to 143 amino acids, but others have only a very short N-terminal extension in addition to the knob domain homology.

Atomic Structure of the HS1 Knob Domain

The Sf6 tail needle knob domain has the very unusual feature that three glutamate molecules are tightly bound in the crevices between the three subunits [26]. To determine if these are unique to Sf6 or are also present in other Sf6-like tail needle knobs, we carried out structural analysis of bacteriophage HS1 tail needle knob. HS1 was identified as an apparently intact P22-like prophage in *E. coli* strain HS [31]. The HS1 tail needle (encoded by locus_tag EcHS_A0316 gene) is characterized by a virion-binding N-terminal domain 98% identical to that of Sf6, and a shaft-forming coiled-coil helical region predicted to be 210 Å long that is only 67% identical in sequence to Sf6. This helical core contains 16 heptad repeats, compared to 11 in the Sf6 and P22 needles. The C-terminus of the HS1 tail needle contains a domain (the knob) that is 99% identical to Sf6 in sequence. We cloned the HS1 prophage needle's C-terminal knob and nine residues of its coiled-coil helical shaft fused to maltose binding protein (MBP) after polymerase chain reaction (PCR) amplification from *E. coli* strain HS DNA. The resulting hybrid protein was affinity purified, and the MBP portion removed by PreScission protease and column chromatography (details in Materials and Methods). The purified HS1 knob was crystallized, and its structure was solved by Molecular Replacement using the phage Sf6 knob domain (pdb

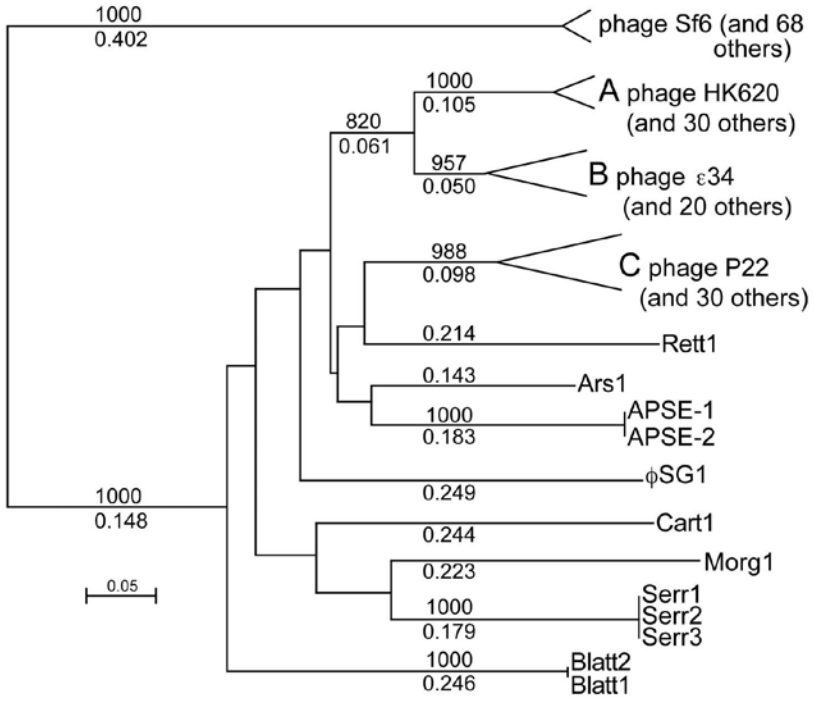


Figure 1. Relationships among tail needle C-terminal domains of the P22-like phages. A neighbor-joining tree (created with Clustal X2 [80]) is shown with selected branch lengths (numbers between 0. and 1) and bootstrap values out of 1000 trials (between 1 and 1000). The nodes far from the branch tips are not well-supported and are not shown. A scale in fractional difference is shown in the lower left. Branches A, B and C have many members and the splits at these branch tips show the regions within which the individual members diverge (the larger tree in figure S1 shows the placement of all the individual sequences). The "phage Sf6" branch is not related to the other branches and its inclusion here is to demonstrate this, and does not imply any phylogenetic relationship with the other branches. The branches not labeled "phage" are from P22-like prophages in the following bacterial genome sequences: Rett1, *Providencia rettgeri* DSM 1131; Ars1, *Arsenophonus nasoniae*; APSE-1/−2, *Hamiltonella defensa*; øSG1, *Sodalis glossinidius*; Cart1, *Pectobacterium carotovorum* PBR1692; Morg1, *Morganella morganii* KT; Serr1/2/3, *Serratia plymuthica* strains AS9, AS12 and AS13; Blatt1/2, *Escherichia blattae* strains DSM 4481 and 105725.
doi:10.1371/journal.pone.0070936.g001

code 3RWN) as search model, and the structure was refined to an $R_{work/free}$ of 15.13%/15.70 at 1.1 Å resolution (Table 1 and Materials and Methods).

The globular HS1 needle knob folds into a homotrimeric TNF-like fold, very similar to that of the Sf6 knob (rmsd 0.137 Å; figure 2). The trimer is built with three identical subunits each displaying a jellyroll topology characterized by seven stacked anti-parallel ß-strands. Like the Sf6 knob [26], the HS1 tail knob structure contains a phosphate ion bound at the distal tip of the knob (figure 2A) and an L-glutamate is present in the electron density at each of the three dimeric interfaces of the homotrimer (figure 2B). The phosphate ion makes polar contacts with HS1 needle residues S266 and N268 near the 3-fold axis at the distal tip of the knob. Alteration of these residues in Sf6 knob did not affect the stability of the protein [26]. The conserved triad, S266-R267-N268, is present in all the known needle knob sequences from P22-like phages; however, although R267 (its side chain points down and is involved in intra-chain stability) is universally present in all the proteins aligned in figure S3, the S and N on either side are not conserved in proteins encoded by the non-P22-like phages, so the bound phosphate is unlikely to be a universal feature of all of these proteins. Although not surprising given the similarity of the two proteins, the presence of the bound glutamate molecules in the HS1 structure (figure 2A) is consistent with their being universally present in the P22-like needle knob structures. Remarkably, in all 21 of the non-P22-like phage knob homologues (above) amino acids that are in intimate contact with the bound L-glutamate molecules are completely conserved (Glu181, K235, S283 and D285 in the HS1 needle protein; figures 2C and S3), so it is very likely that all these proteins have bound L-glutamate. Finally, the Sf6 and HS1 tail needle knob structures have one amino acid difference within the knob domain, HS1 Ala305 (neutral hydrophilic) corresponds to Sf6 Ser270 (weak hydrophobic), and one difference in the common shaft portions whose structures are known, residue HS1 Ser169 (acidic) to Sf6 Asp134 (neutral hydrophilic) (both are indicated in Figure 2D). These differences have no effect on the overall three dimensional structure of the tail needle knob and slightly decease the negative charge of HS1 helical shaft.

Is the C-terminal Knob of the Tail Needle Host Species-specific?

As mentioned above, the Sf6 type needle tip domain has a different fold from the P22 needle tip (figure 3). Since proteins with homology to this domain have been found only in phage that infect *E. coli*, *Escherichia fergusonii* and *S. flexneri* (and it has been argued that at least *E. coli* and *Shigella* are actually one species [32]), it seemed possible that this domain participates in

Table 1. Crystallographic data collection and refinement statistics.

Data collection statistics	
Wavelength (Å)	0.9537
Space group	p212121
Unit cell dimensions (Å)	$a = 56.78$, $b = 87.36$, $c = 88.84$
Angles (°)	$\alpha = \beta = \gamma = 90$
Resolution range (Å)	20-1.1
Wilson *B*-factor (Å^2)	8.94
Total observations	1,053,320
Unique observations	176,986
Completeness[a] (%)	99.1 (92.8)
Redundancy[a]	6.0 (5.3)
R_{sym}[a,b] (%)	7.2 (49.9)
$<I>/<\sigma(I)>$[a]	29.96 (2.4)
Refinement statistics	
Number of reflections (10-1.1 Å)	176,905
R_{work}/R_{free}[c] (%)	15.13/15.70
Number copies in asymm. unit	1
Number of water molecules	754
B value of model (Å^2) chains A/B/C/waters	10.6/10.37/10.91/24.22
r.m.s. deviation from ideal bond length (Å)/angles (°)	0.008/1.3
Ramachandran plot (%)core/allowed/generously allowed/disallowed	92.7/7.3/0.0/0.0

[a]Highest resolution shell is shown in parenthesis.
[b]$R_{sym} = \sum_{i,h}|I(i,h) - <I(h)>|/\sum_{i,h}|I(i,h)|$ where $I(i,h)$ and $<I(h)>$ are the *i*th and mean measurement of intensity of reflection *h*.
[c]The R_{free} value was calculated using 2,000 reflections.
doi:10.1371/journal.pone.0070936.t001

adsorption and/or DNA delivery in a host species-specific manner. The tailspike is known to confer host primary adsorption specificity for the P22-like phages by binding to the O-antigen surface polysaccharide [31], but to test the idea that the needle protein knob might also contribute the ability to infect different hosts, we constructed two hybrid P22 *Salmonella* phages in which the distal C-terminal tip of gp26 is replaced by the Sf6 or HS1 sequence (figure 3; see Methods and Materials). These constructs were made in a P22 prophage since modifications can be made there even if the change is lethal to phage growth, and the functionality of the mutant phage can be tested by analyzing the phenotype after induction of the prophage to lytic growth. The prophage used for these tail needle modifications was P22 *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$*am*H101 (phage strain UC-0911), in which the three mutations allow efficient tailspike gene expression after induction, easy kanamycin selection for lysogens, and control of lysis, respectively, is capable of a normal lytic growth cycle [33]. Modifications of this prophage in *Salmonella* strain UB-1790 were made using recombineering technology (see Materials and Methods; bacterial and phage strains used are listed in Table 2).

The fusion point of the above P22 and Sf6 hybrid tail needle protein was made in the coiled-coil shaft domain in a region of very high sequence similarity between the two genes, so no disruptions of trimerization heptad number or frame were made; the resulting hybrid protein should thus be very likely to fold normally into a functional needle protein. In this hybrid phage, P22 gene *26* codons 69–233 are replaced by phage Sf6 gene *9* codons 69–282, so the C-terminal knob and the C-terminal approximately two-thirds of the coiled-coil shaft have Sf6 sequence. In the HS1 hybrid needle construct, only the C-terminal knob-encoding region (codons 174 to 317 of the phage HS1 gene) neatly replaces the P22 C-terminal domain (codons 141–233) (figure 3). When the *Salmonella* strains carrying these two prophages (strains UB-1918 and UB-2083, respectively) were induced with mitomycin C, a yield of phage particles that make approximately normally sized plaques on indicator strain UB-0002 was produced in both cases that was similar to that of the isogenic strain UB-1790 whose prophage has an all-P22 tail needle (Table 3). Plating for plaques at 30°C or 37°C gave the same results. Thus, the virions of these two hybrid phages are stable and functional under laboratory conditions. We also found that both of these hybrid phages (UC-0918 and UC-0926) infect *S. enterica* serovar Typhimurium normally in liquid culture. Comparison of the proteins present in CsCl gradient purified virions with P22, Sf6 or HS1 needle tips (UC-0911, UC-0918 and UC-0926, respectively) showed identical virion proteins except that, as expected, the needle protein's apparent molecular weight is commensurately larger when the larger Sf6 or HS1 knob domain is present (shown for the HS1 hybrid in figure S4). Since these two hybrid phages infect *Salmonella* apparently normally, we conclude that the Sf6 knob domain function does not confer species specificity to P22 in the laboratory. More formally, this shows that the P22 tip domain is not essential, and its replacement by the Sf6 needle knob does not prevent infection of *Salmonella* by P22 by adding a *Shigella*-specific component to the infection process. However, the observation that in the face of the extensive horizontal exchange among the P22-like phages [31], none of the 48 known P22-like phages that infect *Salmonella* carry the Sf6 type needle tip domain

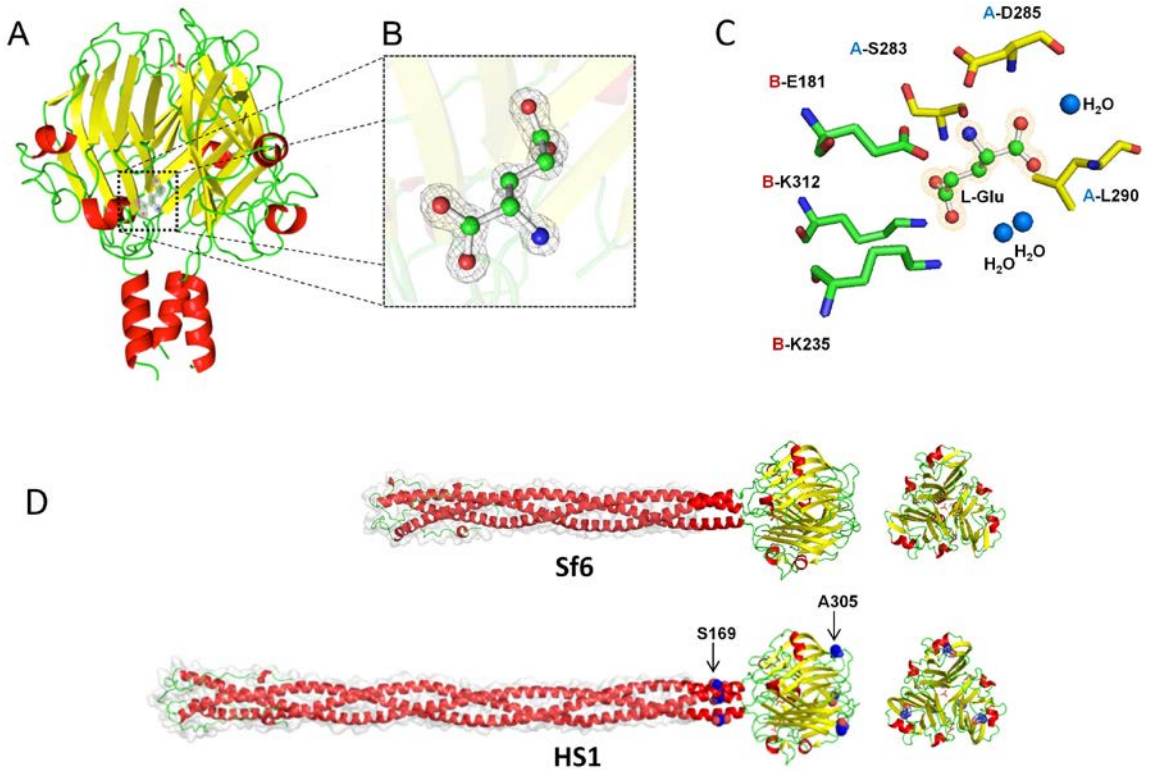


Figure 2. Atomic structure of the phage HS1 tail needle knob. A. Ribbon diagram of bacteriophage HS1 tail needle knob determined crystallographically to 1.1 Å resolution; the N-termini are at the bottom of the diagram. Helices are shown in red, ß-sheets in yellow, and random coil in green; the bound L-glutamate is shown as sticks-and-balls and phosphate is shown as a small red sticks. **B.** Magnified view of L-glutamate trapped at the HS1 needle knob dimeric protomer:protomer interface. L-glutamate (in stick-and-balls) is overlaid to the final 2Fo-Fc electron density map (gray) contoured at 1.5σ above background. **C.** Side chains (sticks) from protomer A (yellow) and protomer B (green) that interact with L-glutamate (stick-and-balls). The indicated HS1 needle amino acids correspond to Sf6 needle amino acids as follows with Sf6 residue numbers in parentheses: Glu181(146), Lys235(200), Ser283(248), Asp285(250), Leu290(255) and Lys312(277). **D.** Structural models of full length Sf6 and HS1 tail needles. The two amino acid differences (from the Sf6 needle) that lie at positions in or near the knob domain, Ser169 and Ala305 of the HS1 tail needle, are shown as blue spheres. The models were obtained by using the Robetta full-chain protein structure prediction server [81]; the N-terminal parts of the needle protein shafts whose structures are modeled from the homologous P22 tail needle have a light gray surface contour behind. In all the panels, α-helices, β-strands and loops are colored in red, yellow and green, respectively.
doi:10.1371/journal.pone.0070936.g002

suggests there may be an evolutionary disadvantage should a *Salmonella* P22-like phage obtain a needle gene with this domain.

Is L-glutamate Binding Essential?

In order to begin to address the biological relevance of the L-glutamate molecules present in the crystal structures of the HS1 and Sf6 needle protein knob domains, recombineering was used to engineer point mutational changes of Sf6 amino acids Glu146, Lys200, Ser248 and Asp250 (corresponding to Glu181, Lys235, Ser283, Asp285 in HS1 knob, respectively; figure 2B) that make close contact with this small molecule (above; see figure 3C in [26]). Changes were made in the Sf6 hybrid needle P22 prophage of bacterial strain UB-1918 (table 2) as follows: Glu146 to Asp or Ala, Lys200 to Glu or Ala, Ser248 to Thr or Ala, Asp250 to Ala, and combinations of the Ala changes (see details in Methods and Materials). Induction of these mutant prophages resulted in approximately normal yields of plaque-forming phage particles in all cases, including the double mutant L-glutamate binding site mutant virions (table 3). It is very likely that these changes, especially in the double mutants, would substantially lower L-glutamate binding. Therefore, the fact that all of these mutational changes do not affect virion function as measured by plaque formation suggests that L-glutamate binding by the hybrid tail needle's distal domain is not essential under our laboratory conditions. Since the experiments here and in the previous section were performed with a hybrid needle protein in the context of P22 infection of *Salmonella*, the possibilities remain that the Sf6 knob could have a specific role in allowing phage Sf6 to infect *Shigella* and that bound L-glutamate could have an important role in a *Shigella* infection.

The C-terminal P22 Tail Needle Domain is not Essential Under Laboratory Conditions

Because the above domain switches and the above amino acid changes did not greatly affect virion function, we wondered whether the tip domain of the gp26 needle is in fact essential for phage P22 virion function. Since removal of the C-terminal domain of the P22 needle lowers the stability of the trimer and

Table 2. Phage and bacteria used in this study.

Bacterial strain[a]	Genotype[b]	Reference
UB-0001	(DB7000) *leuA*$^-$414, *sup*$^\circ$	[62]
UB-0002	(DB7004) *leuA*$^-$414, *supE*	[62]
UB-0020	(MS1868) *leuA*$^-$414, *hsdSB* (r^- m^+) Fels2$^-$, *sup*$^\circ$; from M.Susskind	[75]
UB-0134	*leuA*$^-$*am*414, Fels2$^-$, *cob*$^-$ΔCRR299 (P22 *sieA*$^-$44, *ant*$^-$*am*222, ΔAp68 (tpfr49 *a1*$^-$, *9*$^-$, *c2*$^+$, *mnt*$^+$)); from J. Roth	[76]
UB-1732	*E. coli* HS; from J. Nataro	[77–80]
UB-1737	(TH2788) *fliY*$^-$5221::Tn10dTc; from K. Hughes	
UB-1790	UB-20 *galK*::TetRA-1 (P22 *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	[33]
UB-1807	UB-20 *galK*::TetRA-1 (P22 *26*::Sf6-3::*galK*-1, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1832	UB-20 (P22 *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1918	UB-20 (P22 *26*::Sf6-3, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1919	UB-20 (P22 *26*::Sf6-3 Glu146Asp(GAC), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1920	UB-20 (P22 *26*::Sf6-3 Glu146Ala(GCG), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1921	UB-20 (P22 *26*::Sf6-3 Ser248Thr(ACC), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1922	UB-20 (P22 *26*::Sf6-3 Ser248Ala(GCC), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1924	UB-20 (P22 *26*::Sf6-3 Asp250Ala(GCT), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1925	UB-20 (P22 *26*::Sf6-3 Lys200Ala(GCG), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1926	UB-20 (P22 *26*::Sf6-3 Lys200Glu(GAG), *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1927	UB-20 (P22 *26*::Sf6-3 Glu146Ala/Asp250Ala, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1928	UB-20 (P22 *26*::Sf6-3 Lys200Ala/Asp250Ala, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1929	UB-20 (P22 *26*::Sf6-3 Lys200Ala/Glu146Ala, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1940	UB-20 (P22 *26*::Sf6-3::TetRA-1, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1941	UB-20 *galK*::TetRA-1 (P22 *26*::foldon-1, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1942	UB-20 (P22 *26*::Sf6-3::TetRA-2, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1943	UB-20 (P22 *26*::Sf6-3::TetRA-3, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-1944	UB-20 (P22 *26*::Sf6-3::TetRA-4, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-2078	UB-20 *galK*::TetRA-1 (P22 *26*::galK-2, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-2083	UB-20 *galK*::TetRA-1 (P22 *26*::HS1-1, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101)	
UB-2130	(TH18984) *fliC*5469::MudK, Δ*hin-fljA*8068::PfliC-cat, *rfbF*::TPOP; from K. Hughes	
Phage P22 strains		
UC-0011	P22 *c1*-7, *26*$^+$, *13*$^-$amH101	[12]
UC-0911	P22 *26*$^+$, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	
UC-0918	P22 *26*::Sf6-3, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	
UC-0926	P22 *26*::HS1-1, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	
UC-0927	P22 *26*::foldon-1, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	
UC-0931	P22 *26*::Sf6-3 Glu146Ala/Asp250Ala, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	
UC-0932	P22 *26*::Sf6-3 Lys200Ala/Asp250Ala, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	
UC-0933	P22 *26*::Sf6-3 Lys200Ala/Glu146Ala, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101	

[a]All strains are *S. enterica* serovar Typhimurium LT2 derivatives, except UB-1732 which is *E. coli*.
[b]Strain names in parentheses are the names used in the laboratory from which the strain was obtained. UB-20 in the "Genotype" column indicates derivatives of parental strain UB-0020. Mutant prophage codons are shown in square brackets; multiple mutants have the same codon changes as single mutants.
doi:10.1371/journal.pone.0070936.t002

because a translation stop at codon 54 (the P22 *26* minus *amber* H204 mutation is a C to A change at position 160 in the *26* gene) gives a *26* null phenotype when not suppressed, we deemed it prudent to replace the needle's C-terminal domain with a "trimer nucleation domain" rather than simply C-terminally truncate the protein. The phage T4 fibritin "foldon" domain has been shown to mediate trimerization and folding of fibritin and other coiled-coil trimers when it is present at the C-terminus of these proteins [34–38]. Fibritin is the neck fiber protein encoded by the T4 *wac* gene. Its role in T4 is to aid in tail fiber assembly, and there is no indication that it interacts with target cells in any way. We have previously shown that fusing the foldon sequence to the C-terminus of N-terminal P22 needle protein fragments results in properly folded and very stable trimers [39]. We therefore used recombineering to replace the P22 tail needle tip domains III and IV (P22 amino acids 141–233) [27] with the 25 amino acid foldon as described in Materials and Methods to create *Salmonella* strain UB-1941 (figure 3). Induction of this prophage by mitomycin C released approximately normal numbers of progeny phages (table 3), showing that the C-terminal domains of the P22 needle

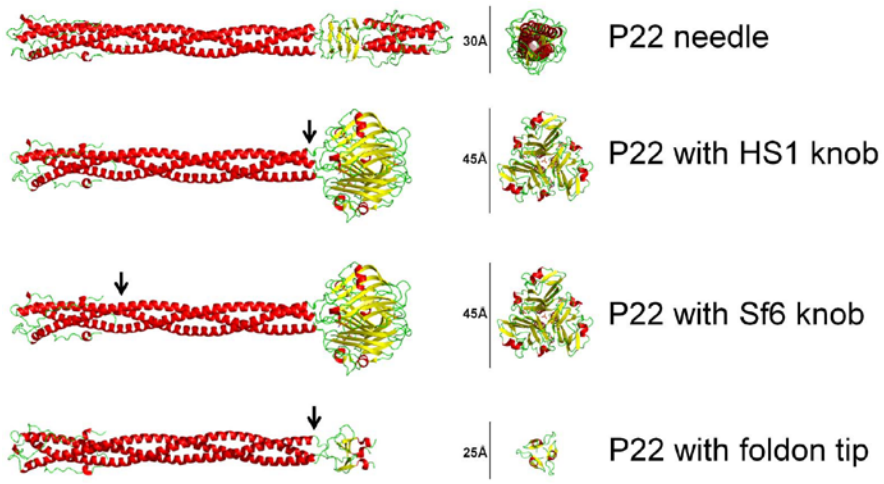


Figure 3. Structural models of the P22 gp26 tail needle and chimeric tail needles. A. Crystal structure of P22 tail needle gp26 (pdb 3C9I). **B–D**. Homology structural models of chimeric P22 needles with Sf6 knob (in phage UC-0911), HS1 knob (in phage UC-0926) and foldon tip (in phage UC-0927). Chimeric models were generated for illustration with align function of PyMol (Version 1.3, Schrodinger, LLC, San Carlos, CA), where C-terminal knob domains of Sf6 and HS1 tail needle (pdb 3RWN, 4K6B), C-terminal foldon domain from fibritin fiber of the bacteriophage T4 (pdb 1AA0) were fused downstream of P22 gp26 tail needle helical core residues 1–140 (pdb 3C9I), respectively. In all three models, arrow indicates point of fusion.
doi:10.1371/journal.pone.0070936.g003

Table 3. Effects of genetic modification the C-terminal needle domain.

Phage source[a]	Lysate titer[b]	Virion infectivity[c]
UB-1790 Parent P22 phage[d]	1.7×10^{10}	1.0
UB-2083 HS1-1 hybrid needle phage	7.5×10^{9}	1.1
UB-1918 Sf6-3 hybrid needle phage	6.2×10^{9}	2.1
UB-1919 Glu146Asp	5.5×10^{9}	0.8
UB-1920 Glu146Ala	5.2×10^{9}	1.7
UB-1925 Lys200Ala	1.4×10^{9}	2.0
UB-1926 Lys200Glu	2.6×10^{9}	1.4
UB-1921 Ser248Thr	6.3×10^{9}	2.0
UB-1922 Ser248Ala	3.8×10^{9}	2.1
UB-1924 Asp250Ala	2.1×10^{9}	2.9
UB-1927 Glu146Ala/Asp250Ala	1.8×10^{10}	1.0
UB-1928 Lys200Ala/Asp250Ala	1.4×10^{9}	1.8
UB-1929 Glu146Ala/Lys200Ala	2.9×10^{9}	1.3
UB-1941 Foldon	8.3×10^{9}	0.13

[a]*Salmonella* lysogens are listed that were induced to prepare stocks of the phages in the table; the amino acid changes in the L-glutamate binding site of the needle are shown after the Sf6-hybrid needle phage strain names (UB-1919 to 1929).
[b]Lysogens were induced with mitomycin C and titered on *Salmonella* strain UB-0002; average results from several replicates are shown.
[c]Phage particles were purified through CsCl step gradients [68], and phage titers were determined on *Salmonella* strain UB-0002. Relative PFU/particle ratios were calculated by normalization to UB-1790 titer (PFU) and to the intensity of quantified coat protein bands in SDS electrophoresis gels and/or OD_{280} (particles). Several replicates were performed for each phage with similar results, and the average value is shown.
[d]The prophages all have the following genetic background: P22 *26*::Sf6-3, *sieA*$^-$Δ1, *15*$^-$ΔSC302::Kan^R, *13*$^-$amH101 (see table 2).
doi:10.1371/journal.pone.0070936.t003

are not absolutely essential in laboratory infections. This phage (UC-0927) does, however, make tiny plaques on indicator strain UB-0002, and its particle/plaque-forming unit (PFU) ratio is about 10-fold lower than the UC-0911 parent with a wild type needle (table 3), suggesting that there is a moderate (but not absolute) defect in these virions under laboratory conditions.

The C-terminal Tail Needle Domain Affects the Rate of DNA Delivery into Cells

Since plaque formation is not a quantitative measure of virion function, DNA might still be delivered into the cell more slowly by the P22 phages with modified tail needles (above) than phages with wild type needles. We therefore sought to measure DNA passage from the virion into the cell on a more nearly real-time scale. Entry of a number of tailed phage DNAs into the host cell cytoplasm during injection is highly correlated with an efflux of K^+ ions out of the cell into the surrounding medium [40–43]. Thus, measurement of K^+ release with an Orion Ionplus potassium electrode (see Materials and Methods) can be used as an approximate surrogate for real-time measurement of DNA entry into the cytoplasm of the cell during most tailed phage infections. P22 had been shown by Ter-Nikogosian *et al.* [44] to exhibit such ion release, and figure 4 confirms that infection by phage P22 UC-0911 causes a rapid release of K^+ ions; however, we did not observe the dependence of K^+ release on the presence of externally supplied Ca^{++} ions reported by those authors (data not shown and N. Cumby, personal communication). We performed the following characterizations of this system to test whether P22-induced K^+ release correlates with its DNA delivery into susceptible cells. The kinetics and extent of K^+ release by P22 is affected by the multiplicity of infection (MOI) (figure 4A), as might be expected from mass action considerations. All subsequent experiments were performed at the more physiologically relevant MOI of 10. P22 mediated K^+ release is more rapid at 37°C than at 30°C (figure 4B), and we used 30°C in subsequent experiments in order

to maximize resolution of any timing differences between wild type and mutant phage infections. As expected, *S. enterica* that lacks P22's O-antigen receptor (UB-02130) shows no K^+ release in response to added phage P22 (figure 4C). In addition, release of K^+ is not affected by the ability or inability of the infecting phage to form a lysogen, since a clear plaque mutant (UC-0011) shows similar K^+ release to phage that can lysogenize (UC-0911) (data not shown). Finally, P22 infection of a host that carries a P22 prophage that is missing its *sieA* and *gtrABC* genes (UB-0134) gives normal K^+ release (figure 4D). This lysogen is defective in superinfection exclusion [45,46] and O-antigen modification [46,47] so DNA is injected normally, but the resident P22 prophage repressor is present in the cell and prevents expression of nearly all of the genes of the infecting P22 genome [45,46]. Thus, expression of the vast majority of P22 genes is not required for K^+ release. These findings are all consistent with the idea that DNA transit into the cytoplasm from adsorbed P22 virions correlates with K^+ release by a mechanism that remains to be elucidated (see for example [48]).

Our measurements indicate that at 30°C DNA K^+ release begins at 7 to 10 min after P22 infection and is complete by 15–20 min (figure 4); strikingly, 60–80% of cellular potassium ions are released during this time period. The lag time is affected by MOI (figure 4A) and temperature (figure 4B), as might be expected if the lag reflects the time it takes for P22 virions to bind the cell surface O-antigen.

polysaccharide and then make their way to the surface of the outer membrane (the latter probably through cleavage of the O-antigen [4]). The period during which K^+ is released should reflect the *maximum* time it takes to transfer the DNA from the virion into the cell, and since K^+ release is a bulk solution measurement and the infections in these experiments are not highly synchronized, DNA entry transit time for individual virions is likely significantly less than this. This transit time value of 8–13 min is sensible in terms of P22's life cycle (reviewed in [46]), which requires that DNA circularize (and therefore have both DNA ends internalized) soon after infection begins [49–52]. The rate of DNA entry into the cell has not been previously measured for P22, but the above *minimum* rate of 55–95 bp/sec at 30°C. Thus the P22 value may not be very different from the approximately 160 bp/sec (about 5 min total transit time) recently measured for individual phage lambda virions at room temperature [53]; lambda is an *E. coli Siphoviridae* phage with a genome of similar size to that of P22 that also must circularize soon after infection begins. We note that DNA cell entry rates are not uniform among tailed phages, and where they have been measured entry rates range from about 3000 bp/sec for phage T4 [42] to 70–250 bp/sec for phage T7 [54–56] and 160 bp/sec for lambda [53]. DNA entry rate appears to be a property that is evolutionarily optimized for each phage's molecular lifestyle.

Parallel infections by P22 with a wild type needle (UC-0911) and the isogenic phage carrying the phage HS1 C-terminal knob domain (the *26*::HS1-1 hybrid needle gene in phage UC-0926) shows significantly delayed K^+ release kinetics for the HS1 hybrid (figure 5A). Phage UC-0926 reproducibly has a slower K^+ "release period" which begins after a lag time that is about the same as wild type; the rate of release ranged from 40–70% of wild type in different experiments. On the other hand, P22 phage with the C-terminal tip domain and most of the needle shaft domain replaced by Sf6 sequences (UC-0918) showed nearly identical kinetics of K^+ release to phage with the fully P22 needle protein (figure 5A). Since there is only one conservative amino acid sequence difference between the HS1 and Sf6 knobs (Ala305 in HS1 is Ser270 in Sf6; figure 2D), the significant difference in K^+ release observed between UC-0918 and UC-0926 is likely due to the presence of the Sf6 shaft residues 69–141 in UC-0918 (which have a number of differences from P22 shaft residues); the mechanism underlying this difference could be due either some shaft function (such as interaction with the bacterial surface), or due to different allosteric coupling between possible conformational changes in the tip with changes that might be propagated through the needle to its N-terminus to effect needle release. We also performed K^+ release measurements with P22 phages whose hybrid Sf6 needle knob domains carry the following alterations: Glu146Asp, Glu146Ala, Lys200Ala, Ser248Thr, Ser248Ala or Asp250Ala, as well as double mutants Glu146Ala/Asp250Ala, Lys200Ala/Asp250Ala and Glu146Ala/Lys200Ala. None of these mutants showed altered K^+ release kinetics relative to the parental phage UC-0926 (figure 4E; data not shown for single mutants and the Glu146Ala/Asp250Ala phage UC-0931). In addition, K^+ release was measured during infection by P22 UC-0926 containing the unaltered HS1 hybrid knob with and without 10 mM L-glutamate present in the cellular resuspension medium (KR buffer), and no significant difference in K^+ release kinetics was observed (figure 4F) (virions were dialyzed for 24 hr at 4°C against several changes of TM buffer with no L-glutamate before use in this experiment). These results make it unlikely that externally added L-glutamate plays an important role in needle function during DNA delivery by these hybrid phages.

P22 virions in which the foldon replaces the C-terminal needle tip domain (UC-0927, above) showed much slower K^+ release kinetics than P22 phages with wild type needles (figure 5A). This is paralleled by this phage's tiny plaques and low PFU/particle ratio (above). Figure 5B shows that even at MOI values as high as 100, K^+ release caused by phage P22 UC-0927 has a longer lag and lower 30 min extent of release than P22 with a wild type needle at an MOI of 10, so its slow K^+ release cannot be explained simply by the low PFU/particle ratio (note that all the K^+ release experiments presented here compare infections in which the MOIs are normalized to physical virion particles rather than PFUs; see table 2 and Materials and Methods).

While K^+ release generally parallels the approximate period of phage DNA entry during the injection process, and in the case of T5 K^+ release has two steps that appear to mirror its two stage DNA injection [42,43], the mechanism of phage mediated K^+ release is not known. It could be the result of opening of the channel for DNA transfer through the cytoplasmic membrane, and leakage through this channel, leakage around the outside of such a channel, or even through some other phage or host protein present in the membrane early in infection could allow K^+ escape. Thus, although for ease of discussion above, we equated DNA entry and K^+ release, we recognize that at present K^+ release and DNA entry are only correlated observations, and they may not be mechanistically coupled.

Concluding Remarks

In this work, we provide evidence that the distal tip of the tail needle is not required to confer host specificity on phage P22 and that the presence of bound L-glutamate in the Sf6 and HS1 hybrid needles does not affect P22 infection of *Salmonella* in the laboratory. It remains likely that in P22 and Sf6 the C-terminal needle domains do have evolutionarily important functions that are not easily measurable under laboratory conditions. For example, moderate differences in the speed of DNA delivery from the virion are not likely to limit the ability to form plaques. Thus, even if DNA release from the virion were slowed considerably (see below), it is possible that a significant overall slowing of the infectious cycle would not be evident. Nonetheless, the existence of functional

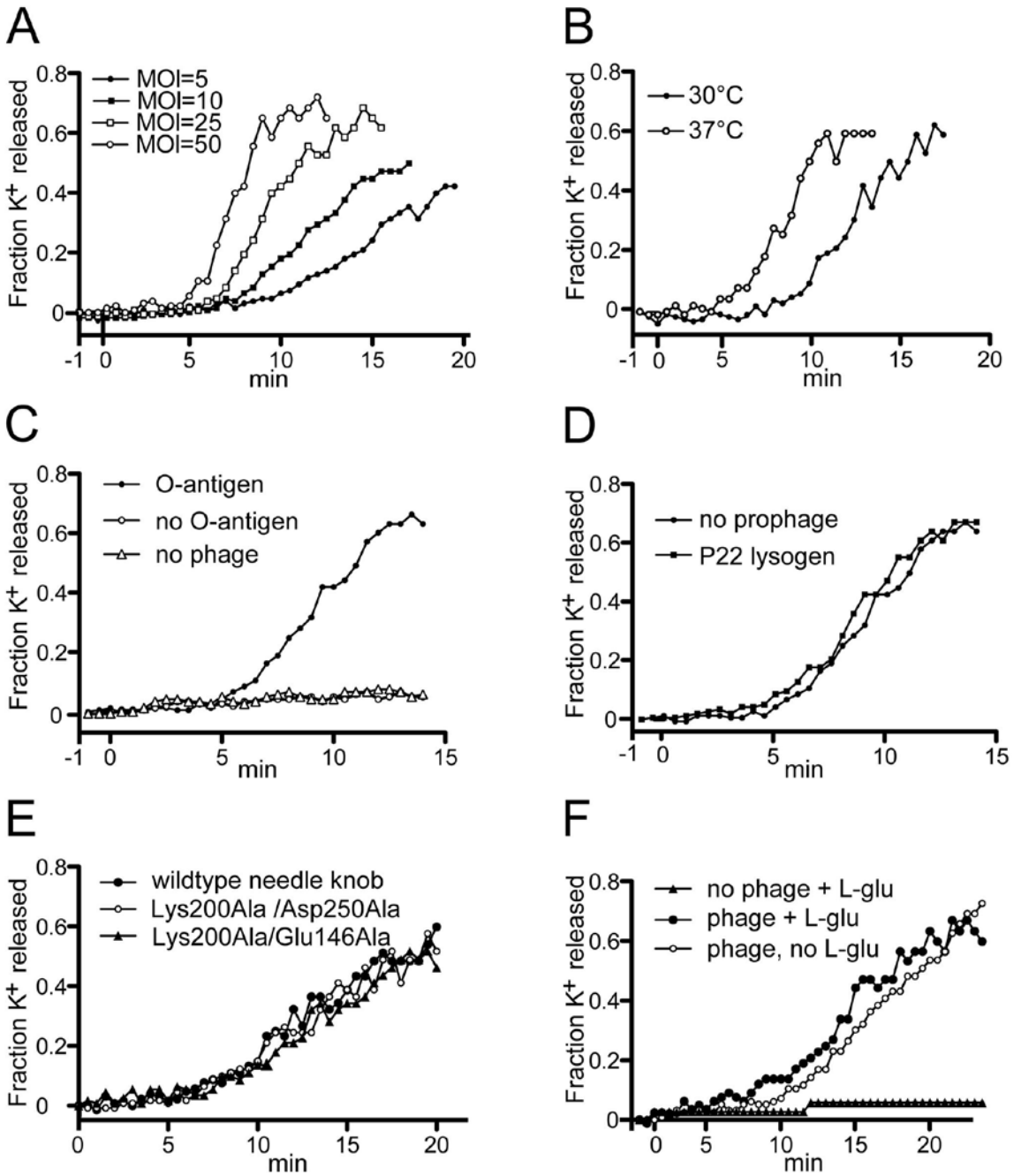


Figure 4. Potassium ion release by phage P22 infection. A. Infection of *Salmonella* strain UB-0001 by P22 clear mutant phage UC-0011 at different multiplicities of infection (MOIs). Full strain genotypes at given in Table 2 of article text. **B.** Infection of *Salmonella* UB-0001 at 30° and 37°C by phage P22 UC-0011. **C.** Infection of *Salmonella* host strains that have (UB-0001) or do not have (UB-2130) P22's O-antigen surface polysaccharide receptor by phage P22 UC-011; potassium ion release by uninfected *Salmonella* UB-0001 is also shown for comparison. **D.** Infection of a *Salmonella* host that has no P22 prophage (UB-0001) and a host that carries a P22 prophage that expresses its repressor (*c2*) gene but is missing the *sieA* and *gtrABC* genes (UB-0134) by P22 UC-0011. **E.** Infection of *Salmonella* UB-0001 by P22 phages that carry two mutations in the tail needle knob that should abrogate L-glutamate binding (see Table 2 of article text for amino acid changes). All infections were carried out at 30°C and MOI of 10 unless otherwise indicated. **F.** Infection of *Salmonella* UB-0001 by P22 UC-0011 with and without 10 mM L-glutamate added to the medium outside of the cells. Potassium ion measurements were performed as described in Materials and Methods of article text.
doi:10.1371/journal.pone.0070936.g004

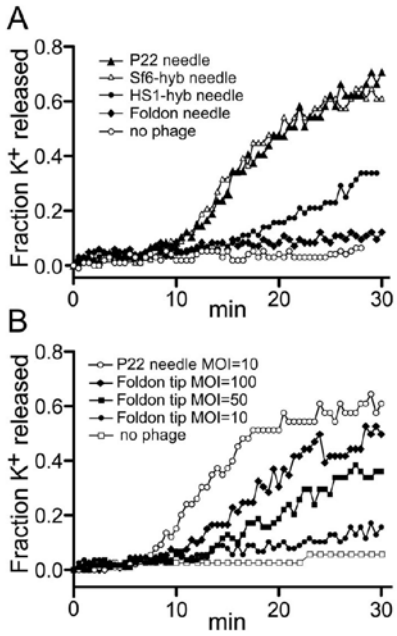


Figure 5. Potassium ion release by P22 phages with modified tail needles. A. Infection of *Salmonella* strain UB-0001 by P22 phages with altered needle proteins at MOI = 10. Potassium ion release was measured at 30°C as described in Materials and methods with a potassium electrode. *Salmonella* host strain UB-0001 was infected by the following phages: UC-0911 fully P22 tail needle (▲); UC-0918, needle has Sf6 C-terminal knob domain and part of the shaft (△); UC-0926, needle has HS1 C-terminal knob domain (●); UC-0927, foldon replaces needle C-terminal domain (◆); no phage infection (○). **B.** Infection of *Salmonella* strain UB-0001 by P22 phages with foldon-tipped needle at various MOIs as follows: P22 UC-0911 with fully P22 tail needle at MOI = 10 (○); P22 UC-0927 where foldon replaces needle's C-terminal domain at MOI = 10 (●), MOI = 50 (■) and MOI = 100 (◆); no phage (□). The horizontal axis is time after infection.
doi:10.1371/journal.pone.0070936.g005

virions that lack the native C-terminal tip of the needle has implications. It suggests that a DNA delivery model in which the needle tip domain makes an *essential* specific contact with a secondary receptor on the cell surface and then signals through the needle shaft to cause the needle's N-terminal portal channel plug domain to release from the virion, thus opening the channel for DNA exit, is not completely correct.

Experiments with phages that have foldon-tipped and HS1 hybrid needles show that modifications of the tip of the phage P22 tail needle can affect the kinetics of DNA release during injection as measured by K^+ release. Several non-mutually exclusive hypotheses for needle function could explain these results. For example, the needle tip could be part of the trigger that signals the virion to release its DNA, and modifying it could alters the kinetic of this signaling process. Although this may be true, the functionality of virions with the foldon tip indicates that the normal needle tip domain cannot be the only such trigger. Another, not mutually exclusive possibility is that the tail needle knob serves a mechanical function during genome ejection, comparable to the tip of a drill bit or hole punch in creating a hole in the inner membrane (the needle is released from the virion during DNA delivery [24], so it need not reach the inner membrane while still part of the virion). A larger globular domain at gp26 distal tip would then result in a "hole" of comparable diameter to the needle tip (~45 Å in Sf6/HS1) (figure 3), and our crystallographic studies have shown that P22's somewhat narrower ~35 Å C-terminal tip can swing by ±18 degrees with respect to the helical shaft [27,57], thus perhaps opening a hole larger than its diameter. Multiple copies of each of the P22 ejection proteins, gp7, gp16 and gp20, are also released from the virion during DNA delivery [24], and they are required for successful DNA entry into the cell cytoplasm. They are thought to perhaps build a structure or conduit that allows DNA transit from the virion through the membranes and periplasmic space into the cell [4]. Thus, if the needle protein in the membrane were to be replaced by the ejection proteins or were to create a hole that could then be occupied by the ejection proteins, the thinner foldon-tipped needle (diameter ~25 Å; figure 3) might be more difficult for the ejection proteins to utilize.

In conclusion, whatever mechanisms P22-like phages have developed to efficiently eject their genomes into Gram negative bacteria, our data make a function of the tail needle distal tip in host specificity unlikely (although the possibility remains of a specific knob function during infection of *Shigella* by Sf6), and provide evidence in support a role of this tip in facilitating the kinetics of genome delivery during infection.

Materials and Methods

Bacterial Strains

S. enterica serovar Typhimurium LT2 strain UB-1790 (*leuA*$^-$414, r$^-$, m$^+$, Fels2$^-$, *sup*$^\circ$ (P22 *sieA*$^-\Delta$1, *15*$^-\Delta$SC302::Kan^R, *13*$^-$amH101)) was used as the parent for alterations in the tail needle gene. The P22 prophage in this strain carries a kanamycin resistance gene to ensure prophage presence [58], a nonsense mutation in gene *13* to allow control of lysis by chloroform after induction to lytic growth [12], and the *sieA*$^-\Delta$1 deletion removes sequences that inhibit tailspike gene expression after induction but does not remove any essential genes [33,59,60]. The Kan^R insertion inactivates gene *15*, so titers were determined on LB plates containing 10 mM citrate [61]. The genotypes of the bacteria used in this study are given in table 2, and their construction is described below. Lytic growth of phages from lysogens was induced by addition of 0.5 μg/ml mitomycin C (Sigma, St. Louis, MO) and continued shaking at 37°C for 3–16 hr, and phage particles were titered on *S. enterica supE amber* mutant suppressing strain UB-0002 [62].

Phage Strains

Two recombineering strategies were used in these manipulations, as described (i) by Karlinsey [63] in which the tetracycline resistance cassette TetRA is selected for by requiring growth in the presence of tetracycline, and loss of TetRA is selected for by growth in the presence of anhydrotetracyline hydrochloride (Thermo Fisher Scientific, Waltham, MA) [33,58], and (ii) by Warming *et al.* [64] in which *galK*$^+$ insertions were selected by requiring growth on galactose as the sole carbon source and *galK*$^-$ bacteria were selected by growth in the presence of 2-deoxygalactose [33]. The phage lambda recombination function expressing plasmid pKD47 [65] was present during recombineering manipulations and was removed by growth at 42°C when it was no longer required. Electroporation was performed with a BIORAD Gene Pulser (25 uF, 2.4 KV, 200 W in 0.2 cm cuvettes). All genetic alterations were sequenced to confirm their structure.

P22 prophages that carry the C-terminal tail needle knob domain of phage Sf6 were constructed as follows: First a "recipient" prophage in which the TetRA cassette resides in P22 gene *26* was constructed. The TetRA cassette was amplified from

DNA of strain UB-1737 (the kind gift of K. Hughes) using oligonucleotides A and B (table S1) and inserted by homologous recombination between codon 64 and the stop codon of the UB-1832 prophage gene *26* resulting in strain UB-1940. Sf6 needle gene DNA was PCR amplified from phage Sf6 clear mutant [66] DNA with oligonucleotides C and D that amplify Sf6 bp 8455–9102 (Accession No. AF547987) and have 3′ P22 sequence tails that allow the amplified fragment to replace P22 bp 84113–8909 when recombined into the prophage genome of UB-1940. The resulting strain UB-1918 carries a prophage in whose hybrid needle gene (*26::Sf6-3*) P22 codons 69–233 are replaced by phage Sf6 gene *9* codons 69 through 282 (Sf6 needle gene *9* is orthologous to P22 gene *26*). This replacement includes part of the coiled-coil shaft domain and all of the C-terminal knob domain. The HS1 knob replacement (hybrid gene *26::HS1-1*) was made in an analogous manner as follows: An *E. coli galK* gene expression cassette was inserted into gene *26* of the prophage of strain UB-1790 by recombineering using DNA amplified from plasmid pGalK [64] (the kind gift of Don Court) with primers E and F whose P22 sequence 5′-tails result in the recombinational replacement of P22 gene *26* codons 142–233 by the *galK* gene and selection for growth on galactose. The resulting strain is UB-2078. DNA amplified from *E. coli* HS (UB-1732) DNA with primers G and H was used to replace the *galK* gene of UB-2078 so that codons 174 to 317 of the phage HS1 needle gene (locus_tag EcHS_A0316) replace P22 needle codons 141–233 to give strain UB-2083.

Modifications of the UB-1918 prophage (above) that contain point mutations in the L-glutamate binding site of the Sf6 needle knob were constructed as follows: A TetRA cassette was inserted into the P22 prophage of strain UB-1918 in three different places so that it replaces (i) codons 145 and 146 (primers I and J) in the 282 codon long, hybrid needle gene to create strain UB-1942, (ii) codon 200 (primers K and L and strain UB-1943) or (iii) codons 248 and 249 (primers M and N and strain UB-1944). Plasmid pPP304 carries the whole Sf6 *9* gene cloned into plasmid pET15b (Novagen, EMD Biosciences, Darmstadt, Germany), and single codon mutations were created in gene *9* of this plasmid with the QUICKCHANGE® site directed mutagenesis kit, Pfu Ultra Polymerase (Stratagene, La Jolla, CA), and restriction enzyme *Dpn*I as recommended by the manufacturer (New England Biolabs, Ipswich, MA). These modified plasmid genes were amplified with primers O and P and the resulting DNA was used to replace the TetRA cassette in UB-1942, -1943 or -1944. The changes made in these strains (UB-1919 through UB-1929) are indicated in table 2.

A prophage in which C-terminal codons 141–233 of P22 gene *26* are replaced by DNA encoding the T4 fibritin foldon [38,39] was constructed as follows: An *E. coli galK* gene expression cassette was amplified from plasmid pGalK [64] using primers Q and R whose P22 sequence 5′-tails result in the recombinational replacement of P22 bp 8403–8913 (gene *26* codons 65 through the stop codon) after electroporation into UB-1790 and selection to be *galK*$^+$. The *galK* gene of the resulting strain (UB-1807) was then replaced by DNA amplified from plasmid pMAL-PP-gp26(1–140)-F [39] with primers S and T; this plasmid carries a P22 gene *26* in which codons 141–233 are replaced by the 25 foldon codons, and the resulting prophage of strain UB-1941 carries this modified gene *26*.

Potassium Ion Efflux Measurement

Potassium ion concentrations were measured using an Orion Ionplus potassium electrode (Thermo Scientific) and a Corning model 430 pH meter. *Salmonella* cells were grown to 2×10^8 cells/ml in LB broth [67], spun down, washed once in KR buffer (10 mM $NaPO_4$ buffer, pH = 7.4, 100 mM NaCl, 10 mM $MgSO_4$) and finally resuspended at their initial concentration in KR buffer; no difference was observed between the two host strains UB-0001 or UB-0002. Aliquots of concentrated stocks of phage particles ($>10^{13}$/ml) that had been purified by CsCl step gradient centrifugation [68] and dialyzed against TM (10 mM TrisCl, pH 7.5, 1 mM $MgCl_2$) were used to infect the cells. Ten ml of the cell suspension was equilibrated to the desired temperature, the electrode was inserted into the cell suspension, phages were added 5 min later from concentrated stocks, and the mixture was briefly vortexed. The concentration of released potassium ions was monitored for 20–60 min after infection, and after these measurements the cells were lysed with Bugbuster reagent (EMD Millipore) or by boiling for 10 min, and total released K^+ was measured. The infections in each panel of figures 4 and 5 were performed on the same batch of cells. Some of the mutant phages used have somewhat different PFU/particle ratios (table 3), and it is unclear whether comparisons between phages should be done with equal numbers of PFU's or physical particles. The reason for these PFU/particle differences are not known, but since the virion proteins appear to be present in the same numbers in the different phages (in particular tailspike appears to be the same) there is no reason to believe that some particles are physically different from others in any given phage genotype. We performed comparative experiments both ways (equal numbers of PFUs or equal numbers of particles) and obtained results that gave the same qualitative conclusions. The data was more reproducible with equal numbers of physical particles, so all the K^+ release curves presented here were obtained this manner, and MOIs were calculated assuming a PFU/particle ratio of one for "wild type" P22 (UC-0911) and the PFU/particle ratios in table 3.

Cloning, Expression and Purification of HS1 Knob

The gene coding prophage HS1 tail needle knob (residues 167–317) was PCR amplified from strain UB-1732 DNA (locus_tag EcHS_A0316, Accession No. CP00802) and cloned in a pMal-c2e expression vector (New England Biolabs) between restriction sites XbaI and HindIII (plasmid pMal-HS1-knob). This plasmid was expressed in *E. coli* BL21(DE3) pLysE at 37°C as follows: Cells were grown to $A_{595} = 0.6$, the culture was induced for 16 h at 22°C by the addition of 0.5 mM isopropyl 1-thio-β-D-galactopyranoside, and the resulting cells were lysed by sonication in lysis buffer (20 mM Tris-HCl pH 8.0, 250 mM NaCl). The protein containing the HS1 knob fused to an N-terminal maltose binding protein (MBP-HS1) was purified by Amylose affinity chromatography (New England Biolabs). The fusion protein was digested with PreScission protease (GE Healthcare) and the resulting free HS1 knob protein was purified by Superdex 200 size exclusion chromatography (GE Healthcare) followed by passage over a 5 ml DEAE column (Sigma) that captured the MBP; pure HS1 tail needle knob was recovered in the DEAE column flow-through. The purified HS1 knob protein (~51 kDa) was concentrated to 10 mg/ml in 20 mM Tris-HCl pH 8.0, 50 mM NaCl using Sartorius ultracentrifugal filter device with a 10,000 Da molecular weight cutoff.

Crystallization and Structure Determination of HS1 Knob

HS1 knob was crystallized by mixing equal volumes of the protein solution and of a reservoir solution composed of 26% PEG 3350, 50 mM Tris-HCl pH 8.0 using the hanging drop vapor diffusion method. Crystals grew to full size over 2 weeks, were cryoprotected by quickly soaking in reservoir solutions also containing 27% ethylene glycol, and were flash-frozen in liquid

nitrogen. Diffraction data were collected at CHESS F1 and NSLS X6A beamlines. HS1 knob data were processed with the HKL2000 software package [69]. These crystals belong to space group P212121 and diffracted to 1.1 Å resolution (table 1). The structure was solved by molecular replacement with Phaser using the Sf6 tail needle knob (pdb code 3RWN) as search model [70]. One copy of trimeric HS1 knob was present in the asymmetric unit. The structure was built with COOT software [71] and refinement procedures were carried out using Phenix.refine [72]. The structure was validated by MolProbity [73] and analyzed by the EBI-PISA server [74] for interface interactions. All 151 residues of HS1 knob were unambiguously traced in the electron density map. The final model also includes 713 water molecules, one phosphate ion at distal tip of knob and three L-glutamate molecules at the dimeric interfaces formed by knob homotrimer (table 1). The coordinates and structure factors for HS1 tail needle knob have been deposited in the protein Data Bank with accession code 4K6B.

Supporting Information

Figure S1 Relationship of host bacterial species to P22-like phage tail needle tip domain type.
(PDF)

Figure S2 Relationships among the phage Sf6 type tail needle C-terminal domains of the P22-like phages.
(PDF)

Figure S3 Conserved residues in Sf6 tail needle knob-like proteins.
(PDF)

Figure S4 The hybrid P22:HS1-1 hybrid needle is incorporated into the virion.
(PDF)

Table S1 Oligonucleotides used in this study.
(PDF)

Acknowledgments

We thank Nicole Cumby and Alan Davidson for advice regarding potassium ion concentration measurements, and James Nataro, John Roth, Miriam Susskind, Don Court, and Kelly Hughes for bacterial strains.

Author Contributions

Conceived and designed the experiments: SRC JCL GC EBG LG AB. Performed the experiments: JCL LG EBG AB. Analyzed the data: SRC GC LG JCL EBG AB. Contributed reagents/materials/analysis tools: SRC GC LG JCL EBG AB. Wrote the paper: SRC GC JCL AB. Performed the bioinformatic analysis: SRC.

References

1. Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. Adv Exp Med Biol 726: 93–114.
2. Liu J, Chen CY, Shiomi D, Niki H, Margolin W (2011) Visualization of bacteriophage P1 infection by cryo-electron tomography of tiny *Escherichia coli*. Virology 417: 304–311.
3. Panja D, Molineux IJ (2010) Dynamics of bacteriophage genome ejection *in vitro* and *in vivo*. Phys Biol 7: 045006.
4. Casjens SR, Molineux IJ (2012) Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol 726: 143–179.
5. Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL (2012) Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol 726: 115–142.
6. Molineux IJ (2001) No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. Mol Microbiol 40: 1–8.
7. Molineux IJ (2006) Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? Virology 344: 221–229.
8. Hu B, Margolin W, Molineux IJ, Liu J (2013) The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339: 576–579.
9. Serwer P, Wright ET, Hakala KW, Weintraub ST (2008) Evidence for bacteriophage T7 tail extension during DNA injection. BMC Res Notes 1: 36.
10. Andres D, Baxa U, Hanke C, Seckler R, Barbirz S (2010) Carbohydrate binding of Salmonella phage P22 tailspike protein and its role during host cell infection. Biochem Soc Trans 38: 1386–1389.
11. Steinbacher S, Baxa U, Miller S, Weintraub A, Seckler R, et al. (1996) Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc Natl Acad Sci USA 93: 10584–10588.
12. Botstein D, Waddell CH, King J (1973) Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. I. Genes, proteins, structures and DNA maturation. J Mol Biol 80: 669–695.
13. Casjens S, King J (1974) P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. J Supramol Struct 2: 202–224.
14. Israel JV, Anderson TF, Levine M (1967) *in vitro* morphogenesis of phage P22 from heads and baseplate parts. Proc Natl Acad Sci U S A 57: 284–291.
15. Israel V, Rosen H, Levine M (1972) Binding of bacteriophage P22 tail parts to cells. J Virol 10: 1152–1158.
16. Andres D, Hanke C, Baxa U, Seul A, Barbirz S, et al. (2010) Tailspike interactions with lipopolysaccharide effect DNA ejection from phage P22 particles *in vitro*. J Biol Chem 285: 36768–36775.
17. Berget PB, Poteete AR (1980) Structure and functions of the bacteriophage P22 tail protein. J Virol 34: 234–243.
18. King J, Lenk EV, Botstein D (1973) Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. J Mol Biol 80: 697–731.
19. Strauss H, King J (1984) Steps in the stabilization of newly packaged DNA during phage P22 morphogenesis. J Mol Biol 172: 523–543.
20. Olia AS, Al-Bassam J, Winn-Stapley DA, Joss L, Casjens SR, et al. (2006) Binding-induced stabilization and assembly of the phage P22 tail accessory factor gp4. J Mol Biol 363: 558–576.
21. Olia AS, Bhardwaj A, Joss L, Casjens S, Cingolani G (2007) Role of gene *10* protein in the hierarchical assembly of the bacteriophage P22 portal vertex structure. Biochemistry 46: 8776–8784.
22. Hoffman B, Levine M (1975) Bacteriophage P22 virion protein which performs an essential early function. II. Characterization of the gene *16* function. J Virol 16: 1547–1559.
23. Hoffman B, Levine M (1975) Bacteriophage P22 virion protein which performs an essential early function. I. Analysis of *16-ts* mutants. J Virol 16: 1536–1546.
24. Israel V (1977) E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. J Virol 23: 91–97.
25. Lenk E, Casjens S, Weeks J, King J (1975) Intracellular visualization of precursor capsids in phage P22 mutant infected cells. Virology 68: 182–199.
26. Bhardwaj A, Molineux IJ, Casjens SR, Cingolani G (2011) Atomic structure of bacteriophage Sf6 tail needle knob. J Biol Chem 286: 30867–30877.
27. Olia AS, Casjens S, Cingolani G (2007) Structure of phage P22 cell envelope-penetrating needle. Nat Struct Mol Biol 14: 1221–1226.
28. Bhardwaj A, Walker-Kopp N, Casjens S, Cingolani G (2009) An evolutionarily conserved family of virion tail needles related to bacteriophage P22 gp26: correlation between structural stability and length of the α-helical trimeric coiled-coil. J Mol Biol 391: 227–245.
29. Clark AJ, Inwood W, Cloutier T, Dillon TS (2001) Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J Bacteriol 311: 657–679.
30. Villafane R, Zayas M, Gilcrease EB, Kropinski AM, Casjens SR (2008) Genomic analysis of bacteriophage epsilon 34 of *Salmonella enterica* serovar Anatum (15+). BMC Microbiol 8: 227.
31. Casjens SR, Thuman-Commike PA (2011) Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology 411: 393–415.
32. Pupo GM, Lan R, Reeves PR (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci USA 97: 10567–10572.
33. Padilla-Meier GP, Gilcrease EB, Weigele PR, Cortines JR, Siegel M, et al. (2012) Unraveling the role of the C-terminal helix turn helix of the coat-binding domain of bacteriophage P22 scaffolding protein. J Biol Chem 287: 33766–33780.
34. Papanikolopoulou K, Teixeira S, Belrhali H, Forsyth VT, Mitraki A, et al. (2004) Adenovirus fibre shaft sequences fold into the native triple beta-spiral fold when N-terminally fused to the bacteriophage T4 fibritin foldon trimerisation motif. J Mol Biol 342: 219–227.
35. Papanikolopoulou K, Forge V, Goeltz P, Mitraki A (2004) Formation of highly stable chimeric trimers by fusion of an adenovirus fiber shaft fragment with the foldon domain of bacteriophage T4 fibritin. J Biol Chem 279: 8991–8998.

36. Sissoeff L, Mousli M, England P, Tuffereau C (2005) Stable trimerization of recombinant rabies virus glycoprotein ectodomain is required for interaction with the p75NTR receptor. J Gen Virol 86: 2543–2552.
37. Stetefeld J, Frank S, Jenny M, Schulthess T, Kammerer RA, et al. (2003) Collagen stabilization at atomic level: crystal structure of designed (GlyPro-Pro)10 foldon. Structure 11: 339–346.
38. Tao Y, Strelkov SV, Mesyanzhinov VV, Rossmann MG (1997) Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. Structure 5: 789–798.
39. Bhardwaj A, Walker-Kopp N, Wilkens S, Cingolani G (2008) Foldon-guided self-assembly of ultra-stable protein fibers. Protein Sci 17: 1475–1485.
40. Boulanger P, Letellier L (1992) Ion channels are likely to be involved in the two steps of phage T5 DNA penetration into *Escherichia coli* cells. J Biol Chem 267: 3168–3172.
41. Cumby N, Edwards AM, Davidson AR, Maxwell KL (2012) The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. J Bacteriol 194: 5012–5019.
42. Letellier L, Plancon L, Bonhivers M, Boulanger P (1999) Phage DNA transport across membranes. Res Microbiol 150: 499–505.
43. Letellier L, Boulanger P, Plancon L, Jacquot P, Santamaria M (2004) Main features on tailed phage, host recognition and DNA uptake. Front Biosci 9: 1228–1339.
44. Ter-Nikogosian VA, Vartanian MK, Trchunian AA (1991) Changes in membrane potential and transport of ions through the *S. typhimurium* LT2 membrane induced by bacteriophages. Biofizika 36: 281–285.
45. Susskind MM, Botstein D, Wright A (1974) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. III. Failure of superinfecting phage DNA to enter sieA+ lysogens. Virology 62: 350–366.
46. Susskind MM, Botstein D (1978) Molecular genetics of bacteriophage P22. Microbiol Rev 42: 385–413.
47. Young B, Fukazawa Y, Hartman P (1964) A P22 bacteriophage mutant defective in antigen conversion. Viology 23: 279–283.
48. Molineux IJ, Panja D (2013) Popping the cork: mechanisms of phage genome ejection. Nat Rev Microbiol 11: 194–204.
49. Botstein D, Matz MJ (1970) A recombination function essential to the growth of bacteriophage P22. J Mol Biol 54: 417–440.
50. Weaver S, Levine M (1977) Recombinational circularization of *Salmonella* phage P22 DNA. Virology 76: 29–38.
51. Weaver S, Levine M (1977) The timing of erf-mediated recombination in replication, lysogenization, and the formation of recombinant progeny by *Salmonella* phage P22. Virology 76: 19–28.
52. Rhoades M, Thomas CA, Jr. (1968) The P22 bacteriophage DNA molecule. II. Circular intracellular forms. J Mol Biol 37: 41–61.
53. Van Valen D, Wu D, Chen YJ, Tuson H, Wiggins P, et al. (2012) A single-molecule Hershey-Chase experiment. Curr Biol 22: 1339–1343.
54. Garcia LR, Molineux IJ (1996) Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli*. J Bacteriol 178: 6921–6929.
55. Kemp P, Gupta M, Molineux IJ (2004) Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. Mol Microbiol 53: 1251–1265.
56. Struthers-Schlinke JS, Robins WP, Kemp P, Molineux IJ (2000) The internal head protein Gp16 controls DNA ejection from the bacteriophage T7 virion. J Mol Biol 301: 35–45.
57. Olia AS, Casjens S, Cingolani G (2009) Structural plasticity of the phage P22 tail needle gp26 probed with xenon gas. Protein Sci 18: 537–548.
58. Cortines JR, Weigele PR, Gilcrease EB, Casjens SR, Teschke CM (2011) Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. Virology 421: 1–11.
59. Adams MB, Brown HR, Casjens S (1985) Bacteriophage P22 tail protein gene expression. J Virol 53: 180–184.
60. Israel V (1967) The production of inactive phage P22 particles following induction. Virology 33: 317–322.
61. Casjens S, Eppler K, Parr R, Poteete AR (1989) Nucleotide sequence of the bacteriophage P22 gene *19* to *3* region: identification of a new gene required for lysis. Virology 171: 588–598.
62. Winston F, Botstein D, Miller JH (1979) Characterization of *amber* and *ochre* suppressors in *Salmonella typhimurium*. J Bacteriol 137: 433–439.
63. Karlinsey JE (2007) lambda-Red genetic engineering in *Salmonella enterica* serovar Typhimurium. Methods Enzymol 421: 199–209.
64. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using *galK* selection. Nucleic Acids Res 33: e36.
65. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640–6645.
66. Casjens S, Winn-Stapley D, Gilcrease E, Moreno R, Kühlewein C, et al. (2004) The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. J Mol Biol 339: 379–394.
67. Chan RK, Botstein D (1972) Genetics of bacteriophage P22. I. Isolation of prophage deletions which affect immunity to superinfection. Virology 49: 257–267.
68. Earnshaw W, Casjens S, Harrison S (1976) Assembly of the head of bacteriophage P22, X-ray diffraction from heads, proheads and related structures. J Mol Biol 104: 387–410.
69. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. Meth Enzymol: 327–326.
70. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674.
71. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
72. Adams PD, Gopal K, Grosse-Kunstleve RW, Hung LW, Ioerger TR, et al. (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. J Synchrotron Radiat 11: 53–55.
73. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12–21.
74. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.
75. Youderian P, Vershon A, Bouvier S, Sauer RT, Susskind MM (1983) Changing the DNA-binding specificity of a repressor. Cell 35: 777–783.
76. Youderian P, Chadwick SJ, Susskind MM (1982) Autogenous regulation by the bacteriophage P22 *arc* gene product. J Mol Biol 154: 449–464.
77. Levine MM, Caplan ES, Waterman D, Cash RA, Hornick RB, et al. (1977) Diarrhea caused by *Escherichia coli* that produce only heat-stable enterotoxin. Infect Immun 17: 78–82.
78. Levine MM, Bergquist EJ, Nalin DR, Waterman DH, Hornick RB, et al. (1978) *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. Lancet 1: 1119–1122.
79. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, et al. (2008) The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190: 6881–6893.
80. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
81. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32: W526–531.

Table S1
Oligonucleotides used in this study

Oligonucleotide[a]	Sequence (5'to 3')
A	ACGAAGCTGGACAGGGTGCTTATGACGCACAGGTAAAAAATGATGTTAAGA CCCACTTTCACATT
B	ACCATTTCTATCAGGTCGATGTTGCGTGTTGGAGTGAATGTAATCCTAAGC ACTTGTCTCCTG
C	GGACAGGGTGCTTATGACGCACAGGTAAAAAATGATGAGCAGGATGTGATT CTCGCTGACCAT
D	CATTTCTATCAGGTCGATGTTGCGTGTTGGAGTGAATGTAATCATTTACTG CTCCGCGATTATC
E	GGGTTACGACTGCCGAGAACAATATTTCGGCATTGCAGGCTGACTACGTAC CTGTTGACAATTAATCATCCGCA
F	CAACCATTTCTATCAGGTCGATGTTGCGTGTTGGAGTGAATGTAATCATTA TCAGCACTGTCCTGCTCCTT
G	GACTGCCGAGAACAATATTTCGGCATTGCAGGCTGACTACGCCACCACTCG CAAGAAATCA
H	CTATCAGGTCGATGTTGCGTGTTGGAGTGAATGTAATCATTACTGCTCCGC GATTATCT
I	CATCGATGCTCTGGAGTATGCAACCACACGCAAGAAGTCATTAAGACCCAC TTTCACATT
J	GCGCTGTCGGGATGGTTACAGATACACCAGAGTAAACAACCTAAGCACTTG TCTCCTG
K	GTGTTCAACGAGAACAAAACCCTGTTCTTCTTAAGACCCACTTTCACATT
L	CACTGGGCCACGTCCCGACAATCGACAGCTAAGCACTTGTCTCCTG
M	GGCGACAACGACCGATAACATCCTGTTAGCTACGTTCTTCTAAGACCCAC TTTCACATT
N	TGGTTAACGTGCTGCCATTTGTGGCAAGAAAGCCGTCTTTCTAAGCACTT GTCTCCTG
O	CCAACGAGGCCGGACAGGGCGCTTATGATGCACAGGTC
P	GCTCCGCGATTATCTTGATGGTTGTGGCAGTAAACGACGC
Q	ACGAAGCTGGACAGGGTGCTTATGACGCACAGGTAAAAAATGATGCCTGTT GACAATTA
R	ACCATTTCTATCAGGTCGATGTTGCGTGTTGGAGTGAATGTAATCTCAGCA CTGTCCTGCTCCTT
S	GGCAGACCCGTCACTTAATAATCC
T	TTGCGTGTTGGAGTGAATGTAATCATTAATCAATCAACCCATGTGCTTATA AAAAGGTAGAAAGCAATACC

a. Oligonucleotides were synthesized by the University of Utah oligonucleotide core facility.

Figure S1 Leavitt *et al.*

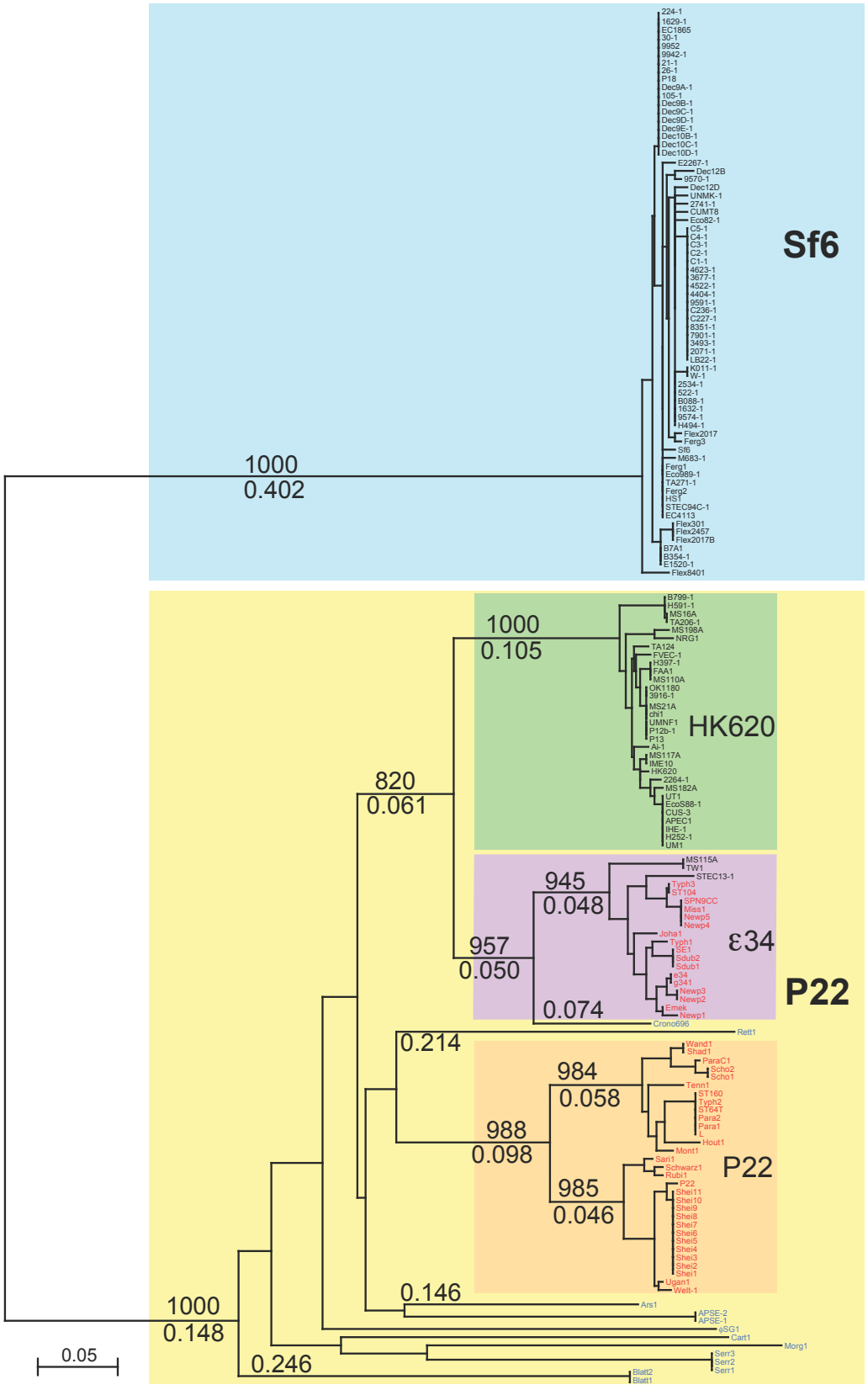


Figure S1. Relationship of host bacterial species to P22-like phage tail needle tip domain type.

A neighbor joining tree (created with Clustal X2 [Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, *et al.* (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948]) is shown with selected branch lengths (numbers between 0. and 1) and bootstrap values out of 1000 trials (between 1 and 1000). Bootstrap values for the nodes that are not well-supported are not shown. A scale in fractional difference is shown in the lower left. The Sf6 type domains are highlighted with a large blue box and P22 type domains with a large yellow box; these two domain types are not homologous but are combined in this tree to demonstrate the large sequence distance between them. *E. coli, Escherichia fergusonii* and *Shigella* phages and prophages are named in red text and *Salmonella enterica* phages and prophages are shown in black text. Other host species are shown in blue text. The names of these other hosts are given in the legend of figure 1, except for Crono696-1, which is a prophage in the genome of *Cronobacter sakazakii* 696. Prophage details can be obtained from the corresponding author.

Figure S2 Leavitt *et al.*

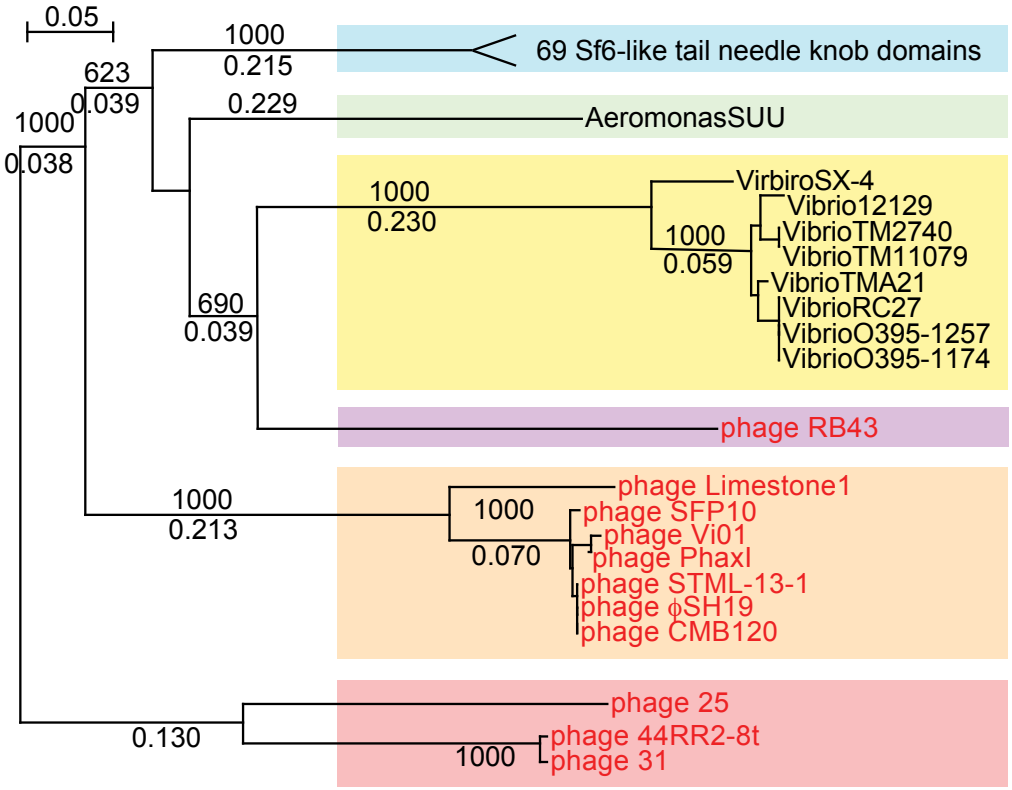
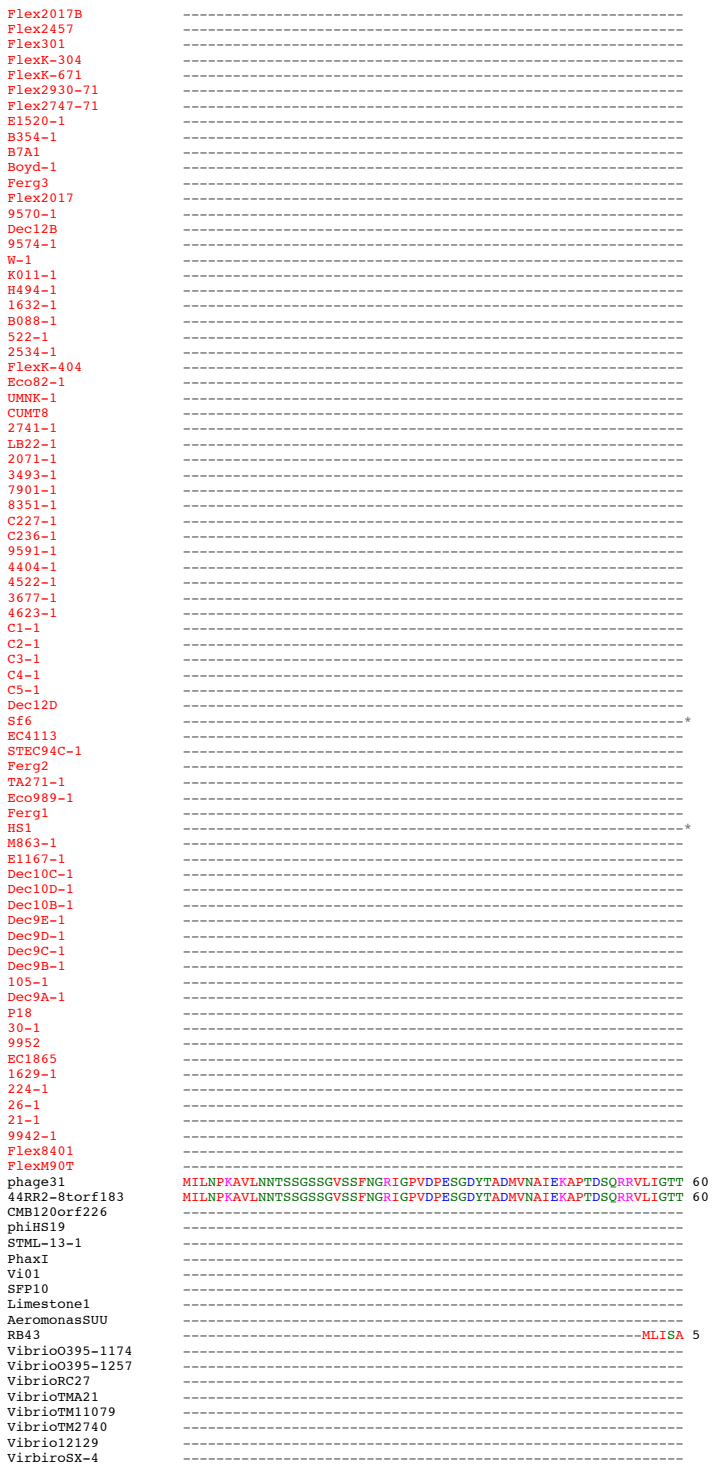


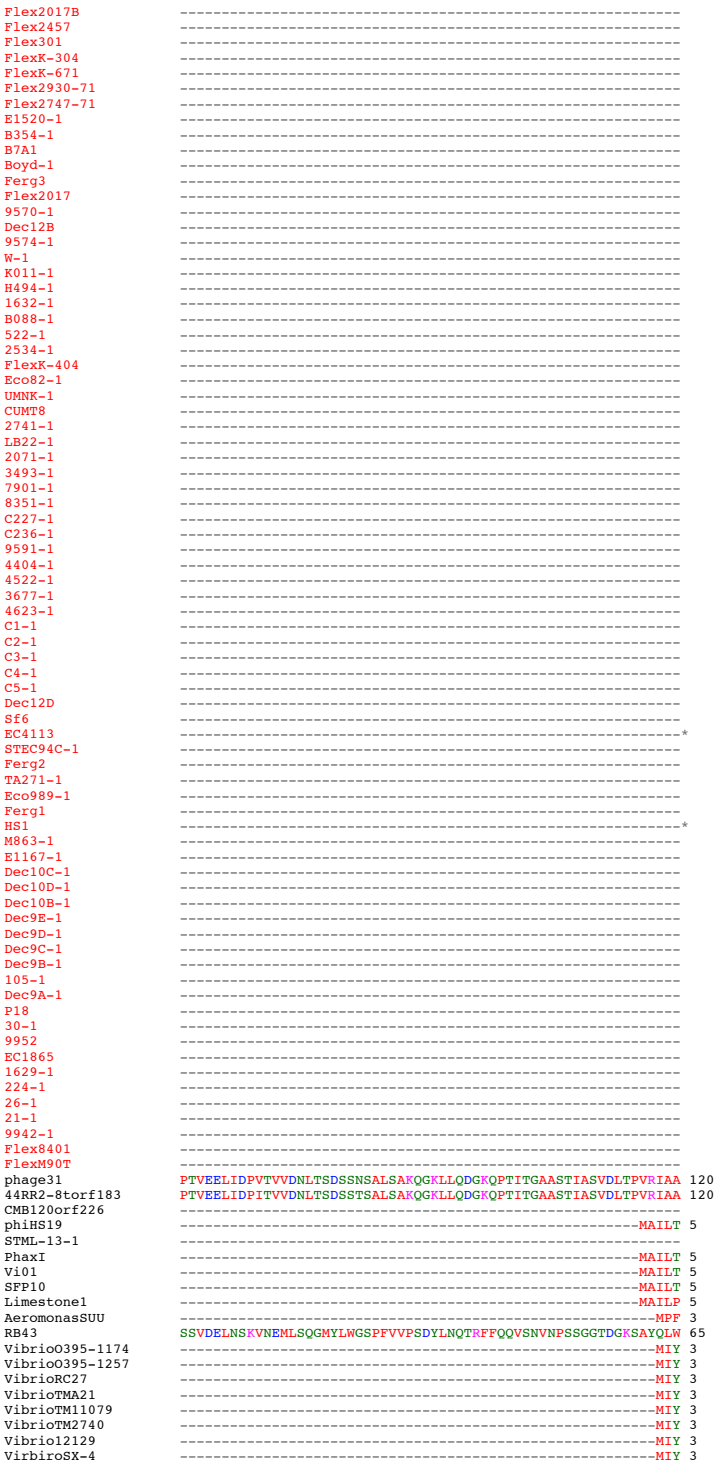
Figure S2. Relationships among the phage Sf6 type tail needle C-terminal domains of the P22-like phages.

A neighbor joining tree (created with Clustal X2 [Larkin et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948]) is shown with selected branch lengths (numbers between 0. and 1) and bootstrap values out of 1000 trials (between 1 and 1000). Bootstrap values for the nodes that are not well-supported are not shown. A scale in fractional difference is shown in the upper left. The six major sequence types are highlighted with colored boxes. Bone fide bacteriophages are labeled in red and prophages in black. Prophage names are provisional and some prophage names were reported in Casjens and Thuman-Commike [(2011) Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology 411: 393-415]. Prophage details can be obtained from the corresponding author.

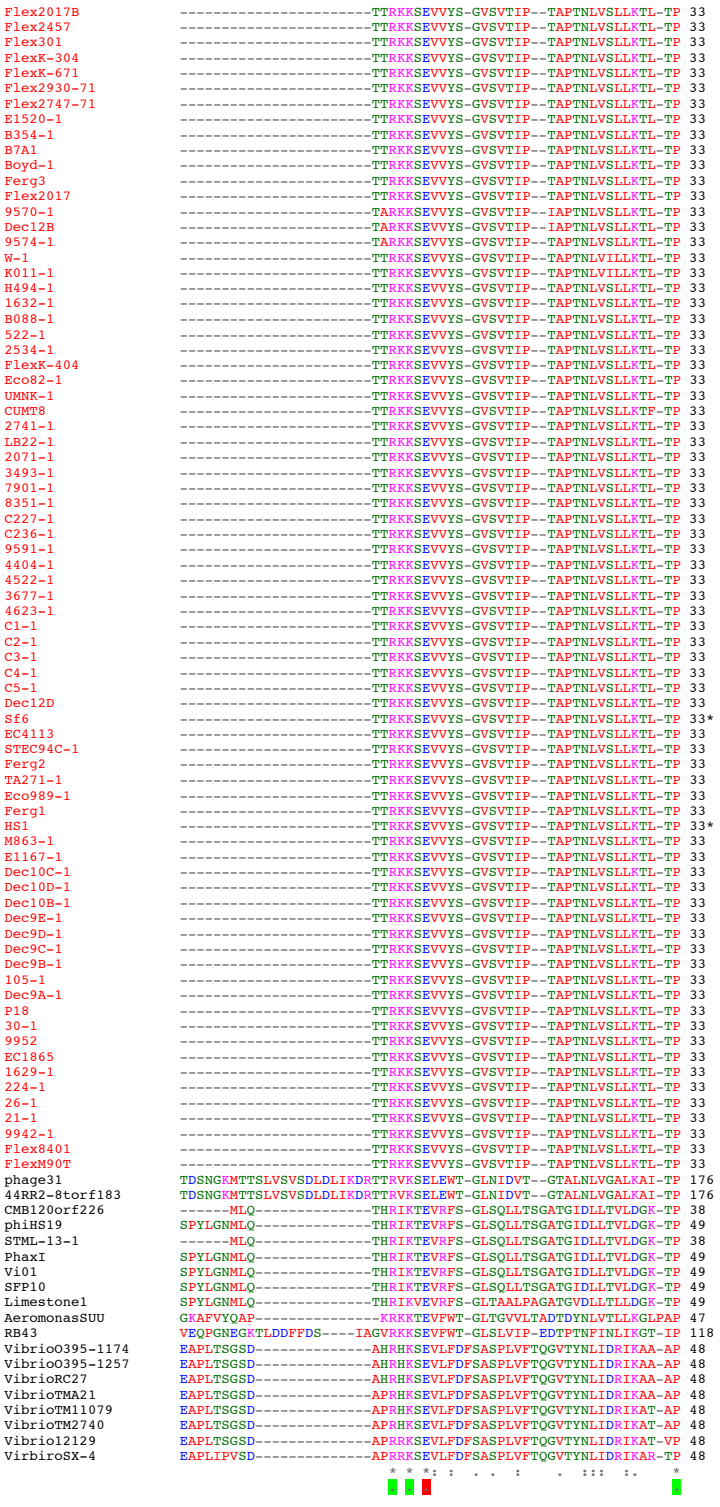
Leavitt *et al*. Figure S3 - page 1



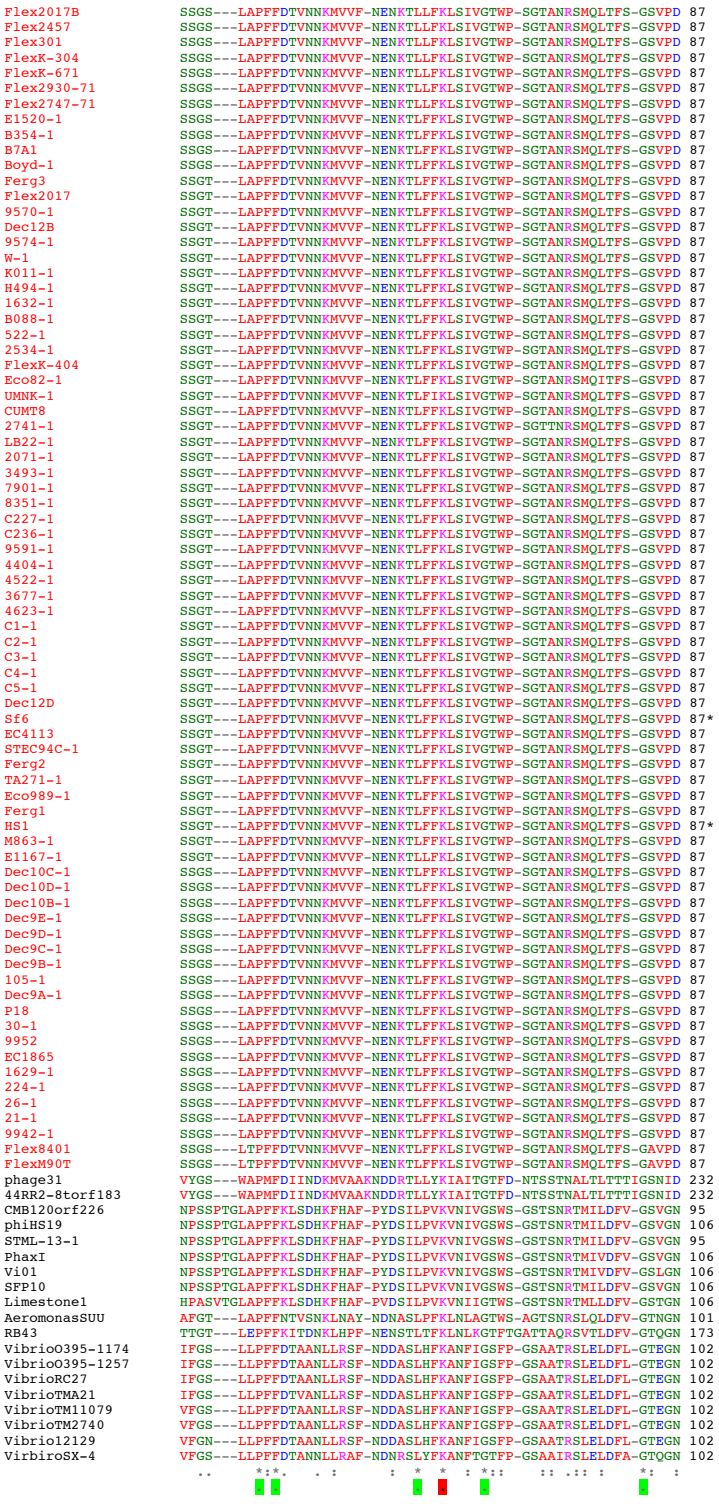
Leavitt *et al*. Figure S3 - page 2



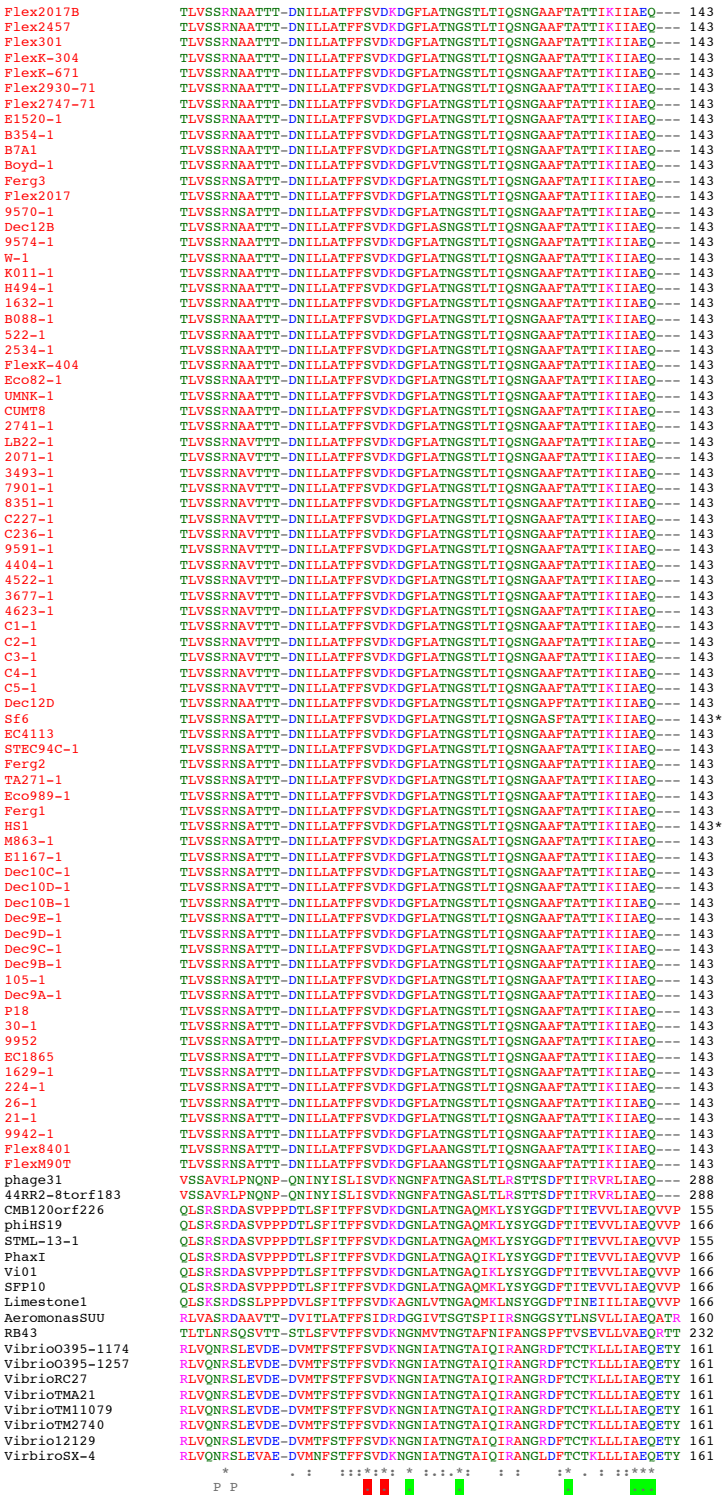
Leavitt *et al*. Figure S3 - page 3



Leavitt *et al*. Figure S3 - page 4



Leavitt *et al*. Figure S3 - page 5



Leavitt *et al*. Figure S3 - page 6

```
Flex2017B           ---------
Flex2457            ---------
Flex301             ---------
FlexK-304           ---------
FlexK-671           ---------
Flex2930-71         ---------
Flex2747-71         ---------
E1520-1             ---------
B354-1              ---------
B7A1                ---------
Boyd-1              ---------
Ferg3               ---------
Flex2017            ---------
9570-1              ---------
Dec12B              ---------
9574-1              ---------
W-1                 ---------
K011-1              ---------
H494-1              ---------
1632-1              ---------
B088-1              ---------
522-1               ---------
2534-1              ---------
FlexK-404           ---------
Eco82-1             ---------
UMNK-1              ---------
CUMT8               ---------
2741-1              ---------
LB22-1              ---------
2071-1              ---------
3493-1              ---------
7901-1              ---------
8351-1              ---------
C227-1              ---------
C236-1              ---------
9591-1              ---------
4404-1              ---------
4522-1              ---------
3677-1              ---------
4623-1              ---------
C1-1                ---------
C2-1                ---------
C3-1                ---------
C4-1                ---------
C5-1                ---------
Dec12D              ---------
Sf6                 ---------*
EC4113              ---------
STEC94C-1           ---------
Ferg2               ---------
TA271-1             ---------
Eco989-1            ---------
Ferg1               ---------
HS1                 ---------*
M863-1              ---------
E1167-1             ---------
Dec10C-1            ---------
Dec10D-1            ---------
Dec10B-1            ---------
Dec9E-1             ---------
Dec9D-1             ---------
Dec9C-1             ---------
Dec9B-1             ---------
105-1               ---------
Dec9A-1             ---------
P18                 ---------
30-1                ---------
9952                ---------
EC1865              ---------
1629-1              ---------
224-1               ---------
26-1                ---------
21-1                ---------
9942-1              ---------
Flex8401            ---------
FlexM90T            ---------
phage31             ---------
44RR2-8torf183      ---------
CMB120orf226        LYMTSI--- 161
phiHS19             LYMTSI--- 172
STML-13-1           LYMTSI--- 161
PhaxI               LYMTSI--- 172
Vi01                LYMTSI--- 172
SFP10               LYMSSI--- 172
Limestone1          LYMNTI--- 172
AeromonasSUU        QTQISAV-- 167
RB43                Q-------- 233
VibrioO395-1174     STEILGGGA 170
VibrioO395-1257     STEILGGGA 170
VibrioRC27          STEILGGGA 170
VibrioTMA21         STEMLGGGA 170
VibrioTM11079       STEMLGGGA 170
VibrioTM2740        STEMLGGGA 170
Vibrio12129         STEILGGGA 170
VirbiroSX-4         STEMLGGGA 170
```

Figure S3. Conserved residues in Sf6 tail needle knob-like proteins.

The proteins shown in the figure S1 tree were aligned by Clustal W at the web site http://www.ebi.ac.uk/Tools/msa/clustalw2/ [Larkin *et al.* (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948; asjens SR, Thuman-Commike PA (2011) Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology 411: 393-415]. The tail needle knob sequences containing only residues homologous to Sf6 amino acids 140 to 282 from P22-like phages are shown above with labels on the left in red text (Sf6 and HS1 are noted with an asterisk (*) at the right end of those lines. The complete sequences of the extant homologues of this domain from outside the P22-like phage group are aligned below with labels in black text of the left. The consensus sequence is shown below with asterisks (*) marking universally conserved amino acids, colons (:) marking amino acids with strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix, and periods (.) marking amino acids with weakly similar properties - scoring ≤ 0.5 in the Gonnet PAM 250 matrix. In the bottom line red boxes indicate the universal conservation of the residues that contact the bound glutamate, green boxes indicate other universally conserved amino acids, and "P" marks amino acids that contact the phosphate ion on the 3-fold axis in Sf6.

Figure S4 Leavitt *et al.*

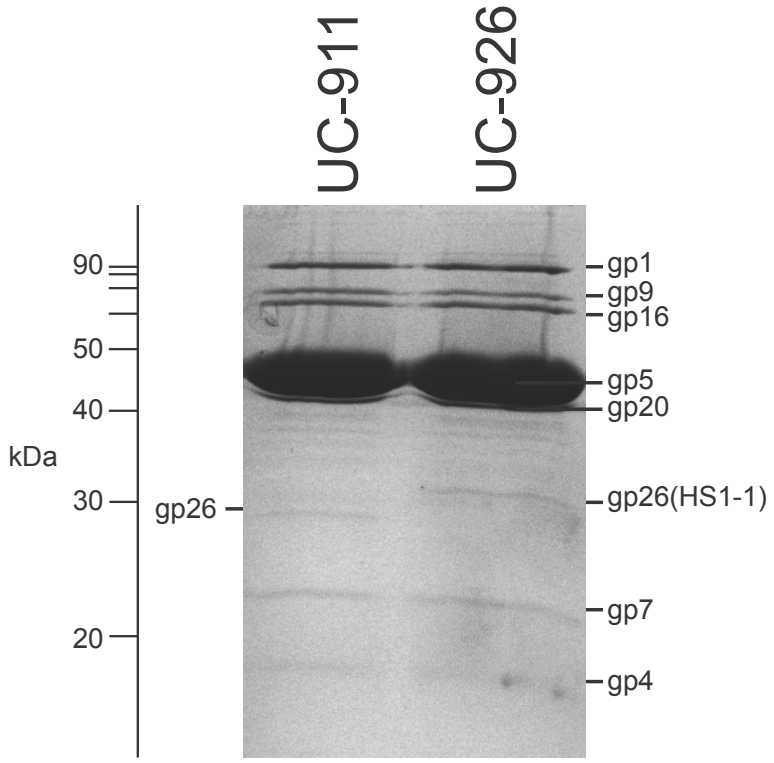


Figure S4. The hybrid P22:HS1-1 hybrid needle is incorporated into the virion.

Virions of P22 UC-911 and P22 UC-926 were purified through two successive CsCl step gradient centrifugations [Earnshaw W, Casjens S, Harrison S (1976) Assembly of the head of bacteriophage P22, X-ray diffraction from heads, proheads and related structures. J Mol Biol 104: 387-410], their proteins were separated in a 12% polyacrylamide SDS electrophoresis gel, and the gel was stained with Coomasie Brilliant Blue. The two virions have identical structural proteins except there is no band in UC-926 that corresponds to the P22 needle protein (gp26); however, there is a band of similar intensity that migrates somewhat slower UC-926 in virions. The P22:HS1-1 needle is expected to be about 6 kDa larger than the P22 protein, and we therefore suggest that this UC-926 protein band is the hybrid gp26. We note that several of the P22 virion proteins do not migrate precisely at the positions predicted by their molecular weights, and the kDa scale at the left was generated from commercially available size standards (BioRad).

CHAPTER 4

IN VIVO P22 *PAC* SITE RECOGNITION BY BACTERIOPHAGE P22 TERMINASE

Abstract

The P22-like bacteriophages package double stranded DNA into a preformed capsid shell in a headful fashion. DNA is recognized and packaged by a terminase ATP-powered packaging motor that is made up of two separate protein components, TerS (small terminase subunit), and TerL (large terminase subunit). Recognition of viral DNA is facilitated by a small 21 bp DNA sequence called the *pac* site that is located within the gene that encodes TerS. It is thought that DNA recognition is carried out primarily by TerS, but the mechanistic details of *pac* site recognition are still unclear. Here, we explore P22 phage production in the absence of a functional *pac* site and describe mutations of the terminase motor complex that have altered DNA packaging specificity.

Introduction

The tailed dsDNA bacteriophages are extremely varied, but all of them utilize a similar mechanism for packaging their DNA into virions (Casjens, 2011; Feiss and Rao, 2012). In this strategy, protein precursor structures called procapsids are assembled first, and a packaging DNA translocase uses energy from ATP cleavage to move the DNA into the preformed procapsid shell. DNA is packed very tightly within such virions, without the aid of DNA condensing proteins such as histones. DNA is not packaged randomly by the viral packaging machinery but virus DNA is specifically selected for packaging, usually through recognition of a specific sequence in the viral DNA. The packaging motor, usually called the "terminase", is typically made up of a complex that contains two different phage-encoded protein subunits; TerL, the terminase large subunit, carries the ATPase activity that powers the DNA translocase, and TerS, the small terminase subunit, is thought to recognize the DNA for packaging. Different tailed phages use several variations of this basic strategy. One of these is termed "headful packaging", so named because the size of the procapsid protein shell (*i.e.*, the phage head) determines the amount of DNA that is packaged (Streisinger et al., 1967). *Salmonella enterica* phage P22, *Escherichia coli* phage T4, and *Bacillus subtilis* phage SPP1 are the best understood phages of this type (Casjens, 2011).

Phage P22 DNA is replicated by a rolling circle mechanism that produces long head-to-tail concatemers of the 41724 bp long P22 genome sequence (Botstein and Levine, 1968). Such concatemers are the substrate for the packaging of DNA into virion particles (Jackson et al., 1978; Tye and Botstein, 1974). DNA packaging is initiated by recognition of a viral DNA sequence termed the *pac* site. In P22, this site is a 21 bp sequence that is both necessary and sufficient to initiate DNA packaging (Leavitt et al., 2013a; Wu et al., 2002). After recognition of the *pac* site, the DNA is cleaved and one of the ends generated is inserted into the procapsid, so packaging proceeds in one direction from the *pac* site (rightwards on the standard P22 genetic map). The procapsid is filled with DNA in a headful manner, *i.e.*, when the procapsid is full, the packaging motor performs a "headful DNA cleavage" that is not specific with regard to nucleotide sequence, and the packaged DNA molecule is released from the remainder of the concatemer. The head is sensed to be full when the packaged DNA is about 43400 bps long (104% of the length of the genome sequence) (Casjens and Hayden, 1988). This headful packaging strategy thus produces a 4% direct terminal redundancy in each packaged DNA molecule. A second packaging headful begins where the previous one ended, and additional headfuls follow in a similar fashion. A "packaging series" of as many as ten cycles of successive packaging events can occur on a single concatemer (Adams et al., 1983).

DNA is translocated into the capsid shell through the portal vertex, the five-fold vertex to which tails will later attach (Casjens et al., 1992; Chang et al., 2006; Lander et al., 2006; Moore and Prevelige, 2002; Tang et al., 2011). In P22, the portal protein (the product of gene *1*, or gp1) forms a dodecamer ring with a channel through its center through which DNA enters during packaging and leaves during injection (Olia et al., 2011; Tang et al., 2011). It also serves as a sensor for headful packing and interacts with terminase (Casjens et al., 1992). The P22 packaging motor is composed of the protein products of gene *2* (TerL) and gene *3* (TerS). These proteins were originally named “terminase” because they contain a nuclease activity that is responsible for creating the ends or termini of the DNA that is packaged. The C-terminal domain of TerL is very likely the endonuclease that cuts the DNA when the head is filled with DNA

(Nemecek et al., 2007; Ponchon et al., 2006; Roy and Cingolani, 2012), and its N-terminal domain contains an ATPase motif that is presumably responsible for powering packaging (Burroughs et al., 2007). Recognition of DNA for packaging is mediated *in vivo* by P22 TerS (Casjens et al., 1987), and a *pac* site is necessary for functional *in vivo* DNA packaging (Schmieger, 1984). The *pac* site is located inside gene *3* (Casjens et al., 1987; Weaver and Levine, 1978; Wu et al., 2002).

Atomic structures have been determined for three full-length TerS proteins of headful packaging phages. P22 TerS forms a nonamer ring (Nemecek et al., 2007; Nemecek et al., 2008; Roy et al., 2011), and the TerS of *Shigella* phage Sf6 (a fairly close relative of P22) forms an octamer ring (Zhao et al., 2010); both have an open channel at their center. Similarly, TerS of *B. subtilis* phage SF6, a much more distantly related phage, also forms a nonamer ring (Buttner et al., 2012). In addition, P22 TerS mutant A112T forms an apparently functional decamer ring (Nemecek et al., 2007), suggesting that the number of subunits in the TerS ring is not critical to the function of the packaging machine. It is not known whether the number of TerS subunits might be variable *in vivo* under normal low level expression conditions during phage infection. These three TerS subunits are not recognizably similar amino acid (AA) sequences and have somewhat related but nonidentical folds. The structure is also known for a dimeric N-terminal fragment of the phage lambda TerS (de Beer et al., 2002). Lambda does not use the headful strategy and the fold of this fragment is quite different from the other known TerS structures. P22 TerS and TerL form a complex *in vivo* and *in* vitro (Leavitt et al., 2013a; Poteete and Botstein, 1979; Roy et al., 2012; H. R. Brown and S. Casjens, unpublished), and TerS is needed for TerL ATPase activity; however, the structure of the functioning packaging motor complex remains unknown for P22, as it does for all dsDNA viruses that utilize this packaging strategy.

It is known that mutations in *terS* can affect the frequency of P22 generalized transduction; these mutants are called high transduction (HT) mutants. Therefore, it is believed that TerS is primarily responsible for specific DNA recognition. Previous work has not given a clear indication of exactly how any viral TerS protein contacts DNA. Both P22 TerS and Sf6 TerS bind to DNA *in*

vitro; however, neither protein has shown any specificity for its *pac* site (Nemecek et al., 2008; Roy et al., 2012; Zhao et al., 2012). Thus, the relevance of this nonspecific binding to *pac* site recognition is unknown. We previously replaced the N-terminal 134 codons of P22 TerS with Sf6 TerS codons 1-114 and found that this hybrid phage has Sf6 *pac* site specificity (Leavitt et al., 2013a). Thus, the N-terminal domain of Sf6 TerS dictates *pac* site specificity. This agrees with the previous observation that an E81K change within P22 N-terminal domain alters TerS target site specificity (Casjens et al., 1992). The location of this latter change suggests that DNA may contact the TerS ring on its outer rim. Work by Tang and coworkers that showed that Sf6 TerS outer rim alterations strongly affect nonspecific *in vitro* DNA binding is also consistent with this notion (Zhao et al., 2012). On the other hand, changes in the 9-stranded C-terminal beta-barrel domain (AAs 129-162) of P22 TerS, outside of the N-terminal *pac* specificity domain, can also affect DNA recognition and binding. Changes in AAs 129 and 152 have effects on packaging specificity *in vivo* (Casjens et al., 1992), and deletion of P22 TerS C-terminal AAs 142-162, but not 152-162, blocks nonspecific *in vitro* DNA binding by purified P22 TerS (Nemecek et al., 2008; Roy et al., 2012). These latter findings suggest that DNA binding may be more complex than simple outer rim contacts.

Two general models have been proposed for *pac* recognition by TerS: (1) TerS acts as a spool that binds the DNA on or around the outer rim of the oligomeric ring, or (2) DNA threads through the central channel of the TerS ring. The channel at the center of the TerS ring is wide enough to allow DNA entry into P22 procapsids, but the Sf6 octamer channel is slightly too small. P22 TerS is thought to interact with the *pac* site to specifically recognize viral DNA *in vivo*, since AA changes in its TerS alter the specificity of *pac* site recognition (Casjens et al., 1987; Casjens et al., 1992), and changes in outer rim AAs affect nonspecific *in vitro* DNA binding by Sf6 TerS (Zhao et al., 2012), but in no virus is the mechanism by which a TerS protein functions understood. In order to gain insight into how specific parts of TerS facilitate *pac* recognition, we devised *in vivo* experiments to examine the effects that changes in TerS AA residues have on DNA packaging and recognition of the *pac* site.

Pac site recognition domain of phage P22

We previously showed that when the N-terminal 134 AA domain of the native P22 TerS is replaced by the parallel domain (N-terminal 144 AAs) of Sf6 TerS, the resulting hybrid phage gives a normal phage burst upon induction. This "specificity swap" phage showed that the Sf6 N-terminal TerS domain is responsible for utilization of the Sf6 *pac* site (which is embedded inside the DNA that encodes this domain) for DNA packaging, and that the P22 C-terminal TerS domain interacts with the P22 TerL protein (Chapter 2 of this dissertation). This result suggests that the same is likely true for the analogous N-terminal P22 TerS domain. In order to experimentally determine whether the P22 N-terminal TerS domain is in fact recognizing the P22 *pac* site *in vivo,* we constructed a reciprocal phage construct whose TerS has a P22 N-terminal domain and an Sf6 C-terminal domain. In order to make this construct, we utilized the P22 hybrid prophage in *Salmonella* strain UB-2290 (diagrammed in Figure 4.1A) in which the entire P22 *terS* gene and the first 474 (out of 499) codons of *terL* are replaced by the Sf6 *terS* gene and the first 453 codons of Sf6 *terL*. This N-terminal Sf6 portion of *terL* is fused in-frame to the C-terminal 25 codons of P22 *terL*. This short P22 C-terminal TerL tail allows the Sf6 terminase (TerS and TerL) to function in the otherwise completely P22 context, presumably because it allows the Sf6 terminase to interact with P22 portal protein of the procapsid (to be described elsewhere; S. Casjens and E. Gilcrease, unpublished results). In the UB-2290 P22 prophage, we used *galK* recombineering to replace the N-terminal Sf6 TerS domain with that of P22 using three different *terL* Sf6-P22 fusion points shown in Figure 4.1B. One of these hybrid prophages, in strain UB-2355, contains the N-terminal 128 codons of P22 TerS fused to the C-terminal 28 codons of Sf6 TerS. It gave a moderate yield of progeny phages upon induction. The other two fusions gave less than 10^{-5} the wild type yield of progeny phage. The functionality of the UB-2355 prophage shows that the P22 N-terminal TerS domain is functional and is utilizing its embedded P22 *pac* site for DNA packaging (this site is essential for phage growth). Restriction analysis of DNA from purified phage particles that result from the induction of UB-2355 showed that it has a normal P22 pac fragment (not shown), which in turn indicates that their DNAs were packaged by normal

packaging series that initiated at the P22 *pac* site. We conclude that in P22, as in Sf6, the N-terminal TerS domain is responsible for recognition of the *pac* site and thus the location of DNA packaging series initiation events.

P22 mutants with a defective *pac* site

The phage P22 TerS protein is involved in the selection of DNA to be packaged into virions *in vivo* (Casjens et al., 1987; Casjens et al., 1992; Schmieger, 1972; Schmieger and Backhaus, 1973). However, although this TerS protein has been found to bind DNA nonspecifically *in vitro*, no evidence for specific *pac* site binding exists. In this study, we therefore examine P22 DNA packaging *in vivo* by genetic means.

Since a P22 phage without a functional *pac* site should not give progeny virions efficiently, we first chose to examine mutants that can overcome a *pac* site defect. However, a phage with such a defect is nonfunctional and cannot be propagated lytically. Therefore, we performed these experiments with prophages that are replicated passively by the bacterial host and do not depend on lytic growth for survival. Prophages can harbor mutations that are lethal to virion assembly, and a *pac* site defect is expected to be such a mutation. Such a mutant prophage can be induced to lytic growth so the growth phenotype of mutations it carries can be determined. We therefore constructed a P22 prophage with a *pac* site defect. Wu *et al.* (2002) identified the *pac* site by creating P22 phages that carry two *pac* sites, genetically modifying one of the two sites, and analyzing the relative usage of the two sites during DNA packaging. In that study, they found a number of changes that greatly lower the effectiveness of the *pac* site to program initiation of packaging. However, since the P22 *pac* site lies inside the *terS* gene, most of the *pac* site changes studied by Wu *et al.* (2002) also alter the AA sequence of TerS. Only a few changes that severely lower *pac* site activity make synonymous codon changes. We selected two of the later changes to create a putatively "null" *pac* site that does not alter the AA sequence of TerS. In these two changes, the native glutamic acid codon 90 GAA codon is changed to GAG and the Serine codon 93 TCT is changed to TCG; these alter P22 bps 6 and 15, respectively, in the P22

genome sequence (see Figure 4.2; bp numbering as in accession No. BK000583).

These two changes were engineered into the prophage of *Salmonella* strain UB-1790 (whose prophage is called P22 UC-911) by recombineering methods as follows: A *galK* gene expression cassette was PCR amplified from plasmid pGalK (Warming et al., 2005) with oligonucleotide primers that have 3'-tails with appropriate P22 sequences that program insertion by homologous recombination between P22 bp 265 and 285 of gene *3* to make strain UB-1947 (for the recombineering methods used, see Leavitt et al., 2013a; Leavitt et al., 2013b; Padilla-Meier et al., 2012; and Appendix of this dissertation). The parental prophage, called P22 phage UC-911 (Padilla-Meier et al., 2012), contains a nonsense mutation in gene *13* to allow control of cell lysis, a deletion of the *Immunity I* region that increases tailspike gene expression after induction, and an indel that replaces gene *15* and neighboring DNA with a Kanamycin resistance cassette that allows easy selection for lysogens but causes a requirement for citrate in the medium to achieve wild type level lysis (Casjens et al., 1989). Using directed mutagenesis of a plasmid carrying a cloned copy of gene *3*, plasmid PP463 was constructed that contains both of the above *pac* site changes. The altered region was PCR amplified from the modified plasmid, and the resulting DNA was used to repair the UB-1947 prophage gene *3* by homologous recombination-mediated replacement of the *galK* cassette through selection by 2-deoxygalatose for loss of the ability to utilize galactose as a carbon source. The resulting modified lysogen is called strain UB-1954. When this prophage was induced by carbodox treatment (see Methods) at 37°C, less than 10^5 plaque-forming units/ml were produced, compared to about $1x10^{11}$/ml for the parental prophage.

The *pac* site region has been previously shown to serve as the initiation site for *in vivo* DNA packaging of P22 DNA, whether this DNA is replicating phage DNA (Jackson et al., 1978; Tye and Botstein, 1974; Wu et al., 2002), is on a prophage inserted into the host chromosome (Kufer et al., 1982; Weaver and Levine, 1978), or is on a plasmid (Schmieger, 1984). However, these experiments did not determine whether the *pac* site sequence is essential for phage P22 growth. We conclude from the above *pac* mutant phenotype that the *pac* site is in fact essential for P22 lytic growth. Curiously, this is in direct contrast to the observation by (Strobel et al., 1984) that in

an *in vitro* P22 packaging reaction, the *pac* site is not required (although TerS protein is required). The reason for this difference is not known, but it correlates with the failure to obtain *pac*-specific DNA binding by P22 TerS *in vitro*. It is possible that, since only the first event in a packaging series requires *pac* recognition (Adams et al., 1983; Casjens and Hayden, 1988; Jackson et al., 1978; Tye et al., 1974), that the i*n vitro* packaging reaction reflects the requirements of subsequent, non-series-initiating packaging events rather than series initiation.

Since the *pac* defective prophage in *Salmonella* strain UB-1954 should express all the functional proteins required for P22 virion assembly, we examined the possibility that virion-like particles are made that contain host and/or part of the P22 genome. Thus, exponentially growing 150 ml, 37°C, L broth cultures of strains UB-1954 and UB-1790 at $2x10^8$ cells/ml were induced by the addition of 1.5 ug/ml carbodox. After 2 hours of shaking at 37°C, cells were concentrated by centrifugation and lysed in 3 ml TM (10 mM TrisCl, 1 mM $MgCl_2$, pH=7.4) by shaking with chloroform. Cell debris was removed by a low speed centrifugation, and particles with P22 density were purified by CsCl density centrifugation. Tables 4.1 and 4.2 show that UB-1954 produces only about 10^{-5} as many infectious phage as UB-1790, but makes 5-15% as many virion-like particles as UB-1790. Although the virion-like particles made by UB-1954 make very few plaques, generalized transduction of the TetR cassette in the *galK* gene by these particles occurs at a frequency that is nearly 100 times greater per particle than by particles made by UB-1790 (Table 4.2). We conclude that (1) the transducing particles made by the *pac* null mutant are functional and can deliver their DNA to a target cell, and (2) the ratio of particles containing host DNA is considerably higher in the UB-1954 particles than in the UB-1790 particles.

A significant amount of P22 DNA is nonetheless packaged in the UB-1954 particles despite the lack of a functional *pac* site. Figure 4.3 shows an agarose gel of equal amounts of DNAs from UB-1790 and UB-1954 particles cut by several different restriction endonucleases. P22-specific DNA bands are present in both the UB-1790 (WT) the UB-1954 (*pac*-) particle DNA digests; however, they are considerably less intense in the latter (see NdeI lanes in Figure 4.3; this was confirmed by digestion with other enzymes, for example EcoRI which, like NdeI, has a number of

well separated P22 DNA fragments; not shown). However, no pac fragments are visible in the digests of UB-1954 particle DNA, suggesting that the P22 DNA in these particles was not packaged by normal packaging series. The latter conclusion is strengthened by the fact that, in addition to the absence of the left-end pac fragment, the first P22 headful right-end fragment bands are visible in NcoI and SacI UB-1790 particle digests, but are also absent from the UB-1954 particle DNA (Figure 4.3). We note that right-end fragments give diffuse gel bands because the P22 headful measuring sensing device is imprecise (see Casjens and Hayden (1988) for a discussion of the origin of such right end fragments). Since DNA packaging is a late phage function, and P22 late functions do not affect early gene expression, UB-1954 prophage induction should result in normal prophage excision and subsequent DNA replication (Susskind and Botstein, 1978). In such a cell under the above induction conditions, we expect there to be very roughly half host and half P22 DNA in the cell (estimated from the number of phage released per cell by the positive control UB-1790 induction). It is therefore reasonable to surmise that the UB-1954 particles contain full-length DNA molecules whose packaging initiated at non-*pac* (but *pac*-like?) sequences on host DNA and on replicating concatemeric phage DNA. We note that P22 DNA-containing particles are present in the UB-1954 lysate in about 10^5-fold higher amounts than the number of plaque-forming units (Table 4.2).

The ability of the UB-1954 particles that contain P22 DNA to inject their DNA into cells was measured by their ability to form prophages in host cells, a process that does not require a *pac* site. Since the UB-1790 and UB-1954 prophages carry a kanamycin resistance cassette, virion-like particles from the induction of these two strains were used to infect *Salmonella enterica* strain UB-20 at a multiplicity of infection (MOI) of 0.5 particles/cell. After 90 minutes, the culture was diluted and plated on kanamycin-containing LB plates. Table 4.2 shows that phage induced from UB-1954 form kanamycin-resistant lysogens as well as the wild type UB-1790 control. This is consistent with the idea that after UB-1954 induction, phage DNA replicates normally and random "accidental" initiation of packaging on phage concatemers creates particles that contain normal length P22 DNAs that were not generated by pac-specific packaging series. Such virions should

contain a full-length P22 genome that has the normal length of terminal redundancy and so might be expected to be functional; however, we suggest that they are not functional because their defective *pac* sites are too inefficiently utilized to allow plaque formation during normal infection conditions. We note that the UB-1954 induction above was performed under conditions where lysis was blocked, and higher amounts of phage DNA might be packaged at very late times (several hours post induction) than on a plate where cells lyses after about an hour of infection.

We also note that there is a clearly visible background smear of non-P22-specific DNA bands in the gel shown in Figure 4.3, and this smear is much more intense in the DNA from particles induced from UB-1954 than in those induced from UB-1790. This background must be due to packaged *Salmonella* host DNA. To confirm the presence of *Salmonella* DNA in these particles, we compared frequency of generalized transduction by the virion-like particles in UB-1954 and UB-1790 phage preps. Since both strains carry the same *tetR* marker that replaces the native *Salmonella galK* gene, transduction was measured by infecting the tetracycline sensitive strain *Salmonella* UB-20 at a cell density of $2x10^8$/ml with an MOI=0.5 of particles/cell, followed by shaking at 37°C for 90 minutes and plating on tetracycline-containing L plates. Although the number of plaque forming units from the UB-1954 induction is $>10^5$ lower than for UB-1790, the number of generalized transductants per particle was nearly 50-fold higher for UB-1954 than for UB-1790 particles (Table 4.2). Thus, the frequency of *tetR* transduction per particle is considerably higher even than that of high frequency of transduction mutants.

Mutations that alleviate the *pac* site defect

In order to find mutations that alter DNA specificity *in vivo*, the *pac* site defective prophage in strain UB-1954 (above) was induced, and plaque-forming phages were selected from the resulting lysates. Plaques appeared at a frequency of 10^4-10^5 per ml of lysate, which is a frequency normally associated with spontaneous single bp mutations (our unpublished observations). We hypothesized that phage from these plaque-forming isolates would have

mutations in the terminase genes that could overcome a defective *pac* site. Twenty-two plaques were chosen for further analysis. Each pac-null suppressing phage was used to lysogenize a naive *Salmonella* UB-20 host and found to give a yield of plaque-forming phages upon induction that is similar to the wild type parent UB-1790 (Table 4.1). Since the phages we re-lysogenized into a naive host, no fortuitous mutation outside the phage genome can be responsible for the growth phenotype. The nucleotide sequence of the pac-null suppressing phage *terS* genes was determined in order to learn whether the starting *pac* site mutations were still present and whether there were additional changes in the *terS* gene. TerS mutations were identified by Sanger sequencing (University of Utah Core Facility) and are listed in Table 4.1.

Twelve of the mutants have altered *pac* sites. Six of the mutant phages have a true reversion of the of the G at position 6 to the original A (see Figure 4.2; bp numbering as in accession No. BK000583), suggesting that this A to G mutation is important in the inactivation of the *pac* site, and that the second T to G mutation at bp 15 is tolerated for plaque formation in spite of the very poor usage of this mutant site observed by Wu *et al.* (2002). We note that their rather insensitive assay would likely not have distinguished between $\leq 5\%$ normal usage and no usage at all. Six of the revertants changed the detrimental *pac* site G at bp 6 to a T. Curiously, Wu *et al.* (2002) found that a T at this position also made usage of the *pac* site undetectable in their assay. There are two possible explanations for this apparent disagreement. (1) The G6T mutation isolated here changes the TerS glutamic acid codon 90 at this location to an aspartic acid codon. This modified TerS protein could have new properties that allow it to recognize the altered *pac* site with a T at position 6 or to recognize some other site in the P22 genome. However, restriction enzyme digested DNA from virion DNA of this mutant has an evident *pac* fragment (see Figure 4.4), indicating that the *pac* site is utilized for packaging by this mutant. (2) Alternatively, in the studies of Wu *et al.* (2002), the G6T mutation was present only in combination with a G19A nucleotide change which itself made the *pac* site only about 20% effective. The decrease to undetectable usage that they observed in *pac* utilization by the G6T+G19A mutant compared to the G19A single mutant could be due to a non-additive effect of

these two *pac* site alterations.

Nine of the pac-null suppressing mutations have alterations in the *terS* gene and retain the two changes that inactivate the *pac* site. These mutations were found to alter four different *terS* codons outside of the *pac* site as follows: E81K, D114N, D114G, and E129K. Curiously, all nine of these mutations alter the same AAs as are altered in previously known high transducing mutants HT12/4a (E81K), HT115/1a (D114N), and HT119/2 (E129K) isolated by the use of a screen for P22 high transduction frequency phages by Schmieger and coworkers (Casjens et al., 1992; Raj et al., 1974; Schmieger, 1972). These mutations could overcome the *pac* site defect by altering TerS specificity so that it (1) utilizes the A6G+T15G *pac* site, (2) utilizes a specific different sequence in the phage genome, or (3) has lowered specificity so that multiple sites are utilized for packaging series initiation. Our previous studies have shown that the E81K change, uniquely, alters the target specificity of TerS (Casjens et al., 1987; Casjens et al., 1992). Analysis of the *pac* fragments generated by these mutants show that none of the mutations outside the *pac* site restore utilization of the mutant *pac* site (see Figure 4.4).

The cryo-electron microscopic structure of the P22 TerS nonamer determined by Nemecek *et al*. (2007; 2008) and a higher resolution atomic structure of P22 TerS determined by Roy *et al*. (2012) show that it is a ring-shaped nonamer with a hole through the center that might accommodate a dsDNA molecule (see Chapter 1 of this work). Models have been proposed for dsDNA binding by P22 TerS (and the related phage Sf6 TerS) in which DNA binds through the central channel of the ring or binds at the outer rim (Nemecek et al., 2007; Nemecek et al., 2008; Zhao et al., 2012). Interestingly, the TerS AAs found to be altered in both the inner channel (D114 and E129) and outer rim (E81) of the TerS nonamer (Figure 4.5), suggesting that *pac* binding by TerS may well be more complex than originally anticipated.

The final pac-null suppressing mutation that gave a nearly normal yield of plaque-forming phage upon induction has an N116K change that lies in the putative N-terminal ATPase domain-encoding portion of the TerL protein (encoded by P22 gene *2*) (see Table 4.1). Again, lysogenization of a naive UB-20 host by this mutant (to create strain UB-1950) showed that the

growth phenotype of this prophage was not generated by a change the host genome. A pac-null revertant change in *terL* is a very surprising finding since the current working model in this field is that tailed phage TerL proteins contain the ATPase that powers the DNA translocase that moves DNA into the procapsid and the nuclease that cuts packaged DNA from the long molecule that is the substrate for packaging. But until now, there has been no evidence that any TerL participates in the choice of DNA to be packaged. To confirm that there is no second mutation elsewhere in the N116K TerL phage's genome, we determined its whole genome sequence using Illumina methodology, and the N116K change is the only change relative to the UB-1954 prophage parent in the entire genome. Additionally, we recombineered the N116K into new P22 UC-911 prophages with a wild type *pac* site (*Salmonella* strain UB-2445) and with the null *pac* site (*Salmonella* strain UB-2444) and confirmed that they were both functional. Thus, this mutant is not dependent on the mutant *pac* site. After digestion by an appropriate restriction enzyme, the DNA from the N116K mutant phage with the pac-null mutation has no pac fragment (not shown, but similar to phage produced by strain UB-1950), whereas DNA from the phage with the N116K mutation and a wild type *pac* site does have a pac fragment. It appears that although this mutant terminase does not require a wild type *pac* site, it nonetheless can use this site when it is present. In addition, the N116K TerL phage is a high transduction mutant (Table 4.3), suggesting that its target specificity is lower than wild type phage.

This change in TerL alone allows successful initiation of packaging in the absence a wild type *pac* site, and the high transduction phenotype suggests a lowered packaging specificity. However, the mechanism by which these occur is not clear. TerL and TerS are known to form a complex that does not require the presence of other proteins (McNulty et al., 2015; Poteete et al., 1979; Roy and Cingolani, 2012; H. Brown and S. Casjens, unpublished). If the two proteins interact before recognition of the *pac* site by TerS, then the interaction with TerL might affect the target specificity of TerS. The part of TerL that interacts with TerS is unknown in all phage systems, but our unpublished evidence (E. Gilcrease and S. Casjens, data not shown) strongly suggests that the extreme C-terminal ~20 AAs of P22 TerL interacts with the procapsid.

Convincing sequence similarity to other biochemically better-characterized phage TerL proteins argues strongly that the N-terminal half of the P22 TerL protein is the ATPase; the x-ray structure of the C-terminal TerL nuclease domain is known for P22 (Roy and Cingolani, 2012), but no structure is available for the N-terminal domain. If the N116K TerL change affects the TerS-DNA interaction, this interaction is likely far from the TerL-procapsid contact region.

Directed mutagenesis of TerS

Since most of the above *terS* mutations that overcome *pac* site inactivation were the same as previously characterized mutations that confer a high transduction phenotype, we constructed an additional 22 mutant P22 phages in which 20 AAs at the surfaces of the outer rim and inner channel are changed. These mutants were made by site-directed mutagenesis and recombineering of the P22 UC-0911 prophage in *Salmonella* strain UB-1790 as described above and in the Methods. The locations of these changes in TerS are listed in Table 4.3 and are shown on a ribbon diagram of the protein in Figure 4.5. In addition, we also constructed mutations in the phage Sf6 TerS as follows: Previously, we constructed a functional Sf6-P22 hybrid prophage (UB-2019) in which the packaging-specificity domain of Sf6 TerS (Sf6 AAs 1-114) replaces the corresponding P22 domain (Leavitt et al., 2013a), and we engineered changes in six outer rim AAs of the Sf6 portion of the *terS* gene in this hybrid. The effects of these various TerS mutations were examined by analyzing (1) the number of functional phage particles produced, (2) the frequency of generalized transduction (*i.e.*, non-phage-specific DNA packaging) and (3) the utilization of the wild type *pac* site.

The plaque-forming virus yields from these P22 TerS mutants in are listed in Table 4.3, and those from the Sf6 TerS mutants are listed in Table 4.4. Of the 22 AA changes engineered into P22 TerS, only deletion of arginine codon 18 and the K48E/L153Q double mutant were completely nonfunctional. Five other changes had yields that were more than ten-fold lower than wild type; Y86H, K128E, Q130R, and the L88A/R89A and C32A/C33A double mutants had PFU yields between $7x10^8$ and $5x10^9$/ml. AAs 128 and 130 are at the base of the beta-barrel domain

and are not required for *pac* site recognition (Leavitt et al., 2013a). AAs 32, 33, 86, 88, and 89 are near the outer rim, but numerous other changes in this region are not detrimental to TerS function.

On the other hand, phages with changes at 21 of the 29 AAs that have been altered in P22 TerS are functional; these 21 functional changes include 15 directed changes made here that do not affect function, 5 previously known AA changes known in high transduction mutants and the above pac-null revertants, and one *amber* mutant N6 (codon 45) that can be suppressed by insertion of several different AAs (Casjens et al., 1991) (see Table 4.3). Most of the 21 functional changes alter AAs in the channel and on the outer rim. Figure 4.5 shows that these changes cover a significant fraction of the AAs at the surface of the outer rim, and drastic outer rim charge reversal changes K21E, E23K, E81R, and E81K give normal phage yields, suggesting that they do not greatly affect P22 TerS function. Similarly, charge reversal inner channel mutations D114K and E129K as well as charge removal mutations K63A, D114A/G/N, and D123C do not inactivate TerS. Given the fact that *pac* site recognition is essential for P22 growth, it would be quite surprising if specific *pac* site binding occurs on either of these surfaces and yet can tolerate such drastic changes.

We also constructed mutants of phage Sf6 TerS in the UB-2019 P22/Sf6 hybrid prophage. Previously, Zhao *et al.* (2012) constructed Sf6 TerS expression plasmids that have several TerS AA changes, purified the encoded mutant proteins, and measured their ability to bind DNA nonspecifically *in vitro*. In order to understand the physiological role of this nonspecific DNA binding *in vivo*, we engineered these same changes into the Sf6-P22 TerS hybrid prophage and determined their *in vivo* functionality. Table 4.4 shows that we find no correlation between the ability to bind DNA nonspecifically *in vitro* with the *in vivo* functionality of Sf6 TerS. For example, K6E and K33E Sf6 TerS proteins fail to bind DNA in the *in vitro* assay, but these changes do not affect the *in vivo* phage yield. We tentatively conclude that, at least for the Sf6 TerS, the nonspecific DNA binding that is observed *in vitro* may not be related to TerS function *in vivo*.

Pac site utilization during virion assembly by TerS mutants

The location at which DNA is recognized and cleaved for initiation of packaging series can be determined by digesting purified P22 DNA with restriction enzymes and observing the pac fragment in agarose electrophoresis gels (see Chapter 1). The presence of this fragment is a direct consequence of utilization and cleavage of *pac* DNA by the terminase at the start of a packaging series. Each of the functional site-directed point mutants was analyzed in this way (except UB-2238 and UB-2244 which do not make a large enough quantity of phage particles). This data is shown in Figure 4.6, and the results are summarized in Table 4.3. The TerS mutant prophages in strains UB-2114 (E81R), UB-2169 (R123C), UB-2115 (E129A), UB-2170 (Y86H), and UB-2248 (E152K) do not generate a pac fragment. Surprisingly, the UB-2240 (R18del) prophage makes a significant fraction of complete particles that contain P22 DNA, despite not forming any visible plaques (Figure 4.6). Cleavage of the DNA from this mutant phage by restriction endonucleases does not generate a pac fragment, and there is a significant smear that is indicative of a large quantity of host DNA.

We find no convincing correlation between location of a mutation on the TerS structure and the presence or absence of a pac fragment; mutations on the outer rim and in the central channel can either utilize the native *pac* site or fail to use it. The different mutants give pac fragments of variable intensity relative to the other phage DNA bands. Low intensity *pac* fragment bands are thought indicate a longer than wild type average length for packaging series, which in turn has been interpreted to mean a lowered rate of series initiation (Adams et al., 1983; Jackson et al., 1978). Several of the TerS mutants, R18K, K48A, Y52C, and K63A, appear to have weak pac fragment bands (Figure 4.6) and so may have lowered initiation rates. The Sf6 *pac* fragment is created by a more imprecise DNA cleavage mechanism and so is more diffuse and difficult to visualize; thus the pac fragment was not analyzed in the Sf6 hybrid phage mutants in this study. We note here that it is possible that P22 TerS mutations in which the *pac* fragment is not visible in gels may in fact have a more imprecisely cut *pac* fragment similar to that of Sf6. Further experiments may be required to analyze this possibility.

Generalized transduction by TerS mutants

Previous studies have shown that about 2% of wild type P22 virions have packaged a phage genome-sized fragment of host DNA, and this DNA can be injected into a new host cell and recombine with its resident genome to result in a transduction event (Ebel-Tsipis et al., 1972a; 1972b). Schmieger and coworkers isolated P22 mutants that have transduction frequencies higher than that of wild type phage; as much as 50% of virions may contain only host DNA (Casjens et al., 1992; Schmieger, 1972). Since it has been argued that generalized transduction by P22 results from utilization of *pac*-like sites in the host genome (Schmieger, 1982), the frequency of generalized transduction was measured for the TerS mutants. The strains isolated here that carry the mutant prophages also have a tetracycline resistance (TetR) cassette that replaces their *galK* gene, so these lysogens were induced with carbodox and the resulting lysates were used to transduce *Salmonella* strain UB-134 to tetracycline resistance as described in the Methods. Table 4.3 shows the results of this study. Nine of the 22 mutants show reproducibly greater than two-fold increase in transduction frequency. Among the outer rim changes, R18K, K21E, and E81R have strong effects on transduction, and among the central channel mutations, D114K and R123C have strong effects. Other changes such as N19A and E23K on the outer rim and Y52C and E129A in the channel have little to no effect on transduction. With the exception of K48A, every change of a charged AA side chain that faces towards the center of the inner channel of TerS, resulted in a significant increase in the frequency of transduction (K63A, D114K/A/G/N, D123C, E129K).

The mutant P22 DNAs that do not generate a visible pac fragment do not correlate with the HT phenotype; HT mutant R123C has no pac fragment while HT mutant K21E does have a pac fragment, and normally transducing mutation E152K has no pac fragment while normally transducing mutant Y52K does have a pac fragment. The HT mutant results can be understood if some TerS proteins with lowered specificity can still use the native *pac* site while others cannot. Similarly, it is not surprising that mutants which transduce normally have a pac fragment, but the failure of normally transducing E152K to generate a wild type pac fragment is difficult to

understand at present, unless it utilizes another specific site on P22 DNA that has not yet been identified.

Mutational phenotypes support a dynamic recognition process

What can we conclude from the mutational analysis of P22 and Sf6 TerS? Unfortunately, a precise location of the TerS-*pac* site DNA binding interface responsible for specificity cannot be unambiguously deduced from the properties of the mutants that we isolated and studied. The complex pattern of residues effecting transduction and of *pac* site utilization (pac fragment generation) suggests that either both inner and outer surfaces of the TerS oligomer are important for DNA recognition or that small changes in the TerS AA sequence at many locations have surprisingly large effects on TerS function. Since wild type function is very sensitive to single AA changes, our results could be explained if the AA changes we studied cause changes in the conformation of TerS and the *pac* site binding activity of TerS requires a very specific conformation. Structural studies with phage Sf6 and phage SF6 TerS have strongly suggested that these proteins are quite flexible and may undergo large conformational changes (Buttner et al., 2012; Zhao et al., 2012). If flexibility and conformational changes are critical to TerS function *in vivo*, then (1) it makes sense that AA changes at many locations might affect the function of TerS as we have observed, and (2) the locations of AAs in the *currently known* atomic structures may not be reflective of their locations on the form of TerS protein that interacts with *pac* site DNA. In addition, the easiest way to understand the phenotype of the N116K TerL mutation that influences the choice of DNA to be packaged is to imagine a dynamic interaction between TerL and TerS in which TerL influences the specific conformation of TerS.

In summary, previous models proposing that simple contact of *pac* site DNA with the outer rim or with the inner channel of the TerS protein is responsible for recognition are not supported by our results. Our results strongly imply that *pac* recognition is more complex than these simple models. In addition, our analysis of several phage Sf6 TerS mutants indicates that the nonspecific DNA binding by TerS proteins that has been observed *in vitro* may not be reflective of *in vivo pac*

site recognition. Our work here also gives the first evidence that TerL is involved in specific DNA recognition.

Materials and methods

Gene *3* and other targeted sequences were determined by dideoxysequencing of PCR amplified DNA by the University of Utah Sequencing Core Facility, and whole genome sequences were determined by Illumina sequencing at Brigham Young University DNA Sequencing Center and at the University of Utah Sequencing Core Facility. Illumina sequencing data were assembled by Geneious (Kearse et al., 2012).

Salmonella strain UB-1954 was induced at $2x10^8$ cells/ml with a final concentration of 1.5 ug/ml carbodox at 37°C in L broth for 4-6 hours. The bacterial cells were then lysed with chloroform, and the cells debris was removed centrifugation. The resulting lysate was titered on *amber* suppressing strain UB-21, and the resulting plaques collected. Each revertant was re-lysogenized in *Salmonella* strain UB-20.

Measurement of the number of P22 virus-like particles isolated from induced UB-1954 cultures was preformed by SDS-PAGE gel analysis as follows: Phage particles were purified through a CsCl gradient and dialyzed against TM (10 mM TrisCl, 1 mM $MgCl_2$, pH=7.4) buffer. The particle number was estimated by comparison of the amounts of Coomassie brilliant blue-stained gene *5* protein (major capsid protein) with a known amount of parent phage UC-911 induced from *Salmonella* UB-1790.

Generalized transduction frequency was measured by growing each prophage-carrying strain to $2x10^8$ cells/ml at 37°C in L broth, inducing by addition of carbodox to a final concentration of 1.5 ug/mL, and shaking at 37°C for 6 hours. Cells were lysed by shaking with chloroform, and cell debris was removed by centrifugation. Phage lysates were titered on host strain UB-21. Strain UB-134 grown to $2x10^8$ cells/ml was infected at an MOI of 0.5 at 37°C in L broth for 90 minutes, and 150 uL of these infected cells was plated on a tetracycline plate. After incubating overnight at 37˚C, the number of tetracycline resistant transductant colonies was counted.

Acknowledgements

The reported work was supported by NIH grant RO1 GM114817 to SRC. We thank Julianne Grose, as well as Desi DeMille and Ed Wilcox at the Brigham Young University sequencing center, for help with whole genome sequencing. We also thank Kelly Hughes and John Roth for *Salmonella* strains and the University of Utah Core Sequencing Facility for nucleotide sequencing. Sherwood Casjens, Eddie Gilcrease, Kassandra Wilson, and Aleaxandra Heitkamp all assisted with this work directly.

References

Adams, M.B., Hayden, M., Casjens, S., 1983. On the sequential packaging of bacteriophage P22 DNA. J Virol 46, 673-677.

Botstein, D., Levine, M., 1968. Synthesis and maturation of phage P22 DNA. II. Properties of temperature-sensitive phage mutants defective in DNA metabolism. J Mol Biol 34, 643-654.

Burroughs, A.M., Iyer, L.M., Aravind, L., 2007. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. Genome Dyn 3, 48-65.

Buttner, C.R., Chechik, M., Ortiz-Lombardia, M., Smits, C., Ebong, I.O., Chechik, V., Jeschke, G., Dykeman, E., Benini, S., Robinson, C.V., Alonso, J.C., Antson, A.A., 2012. Structural basis for DNA recognition and loading into a viral packaging motor. Proc Natl Acad Sci U S A 109, 811-816.

Casjens, S., Eppler, K., Parr, R., Poteete, A.R., 1989. Nucleotide sequence of the bacteriophage P22 gene *19* to *3* region: identification of a new gene required for lysis. Virology 171, 588-598.

Casjens, S., Eppler, K., Sampson, L., Parr, R., Wyckoff, E., 1991. Fine structure genetic and physical map of the gene *3* to *10* region of the bacteriophage P22 chromosome. Genetics 127, 637-647.

Casjens, S., Hayden, M., 1988. Analysis *in vivo* of the bacteriophage P22 headful nuclease. J Mol Biol 199, 467-474.

Casjens, S., Huang, W.M., Hayden, M., Parr, R., 1987. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. J Mol Biol 194, 411-422.

Casjens, S., Sampson, L., Randall, S., Eppler, K., Wu, H., Petri, J.B., Schmieger, H., 1992. Molecular genetic analysis of bacteriophage P22 gene *3* product, a protein involved in the initiation of headful DNA packaging. J Mol Biol 227, 1086-1099.

Casjens, S.R., 2011. The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol 9, 647-657.

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.

Chang, J., Weigele, P., King, J., Chiu, W., Jiang, W., 2006. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. Structure 14, 1073-1082.

de Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M., Catalano, C.E., 2002. Insights into specific DNA recognition during the assembly of a viral genome packaging machine. Mol Cell 9, 981-991.

Ebel-Tsipis, J., Botstein, D., Fox, M.S., 1972a. Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. J Mol Biol 71, 433-448.

Ebel-Tsipis, J., Fox, M.S., Botstein, D., 1972b. Generalized transduction by bacteriophage P22 in *Salmonella typhimurium*. II. Mechanism of integration of transducing DNA. J Mol Biol 71, 449-469.

Feiss, M., Rao, V.B., 2012. The bacteriophage DNA packaging machine. Adv Exp Med Biol 726, 489-509.

Jackson, E.N., Jackson, D.A., Deans, R.J., 1978. EcoRI analysis of bacteriophage P22 DNA packaging. J Mol Biol 118, 365-388.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

Kufer, B., Backhaus, H., Schmieger, H., 1982. The packaging initiation site of phage P22. Analysis of packaging events by transduction. Mol Gen Genet 187, 510-515.

Lander, G.C., Tang, L., Casjens, S.R., Gilcrease, E.B., Prevelige, P., Poliakov, A., Potter, C.S., Carragher, B., Johnson, J.E., 2006. The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science 312, 1791-1795.

Leavitt, J.C., Gilcrease, E.B., Wilson, K., Casjens, S.R., 2013a. Function and horizontal transfer of the small terminase subunit of the tailed bacteriophage Sf6 DNA packaging nanomotor. Virology 440, 117-133.

Leavitt, J.C., Gogokhia, L., Gilcrease, E.B., Bhardwaj, A., Cingolani, G., Casjens, S.R., 2013b. The tip of the tail needle affects the rate of DNA delivery by bacteriophage P22. PLoS One 8, e70936.

McNulty, R., Lokareddy, R.K., Roy, A., Yang, Y., Lander, G.C., Heck, A.J., Johnson, J.E., Cingolani, G., 2015. Architecture of the complex formed by large and small terminase subunits from bacteriophage P22. J Mol Biol 427, 3285-3299.

Moore, S.D., Prevelige, P.E., Jr., 2002. Bacteriophage p22 portal vertex formation *in vivo*. J Mol Biol 315, 975-994.

Nemecek, D., Gilcrease, E.B., Kang, S., Prevelige, P.E., Jr., Casjens, S., Thomas, G.J., Jr., 2007. Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. J Mol Biol 374, 817-836.

Nemecek, D., Lander, G.C., Johnson, J.E., Casjens, S.R., Thomas, G.J., Jr., 2008. Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. J Mol Biol 383, 494-501.

Olia, A.S., Prevelige, P.E., Jr., Johnson, J.E., Cingolani, G., 2011. Three-dimensional structure of a viral genome-delivery portal vertex. Nat Struct Mol Biol 18, 597-603.

Padilla-Meier, G.P., Gilcrease, E.B., Weigele, P.R., Cortines, J.R., Siegel, M., Leavitt, J.C., Teschke, C.M., Casjens, S.R., 2012. Unraveling the role of the C-terminal helix turn helix of the coat-binding domain of bacteriophage P22 scaffolding protein. J Biol Chem 287, 33766-33780.

Ponchon, L., Boulanger, P., Labesse, G., Letellier, L., 2006. The endonuclease domain of bacteriophage terminases belongs to the resolvase/integrase/ribonuclease H superfamily: a bioinformatics analysis validated by a functional study on bacteriophage T5. J Biol Chem 281, 5829-5836.

Poteete, A.R., Botstein, D., 1979. Purification and properties of proteins essential to DNA encapsulation by phage P22. Virology 95, 565-573.

Poteete, A.R., Jarvik, V., Botstein, D., 1979. Encapsulation of phage P22 DNA *in vitro*. Virology 95, 550-564.

Raj, A.S., Raj, A.Y., Schmieger, H., 1974. Phage genes involved in the formation generalized transducing particles in Salmonella--Phage P22. Mol Gen Genet 135, 175-184.

Roy, A., Bhardwaj, A., Cingolani, G., 2011. Crystallization of the nonameric small terminase subunit of bacteriophage P22. Acta Crystallogr Sect F Struct Biol Cryst Commun 67, 104-110.

Roy, A., Bhardwaj, A., Datta, P., Lander, G.C., Cingolani, G., 2012. Small terminase couples viral DNA binding to genome-packaging ATPase activity. Structure 20, 1403-1413.

Roy, A., Cingolani, G., 2012. Structure of P22 headful packaging nuclease. J Biol Chem 287, 28196-28205.

Schmieger, H., 1972. Phage P22-mutants with increased or decreased transduction abilities. Mol Gen Genet 119, 75-88.

Schmieger, H., 1982. Packaging signals for phage P22 on the chromosome of *Salmonella typhimurium*. Mol Gen Genet 187, 516-518.

Schmieger, H., 1984. pac sites are indispensable for in vivo packaging of DNA by phage P22. Mol Gen Genet 195, 252-255.

Schmieger, H., Backhaus, H., 1973. The origin of DNA in transducing particles in P22-mutants with increased transduction-frequencies (HT-mutants). Mol Gen Genet 120, 181-190.

Streisinger, G., Emrich, J., Stahl, M.M., 1967. Chromosome structure in phage T4, iii. Terminal redundancy and length determination. Proc Natl Acad Sci U S A 57, 292-295.

Strobel, E., Behnisch, W., Schmieger, H., 1984. *In vitro* packaging of mature phage DNA by *Salmonella* phage P22. Virology 133, 158-165.

Susskind, M.M., Botstein, D., 1978. Molecular genetics of bacteriophage P22. Microbiol Rev 42,

385-413.

Tang, J., Lander, G.C., Olia, A.S., Li, R., Casjens, S., Prevelige, P., Jr., Cingolani, G., Baker, T.S., Johnson, J.E., 2011. Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in P22. Structure 19, 496-502.

Tye, B.K., Botstein, D., 1974. P22 morphogenesis. II: Mechanism of DNA encapsulation. J Supramol Struct 2, 225-238.

Tye, B.K., Huberman, J.A., Botstein, D., 1974. Non-random circular permutation of phage P22 DNA. J Mol Biol 85, 501-528.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., Copeland, N.G., 2005. Simple and highly efficient BAC recombineering using *galK* selection. Nucleic Acids Res 33, e36.

Weaver, S., Levine, M., 1978. Replication in situ and DNA encapsulation following induction of an excision-defective lysogen of *Salmonella* bacteriophage P22. J Mol Biol 118, 389-411.

Wu, H., Sampson, L., Parr, R., Casjens, S., 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol Microbiol 45, 1631-1646.

Zhao, H., Finch, C.J., Sequeira, R.D., Johnson, B.A., Johnson, J.E., Casjens, S.R., Tang, L., 2010. Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. Proc Natl Acad Sci 107, 1971–1976.

Zhao, H., Kamau, Y.N., Christensen, T.E., Tang, L., 2012. Structural and functional studies of the phage Sf6 terminase small subunit reveal a DNA-spooling device facilitated by structural plasticity. J Mol Biol 423, 413-426.

Table 4.1

Mutations that alleviate the *pac* site defect

Number of Isolates	*pac* site sequence[a]	Nonsynonymous codon change[b]	Lysate titer[c]	HT [d]	Pac fragment?	Strains[e]
–	GA**A**GATTTATC**T**	parental strain	1.6 x 10^{11}	–	Yes	UB-1790
–	GA**G**GATTTATC**G**	none (*pac*⁻ A6G/T15G)	<10^5	–	No	UB-1954
6	GA**A**GATTTATC**G**	none (*pac*⁻ T15G)	1.6 x 10^{11}	–	Yes	UB-1997, 1998, 2001, 2002, 2008, 2011
6	GA**T**GATTTATC**G**	TerS E90D (GAA->GAT)	1.8 x 10^{11}	–	Yes	UB-1948, 1949, 1951, 2005, 2007, 2013
5	GA**G**GATTTATC**G**	TerS E81K (GAA–>AAA)	7.0 x 10^{10}	HT12/4a	No	UB-1996, 2009, 2010, 2012, 2006
1	GA**G**GATTTATC**G**	TerS D114N (GAC–>AAC)	1.7x 10^{11}	HT115/1	No	UB-2004
2	GA**G**GATTTATC**G**	TerS D114G (GAC–>GGC)	1.1x 10^{11}	HT13/4b	No	UB-1953, 2003
1	GA**G**GATTTATC**G**	TerS E129K (GAG–>AAG)	1.3 x 10^{11}	HT119/2	No	UB-1952
1	GA**G**GATTTATC**G**	TerL N116K (AAC–>AAA)	2.0 x 10^{10}	–	No	UB-1950

a. Non-wild type nucleotides shown in bold; the wild type *pac* site region sequence is GAAGATTTATCT.
b. Changes relative to the wild type P22 genome are shown. The parental lysogen UB-1954 is strain UB-20 with a P22 *pac*⁻A6G/T15G, Δsc302(KanR), 13⁻*am*H101, Δ*sieA*-1 prophage.
c. Induced cultures were shaken for 6 hr at 37°C, and the values shown are an average of several determinations on host strain UB-0002.
d. Previously known high transduction (HT) mutation with same AA change (Casjens et al., 1992; Raj et al., 1974).
e. Parental wild type lysogens in first two lines. The following lines are strains in which the revertants of the defective *pac* site prophage were re-lysogenized.

Table 4.2

Transduction by a P22 phage with a defective *pac* site

Strain	Time (hr) [a]	TetR/ml [b]	KanR/ml [c]	PFU [d] /ml	Particles/ml [e]
UB-1790	2	4.0×10^{2}	8.6×10^{6}	2.1×10^{12}	2.1×10^{12}
UB-1790	6	–	–	1.3×10^{13}	1.3×10^{13}
UB-1954	2	1.5×10^{4}	6.0×10^{4}	–	1×10^{11}
UB-1954	3	1.7×10^{4}	3.8×10^{5}	–	5×10^{11}
UB-1954	4	1.5×10^{4}	2.5×10^{6}	–	7×10^{11}
UB-1954	5	1.6×10^{4}	2.6×10^{6}	–	1×10^{12}
UB-1954	6	0.8×10^{4}	7.5×10^{6}	$<10^{7}$	2×10^{12}

a. Hours after induction by carbodox addition.
b. Number of tetracycline resistant generalized transductants of *S. enterica* strain (see Methods).
c. Number of lysogens formed by integration of the kanamycin resistance-carrying UB-1790 prophage (P22 UC-911) after infection of *S. enterica* strain UB-20 at 2×10^{8} cells/ml at an MOI of 0.5 particles/ml.
d. PFU, plaque forming units
e. Concentration of purified P22 virion-like particles quantified by amount of major capsid protein (gp5) present in SDS-PAGE and assuming a particle/PFU ratio of 1.0 (data not shown).

Table 4.3
P22 Terminase mutations

TerS AA change	Codon change	Lysate titer[a]	Pac Transduction[b]	Lysogen fragment[c]	Strain
Wild type parent					
none	none	$1.0x10^{11}$	Yes	UB-1790	
TerS central channel					
K48A	GAA–>GCG	$1.0x10^{11}$	0.6 (WT)	Yes vw	UB-2112
K48E,L153Q	GAA–>GGA CTA–>CAA	$<10^5$	nd[e]	nd	UB-2238
Y52C	TAT–>TGT	$1.0x10^{11}$	1.5 (WT)	Yes w	UB-2163
Q53A	CAG–>GCG	$3.6x10^{11}$	1.8 (WT)	Yes	UB-2107
K63A[f]	AAG–>GCG	$3.4x10^{11}$	3.9 (HT)	Yes w	UB-2090
D114A	GAC–>GCG	$1.0x10^{11}$	5.3 (HT)	Yes	UB-2108
D114K	GAC–>AAA	$8.0x10^{10}$	6.6 (HT)	Yes	UB-2117
R123C	CGT–>TGG	$3.0x10^{10}$	11.4 (HT)	No	UB-2169
K128E	AAA–>GAG	$2.0x10^{8}$	nd	nd	UB-2244
E129A	GAG–>GCA	$1.5x10^{11}$	0.7 (WT)	Yes vw	UB-2115
Q130R[d]	CAG–>CGG[5]	$1.9x10^{9}$	nd	Yes	UB-2171
TerS outer rim					
R18del	AATdel	$<10^5$	nd	No	UB-2240
R18K	AAT–>AAA	$1.1x10^{11}$	5.5 (HT)	Yes w	UB-2227
N19A	AAT–>GCG	$3.2x10^{11}$	1.2 (WT)	Yes	UB-2113
K21E	AAA–>GAG	$3.0x10^{11}$	8.0 (HT)	Yes	UB-2088
E23K	GAA–>AAA	$1.1x10^{11}$	2.0 (WT)	Yes	UB-2243
E81R	GAA–>CGA	$3.8x10^{11}$	7.4 (HT)	No	UB-2114
Y86H	TAT–>CAT	$1.5x10^{9}$	nd	No	UB-2170
L88A,R89A	CTG–>GCG AGA–>GCG	$2.2x10^{11}$	3.8 (HT)	Yes	UB-2246
Other TerS mutations					
C32A,C33A	TGT–>GCG TGT–>GCG	$4.0x10^{9}$	nd	yes	UB-2230
Q53K,V39A	CAG–>AAA, GTG–>GCG	$3.1x10^{11}$	3.4 (HT)	Yes	UB-2242
E152K	GAG–>AAA	$2.0x10^{11}$	1.7 (WT)	No	UB-2248
TerL mutation					
TerL N116K	AAC–>AAA,*pac*$^+$	$4.0x10^{10}$	7.0 (HT)	Yes	UB-2445
TerL N116K	AAC–>AAA,*pac*$^-$	$2.0x10^{10}$	13.2 (HT)	No	UB-2444

Table 4.3 legend

a. Plaque-forming units/ml on *Salmonella* strain UB-21. A representative value is shown out of several determinations that gave similar results.
b. Frequency of transduction normalized to the parental phage P22 UC-0911, the prophage in *Salmonella* strain UB-1790; see Methods. If the value is greater than or equal to two times the UC-0911 value it is indicated as high transducing (HT) in parentheses.

Table 4.3 legend continued

c. Pac fragment determination electrophoresis gels are shown in Figure 4.6; w, weak pac fragment band; vw, very weak pac fragment band.

d. Also carries synonymous change GCT–>GCC codon 31

e. nd, not determined

f. Also carries synonymous change CGA–>CGG codon 65

Table 4.4

Alterations of phage Sf6 TerS

Sf6 TerS mutant	Phage yield[a]	DNA binding[b]	Transduction[c]	Strain Name[d]
WT[e]	$1.0x10^{11}$	yes	1.0	UB-2019
K6E[f]	$1.4x10^{11}$	no	0.3	UB-2440
D19R	$6.0x10^{10}$	enhanced	1.3	UB-2404
K33E	$1.4x10^{11}$	no	0.4	UB-2399
R48A	$>10^5$	weak	–	UB-2397
K59E[g]	$>10^5$	no	–	UB-2393
K59A[g]	$>10^5$	weak	–	UB-2401

a. Plaque-forming units/ml of culture grown and induced as described in Table 4.3.
b. Nonspecific *in vitro* dsDNA binding by purified mutant Sf6 TerS as reported by Zhao *et al*. (2012).
c. Transduction of tetracycline resistance presented as the ratio of the number of mutant transductions to the parental phage with unaltered Sf6 TerS.
d. Lysogen carrying the mutant prophages.
e. Sf6-P22 hybrid prophage is resident in *Salmonella* strain UB-2019.
f. This mutant also carries a synonymous codon 9 CGC to CGA change that should not alter the function of the protein.
g. These changes could lie in the Sf6 *pac* site (Leavitt et al., 2013a) and so failure to grow could be due to a defective *pac* site.

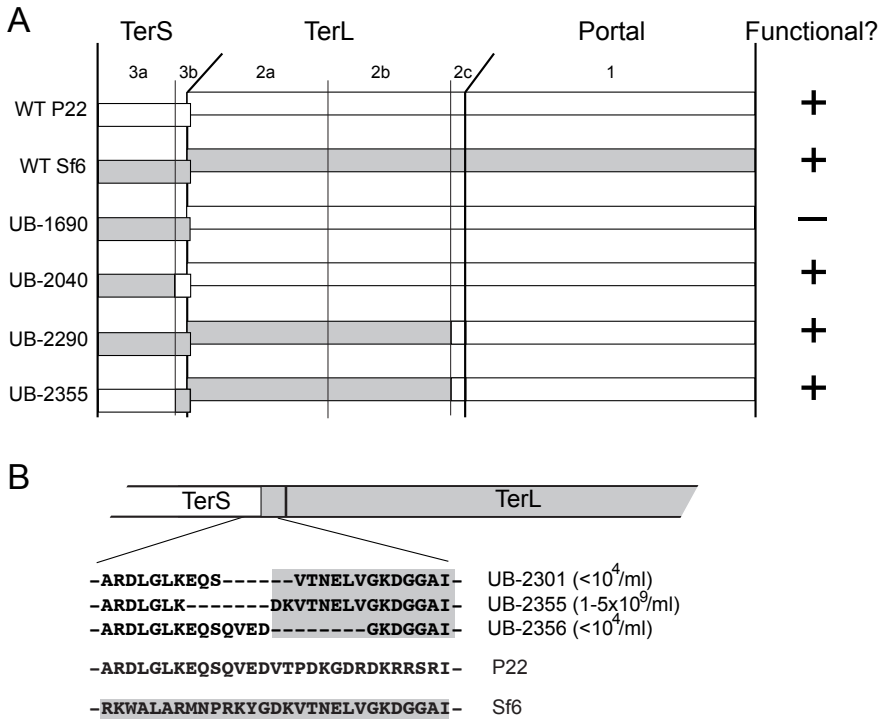


Figure 4.1 The N-terminal domain of P22 TerS determines *pac* site specificity.

A. Large (TerS) and small (TerL) subunit terminase and portal genes are shown by horizontal rectangles where white regions are P22 sequence and gray regions are Sf6 seauence. Thick vertical lines mark gene boundaries and thin vertical lines mark domain boundaries identified in the atomic structures of the encoded proteins and as mosaic boundaries in the phage genome sequences (Casjens and Thuman-Commike, 2011). The strain name of *Salmonella* that carries each prophage is shown on the left and the functionality of the construct is indicated on the right; the functionality of UB-1790 and -2040 is from Leavitt *et al.* (2013) and is shown here for comparison. **B.** The *terS-terL* hybrid gene boundary is shown diagrammatically above and the AA sequence of the TerS protein encoded at the hybrid junction is shown below. Phage titers of the induced lysogens are shown at the right.

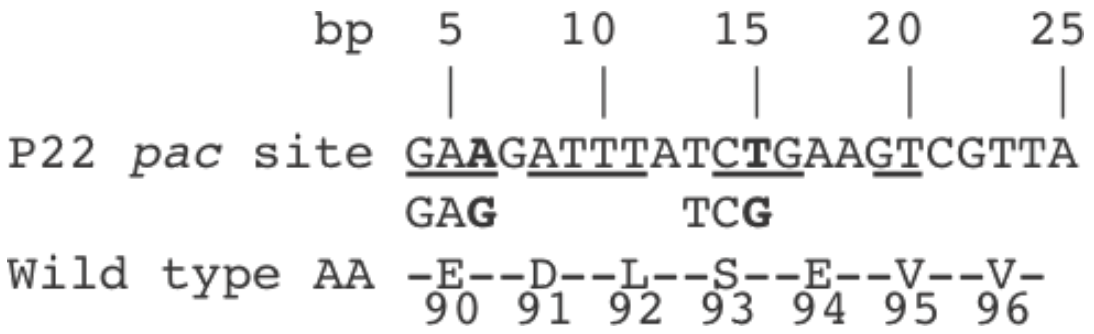


Figure 4.2 Nucleotide changes that inactivate the P22 *pac* site but do not alter the amino acid sequence of TerS.

The P22 *pac* site sequence is shown with bp numbering as in Accession No. BK000583 above. Below, two A6G and T15G synonymous codon changes that inactivate the *pac* site are shown with the TerS AA sequence encoded by this region is shown at the bottom.

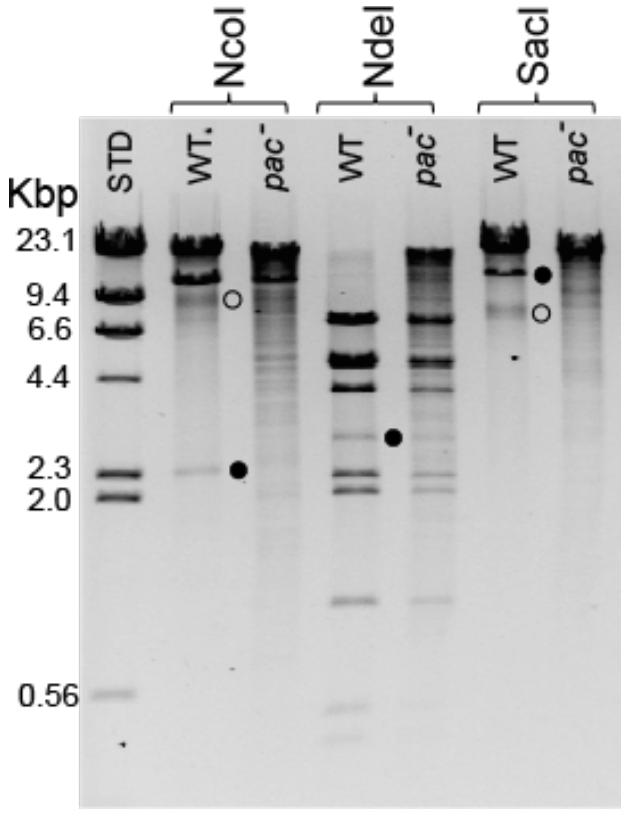


Figure 4.3 Restriction analysis of DNA from virion-like particles produced after induction of a *pac*-defective prophage.

Virion-like particles were isolated 6 hr after induction of *Salmonella* strains UB-1790 (wild type (WT) prophage) and UB-1954 (*pac*⁻ A6G/T15G prophage), and equal amounts of DNA were cleaved by NcoI restriction endonuclease. The resulting fragments were separated in a 1.0% agarose electrophoresis gel, and the gel was stained with ethidium bromide. The black circles to the right of each WT lane marks the position of the 2320, 2789, and 12850 bp NcoI, NdeI, and SacI pac fragments, respectively. The open circle to the right of the WT NcoI and SacI lanes marks the positions of the imprecise right end fragments of the first headfuls of DNA whose packaging initiated at *pac*; the NdeI right end fragment is shorter and so is too diffuse to be seen in the gel. The right lane contains HindIII cleaved phage lambda DNA.

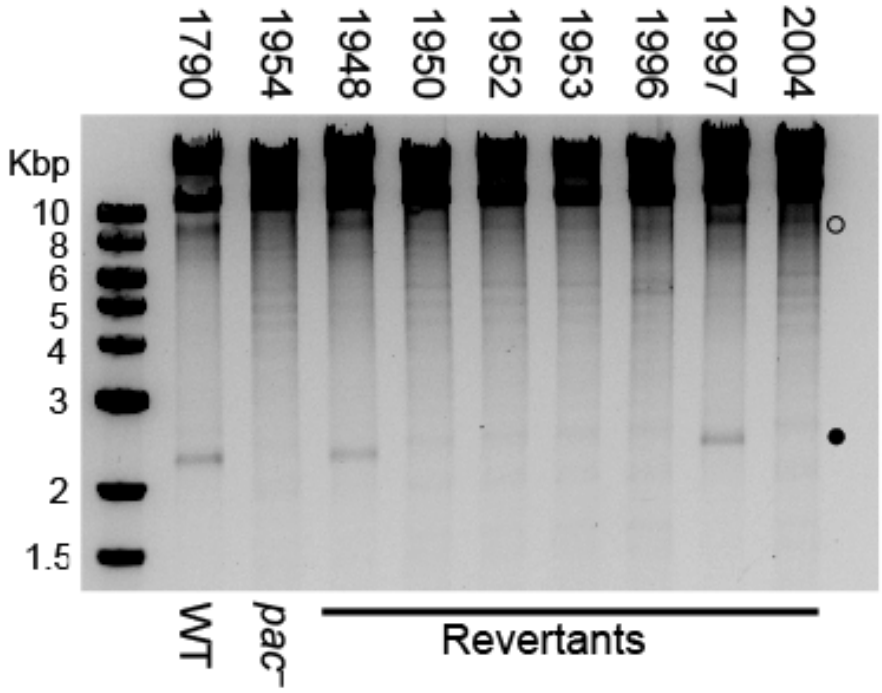
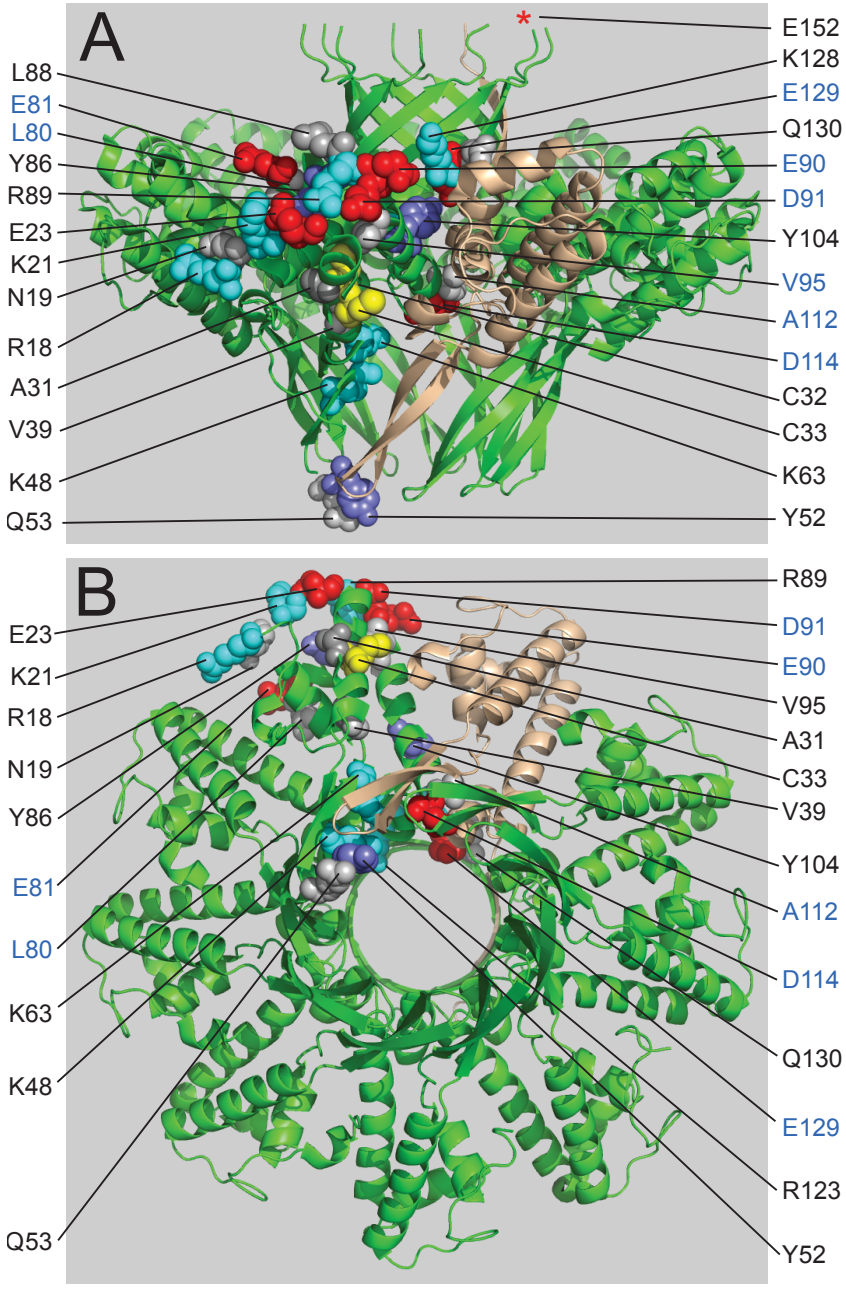


Figure 4.4 Pac fragments of *pac*⁻ revertant mutants

Virion-like particles were isolated 6 hr after induction of *Salmonella* strains UB-1790 (WT prophage) and UB-1954 (*pac* A6G/T15G prophage), and equal amounts of DNA were cleaved by NcoI restriction endonuclease. The resulting fragments were separated in a 1.0% agarose electrophoresis gel, and the gel was stained with ethidium bromide. The black circle on the right marks the position of the 2320 bp NcoI pac fragment. The open circle on the right marks the position of the imprecise right end fragment of the first headful of DNA whose packaging initiated at *pac*; its position is partially obscured by the smear of host DNA, but can be seen in UB-1790, -1948, and -1997. A scale in Kbp is given on the left.

Figure 4.5 Locations of P22 TerS amino acids.

Side (**A**) and bottom (**B**) views of a ribbon diagram of the P22 TerS nonamer are shown, with eight green subunits and one tan subunit. The AA residues that are mentioned in the text are indicated in one green subunit (black labels, AAs changed in mutations isolated in this study; blue labels, AAs changed in previously known high transduction mutants (Casjens et al., 1992); some of the latter were also changed in mutations isolated in this study. These AAs are highlighted as sphere depictions that are colored as follows: red, negatively charged; blue, positively charged; gray, uncharged and aliphatic hydrophobic; yellow, cysteine; purple, aromatic. The red asterisk marks an AA in the portion of P22 TerS that was not present in the x-ray structure, presumably because of its flexibility. The depiction was created with MacPyMol (Copyright Schrodinger, LLC) and Protein Data Bank P22 TerS ID code PDB 3P9A.


A
L88
E81
L80
Y86
R89
E23
K21
N19
R18
A31
V39
K48
Q53
*
E152
K128
E129
Q130
E90
D91
Y104
V95
A112
D114
C32
C33
K63
Y52
B
E23
K21
R18
N19
Y86
E81
L80
K63
K48
Q53
R89
D91
E90
V95
A31
C33
V39
Y104
A112
D114
Q130
E129
R123
Y52

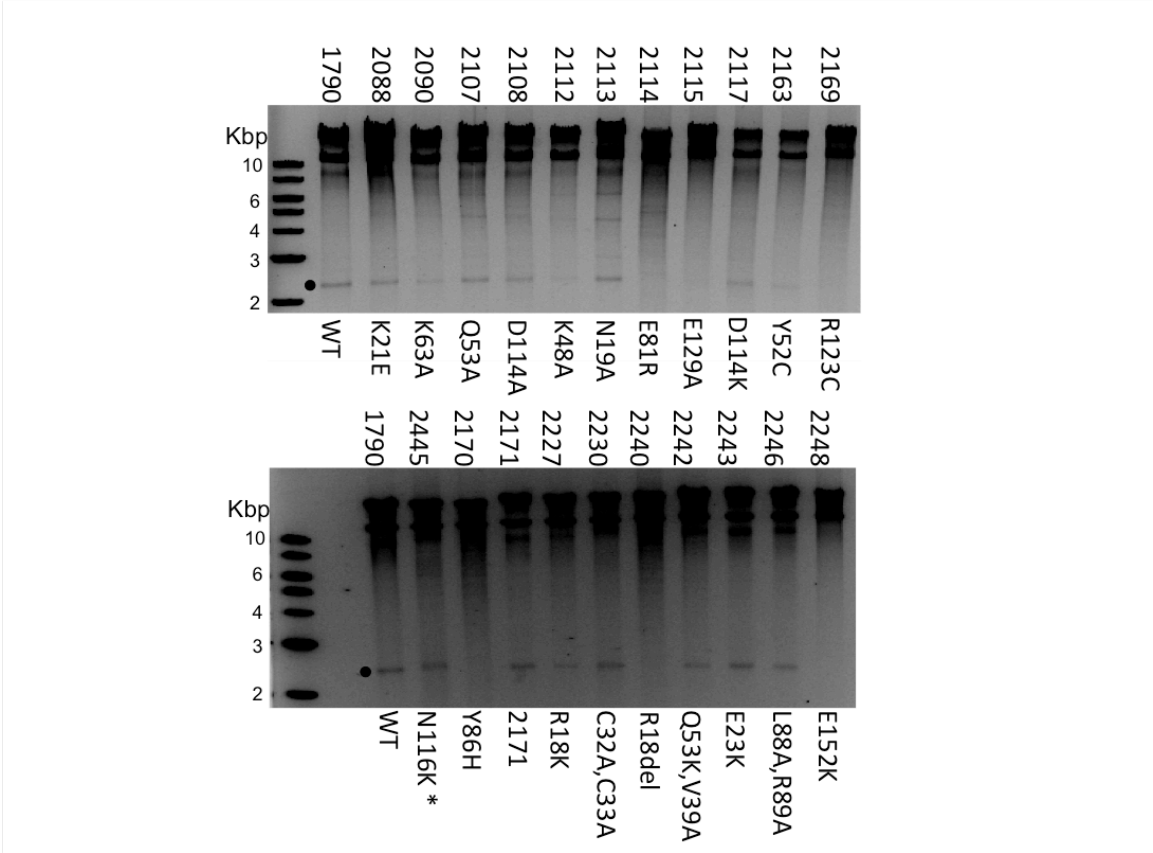


Figure 4.6 Pac fragments of TerS directed point mutants.

The *Salmonella* strains harboring P22 prophages that carry the point mutant TerS proteins in Table 4.3 were induced to lytic growth and virions were isolated 6 hr after induction. Equal amounts of DNA were cleaved by NcoI restriction endonuclease. The resulting fragments were separated in a 1.0% agarose electrophoresis gel, and the gel was stained with ethidium bromide. The strain "UB-" numbers are given above each lane and the TerS mutation is given below. The black circle on the left marks the position of the 2320 bp NcoI pac fragment. A scale in Kbp is shown on the left. The asterisk marks the *terL* (gene *2*) mutation; this strain UB-2445 prophage has a wild type *pac* site.

CHAPTER 5

LOCATION AND QUANTIFICATION OF EJECTION PROTEINS IN BACTERIOPHAGE P22

Abstract

The P22 virion has a T=7 icosahedrally symmetric protein shell with a portal protein dodecamer at the five-fold vertex to which the tail is attached. Extending inwards along the tail axis is a 20 nm-long helical barrel formed by the C-terminal domains of the portal protein subunits. In addition to the densely packed genome, the capsid contains three "ejection proteins" (gp7, gp16, gp20) that are destined to exit from the tightly sealed capsid during the process of delivery of DNA into a target cell during injection. We determined the number of molecules of the ejection proteins present in the virion by quantitative SDS-polyacrylamide gel electrophoresis to be approximately twelve molecules per virion of both gp16 and gp7 and 30 molecules of gp20. In addition, we showed that when the ejection proteins are missing, a longer DNA molecule is packaged in the virion. We thus show for the first time that the ejection proteins reside inside the phage capsid in a location that can be occupied by DNA when they are absent.

Introduction

Although Hershey and Chase (1952) used the fact that phage T2 virion DNA and not "the bulk of the sulfur-containing protein" is injected into target cells to identify DNA as the genetic material, we now know that phage virions also contain proteins that are expelled from the virion into the cell during infection, and they have functions in the cell that are needed for the infection to be productive (Casjens and Molineux, 2012; Mullaney and Black, 1998). In the P22-like phages, these proteins are referred to as the "ejection proteins" because they are essential for proper DNA ejection into the target cell (Israel, 1977). There are three P22 ejection proteins: gp7, gp16, and gp20 (the products of genes *7*, *16*, and *20*, respectively). These proteins are not required for virion assembly, but are *absolutely necessary* for delivery of the viral DNA from the virion into the host cell (King et al., 1973; Poteete and King, 1977).

The P22 ejection proteins are recruited to the procapsid before DNA is packaged (Botstein et al., 1973), at least in part with the aid of the scaffolding protein, a protein required for procapsid assembly but which is completely absent from the completed virion (Greene and King,

1994; Weigele et al., 2005). Thus, the ejection proteins are thought to be localized to the phage head, but their precise location is unknown. They are present in empty virions after DNA has been artificially released (Botstein et al., 1973), so it is unlikely that they are bound to DNA in the virion. Cryo-EM reconstructions of the capsid and virion, and atomic structures of several virion proteins have given a detailed ~8 Å resolution picture of the structure of the virion (Chang et al., 2006; Tang et al., 2011). Although rather precise locations within this structure are known for all other virion proteins, the ejection proteins remain unaccounted for since the other proteins occupy essentially all of the protein electron density of the structure. Thus, the ejection proteins either occupy many alternate locations or they are so flexible that they do not show up in such reconstructions that are made by superimposing many individual virion images. The detailed roles that the P22 ejection proteins play in mediating DNA injection are also not known. It has been postulated that they form a conduit through the cell's outer membrane, peptidoglycan cell wall, and inner membrane that allows passage of the DNA from the virion into the cell cytoplasm (Casjens and Molineux, 2012; Hu et al., 2013). Hoffman and Levine (1975a, b) showed that gp16 from one P22 infecting virion can help DNA release from a particle that lacks gp16 and concluded that gp16 is diffusible after its release from the virion.

The ejection proteins exit the capsid with the DNA during normal delivery of DNA into target cells, and it seems reasonable that they leave through the portal protein channel. The sizes of the gp7, gp16, and gp20 are 21.1, 64.4, and 50.1 kDa, respectively, and it is not known if they might need to be unfolded to leave the virion. The diameter of the portal vertex channel varies between 35 and 75 Å (Olia et al., 2011), and fitting nucleic acid or unfolded peptide chains through it appears quite possible. However, fitting both a peptide chain and nucleic acid though this channel, especially if peptide were folded, would seem much more difficult. A recent report indicates that under nonphysiological conditions, the P22 ejection proteins can be released from purified virions before the DNA is released, and the authors suggest that they therefore could leave the virion before the DNA during a normal infection (Jin et al., 2015). This result suggests that the ejection proteins may be localized in or very near the portal channel.

It is possible that the ejection proteins are localized in the same interior virion space as the DNA, inside the portal protein channel in space not normally taken by DNA, or even on the outside of the viral particle organized in such a way as to be unavailable to proteases (protease treatment of intact virions did not cause degradation of any ejection protein; S. Casjens, unpublished results). Here we show that the ejection proteins occupy internal space that can be taken up by DNA when they are missing, and we accurately measure the number of each of the ejection protein molecules in P22 virions.

Quantification of the ejection proteins in P22 virions

The number of ejection proteins associated with P22 virions has been only very approximately estimated (Casjens and King, 1974). It has also been postulated that the three ejection proteins assemble into procapsids completely independently of one another (Botstein et al., 1973), since a nonsense mutation in any one of the three genes does not block incorporation of the other two. However, Adhikari and Berget (1993) reported that the N-terminal 314 amino acid fragment produced by the gene *20* mutation *am*H1025 under nonsuppressing conditions is incorporated into virions, and the 154 amino acid fragment of gene *7 am*H1375 has also been shown to be incorporated in virions (E. Gilcrease and S. Casjens, unpublished results). The *amber* mutations originally used in the study of the incorporation of the ejection proteins were assumed to be null mutations. This complicates the conclusion that the ejection proteins are assembled independently of one another, since the N-terminal fragment of one protein could aid in the assembly of a second protein. In order to bring clarity to this issue, we constructed complete deletions of each ejection protein gene and a deletion that removes all three genes.

We used quantitative sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) to determine the number of molecules of gp7, gp16, and gp20 per virion. Virion-like particles completely missing each of the three individual ejection proteins and particles missing all three were purified as described in the Methods. Particles containing *amber* mutant ejection protein genes were prepared after infection of a nonsuppressing host by phage with *amber* mutations in

each of the three genes, and after induction of prophages with each of the genes individually deleted or all three genes deleted. The construction of these mutant phages is described in the Methods. The structural proteins of these purified virion-like particles were separated by SDS-PAGE (10.0% acrylamide to maximally separate gp1, gp16, and gp20 and 12.5% to separate gp4 and gp7). The gels were stained with Coomasie Brilliant Blue, and gel images were digitized with a Bio-Rad GelDoc image capture apparatus. The intensity of dye associated with each protein band was quantified using ImagJ 1.48v (http://imagej.nih.gov/ij) and normalized to the cognate ejection protein molecular weight and to the known amounts of gp1 (portal protein) or gp4 (a tail protein) in the virion. The latter two proteins are known to be present in virions in exactly twelve copies each (Tang et al., 2011).

Quantitation of these proteins in the different ejection mutant virus-like particles (a complete deletion of all three genes, *7* deletion, *20* deletion, and *16* deletion) was performed to determine the relative abundance of each ejection protein when one of the other ejection proteins is completely absent. In addition, particles made by an *amber* mutant in each ejection protein gene was examined to attempt to determine whether the N-terminal *amber* fragments affect assembly of the other ejection proteins. A typical gel trace is shown in Figure 5.1, and the measured amount of each individual ejection protein is given in Table 5.1. We find that gp16 and gp7 are present as 12±2.7 molecules per wild type virion, the same copy number (within our experimental error, see Table 5.1) as the portal protein. The gp20 protein is somewhat more abundant at about 30 copies in the wild type particle. Thus, there are substantial amounts of ejection proteins present in the wild type virion - 1.50 MDa of gp20, 0.77 MDa of gp16, and 0.25 MDa of gp7 for a total of 2.52 MDa.

These results show for the first time that stable virus particles are assembled apparently normally by the deletion mutant that completely lacks all three ejection proteins. Furthermore, analysis of the particles made by the single gene deletions shows that the ejection proteins can assemble to the virion independently of one another. Although no ejection protein gene is absolutely required for the assembly of the others, removal of an ejection protein gene can

substantially affect the amounts of the other ejection proteins in the virion. The deletion of gene *7* results in significantly less gp20 and gp16, deletion of gene *16* lowers the amount of gp20 modestly but does not affect the amount of gp7, and deletion of gene *20* lowers gp16 modestly but does not affect the amount of gp7. The *amber* and deletion mutations in the same gene generally contain about the same levels of the non-mutant ejection proteins, with one exception; *amber 7* mutant particles have significantly more gp20 than particles with a completely deleted gene *7* (15.5±1.9 vs. 4.7±1.3 molecules/virion, respectively), so it is possible that the N-terminal gp7 fragment causes more efficient assembly of gp20.

Ejection proteins compete with DNA for space inside the capsid

The length of dsDNA packaged inside the P22 capsid is determined by a headful sensing mechanism. Therefore, DNA packaging length is determined by how much space is available in the capsid (Casjens and Hayden, 1988; Casjens et al., 1992; Tye et al., 1974; Weigele et al., 2005). If the ejection proteins are inside the capsid and occupy space that could be taken by DNA, then the absence of these proteins should cause the virus particle to package a correspondingly longer strand of DNA. We therefore analyzed the Δ7, Δ16, Δ20, and triΔ ejection protein mutants to determine whether the length of DNA packaged is affected by these proteins' absence. P22 packaging series begin near the *pac* site (Casjens and Hayden, 1988; Jackson et al., 1978; Wu et al., 2002). Our electrophoretic measurement of the length of the whole, uncleaved P22 chromosome is 43500 bp with an estimated uncertainty of about ±5% (±1000 bp). For the current purpose, we wished to determine the chromosome length considerably more accurately than this, so we determined the position of the right-end headful cleavage site for the first packaging event in a series by measuring the length of right-end restriction fragments. The length of the packaged DNA can then be calculated from the known P22 genome sequence and the size of the right-end restriction fragment (diagrammed in Figure 5.2B). This fragment is seen as a diffuse band in electrophoresis gels because the headful measuring sensor is imprecise (Casjens and Hayden, 1988). Using enzymes that cleave near the right-end of the genome to

produce a relatively short several kbp right-end fragments allows measurements of the average length of DNA packaged to about ±100 bp.

Enzymes that cleave P22 once or a small number of times and include a cleavage site near the right-end of the genome sequence are suitable for such an analysis, and Figure 5.2B shows the locations of the XhoI and SpeI cleavage sites used. Not surprisingly, first headful right-end fragment length measurements of the "wild type" P22 (strain UC-937) DNA indicate that it packages 43500±775 bp (Figure 5.2, Table 5.2), where the uncertainty is the width of the diffuse right-end fragment, not error in the measurement. This intra-virion chromosome DNA length is within experimental error of the previously reported value for truly wild type P22 (Casjens and Hayden, 1988). Parallel determinations of first headful cleavage locations for phages Δ7, Δ16, Δ20, and triΔ gave packaged DNA lengths of 45200, 44000, 43700, and 45100, about 1700, 500, 200, and 1600 bp longer than wild type, respectively (Figure 5.2B, Table 5.2). Analysis of P22 phages with nonsense mutations in the ejection protein genes gave similar values (Table 5.2).

We calculate that the ejection proteins in the P22 virion should occupy about 3.0×10^6 Å^3 assuming an average value of 1.2 Å^3/Daltons of protein, a value that ignores protein hydration and therefore is likely an underestimate of the space they actually require in the virion. The internal volume of the virion is 9.7×10^7 Å^3 (approximated by a sphere of inner radius ~285 Å (Tang et al., 2011) and subtracting the estimated portal barrel volume). The 43500 bp packaged chromosome thus implies that one hydrated bp occupies ~2200 Å^3 in the head. All the single ejection protein mutants have longer than wild type DNAs, but analysis of the single mutants is complicated by the fact that absence of one protein affects the amounts of the others present (above); however, the space occupied by all the ejection proteins (without their hydration) is equivalent to ~1400 bp of DNA at its normal hydrated intravirion density. Thus, the volume of the 1600-1700 bp of extra DNA present in the triΔ virions which have no ejection proteins is about 3.6×10^6 Å^3, which is quite comparable to the 3.0×10^6 Å^3 volume of the missing ejection proteins. We conclude that the ejection proteins are inside the phage capsid in a location that can be occupied by DNA when they are absent.

Discussion

We determined that the mass of P22 ejection proteins in wild type virions is about 2.5 MDa, which corresponds to a few percent of the internal volume of the virion. Analysis of the virion-like particles made by mutants in which one or all of the ejection protein genes are deleted showed conclusively for the first time that (1) the ejection proteins can assemble into procapsids independently from one another, and (2) that stable virion-like particles are produced normally in the complete absence of all ejection proteins. The latter refutes an earlier report that gp16 is involved in procapsid assembly (Thomas and Prevelige, 1991). On the other hand, we find that when one ejection protein is missing, another can be present in lower amounts (Table 5.1). This observation is not currently understood, but for example, the gp16 and gp20 ejection proteins may have affinity for one another (C. Teschke, personal communication) and assembly of the complex into the procapsid could be more efficient than assembly of the individual proteins by themselves. Alternatively, these results could in theory be accounted for by differences in the expression level of the intact ejection protein genes by the various mutant genomes. However, there is no experimental support for this, and it seems unlikely since each of the deletion mutants was designed not to disrupt the translation initiation regions of the downstream gene. In addition, nonsense mutant polarity on the expression of downstream genes is suppressed by the late operon transcription activator, so such polarity is unlikely to affect expression in the *amber* mutant phages.

We also measured the lengths of DNAs packaged in P22 virion-like particles when ejection proteins are missing and found that longer DNAs are packaged in their absence. We conclude that the P22 ejection proteins are located inside the virion, where they occupy space that is also accessible to DNA. This is in agreement with the observation that the ejection proteins are not degraded when the mature virions are exposed to various proteases (S. Casjens, unpublished results). It also fits the (very reasonable but unproven) model of ejection protein action in which they are released from the virion through the portal vertex channel. The volume of the total ejection protein mass in the virion corresponds rather well with the volume of DNA that replaces

them when they are absent. However, the volume of each individual ejection protein does not correlate as accurately with the volume of extra DNA present when that protein is missing. For example, when the gene *20* protein is missing from the virion, the volume of 'extra' DNA is less than expected. The reason(s) for this are not known, but it is possible that the conformation or positioning of other ejection proteins in the virion could be altered when another ejection protein is not present.

Why is there essentially no electron density available for the ejection proteins in the 7.8 Å P22 virion structure determined by Tang *et al*. (2011)? If they were identically arranged in every individual virion, their mass is such that they should be unambiguously present in the electron density map of the particle. There are at least three possible explanations as follows: (1) The ejection proteins are "floating" and thus can move in the virion interior; (2) in a given particle, they occupy only a subset of possible precisely located sites, so that during the cryo-EM reconstruction process, they occupy only a fraction of any individual site and so are "averaged out" to below the electron density cutoff used to display the particle; and (3) the ejection proteins are each bound to unique and precise positions in the virion, but the bulk of their polypeptide chains are unfolded or otherwise very flexible so they do not occupy exactly the same location in every particle. One could imagine that, if the latter were true and a small part of each protein is responsible for tethering it to a unique site in the virion and this small part is too small to identify even in the 7.8 Å resolution virion structure, the ejection proteins would not be seen even though they occupy specific locations in the virion. Our preliminary collaborative results with the Dr. Alasdair Steven laboratory (NIH, Bethesda, MD) using bubblegram technology (Cheng et al., 2014) have shown that the P22 ejection proteins very likely reside at a particular intravirion location on portal protein barrel, suggesting that explanation 3 above may be correct. This agrees with the ideas that the ejection proteins may have to be unfolded to pass through the portal channel during ejection and that they should reside close to the portal channel so they can exit through the channel before the DNA exits the virion.

Materials and methods

Ejection protein deficient mutants were constructed using *galK* recombineering methods which are described in Warming et al. (2005) and the Appendix of this work. The pKD46 plasmid (Datsenko and Wanner, 2000) was present to promote recombinational replacement, but was removed by growth at 42°C before prophage induction. *Salmonella enterica* serovar Typhimurium LT2 strain UB-2235 is *galK*⁻ (a tetracycline resistance cassette replaces the *galK* gene), lacks the LT2 Fels-1, Fels-2, Gifsy-1, and Gifsy-2 prophages, and carries a P22 UC-937 prophage (genotype P22 *c1*-7, *13*⁻*am*H101, Δ*sieA*-1, Δorf25::CamR-EG1; its detailed construction will be described in a future publication; S. Casjens, J. Leavitt and E. Gilcrease., unpublished). The *c1*-7 mutation greatly lowers the frequency of lysogen formation so that the phage grows lytically in liquid culture, *13*⁻*am*H101 blocks normal lysis and gives higher phage yield, Δ*sieA*-1 removes the Mnt repressor and its binding sites that lower tailspike protein production, and Δorf25::CamR-EG1 is a chloramphenicol resistance cassette that replaces nonessential DNA between genes *15* and *3.* The *galK* cassette was amplified from plasmid pGalK (Warming et al., 2005) with sets of primers that allow recombinational insertion of the amplified DNA into the P22 UC-937 prophage of UB-2235 to generate strains UB-2276, UB-2274, UB-2272, and UB-2278 in which the *galK* cassette replaces gene *7, 16,* or *20* or all three genes, respectively. Synthetic oligonucleotides were designed so that recombinational replacement of these *galK* insertions generate prophages missing essentially all of each of the ejection protein genes (strains UB-2285, UB-2288, UB-2289, respectively). The single gene deletions each remove the entire coding region of each gene except for the 3'-terminal 60 bp of the coding region. This was done in order to not interfere with the translational expression of the downstream gene. To generate isogenic strains that carry nonsense mutations in the ejection protein genes, several kbp DNA regions were PCR amplified from P22 strains that contain *amber* mutations 7⁻*am*H1375, 16⁻*am*N121, and 20⁻*am*N20 (Botstein et al., 1973; Poteete and King, 1977). These DNAs were used to replace the *galK* cassettes in UB-2276, UB-2274, and UB-2272 to generate versions of strain UB-2235 that carry these alleles (UB-2387, UB 2350, and UB-2366, respectively). The structure of each of the above phage

genomes was confirmed by DNA sequencing of the modified regions.

Purified phage particles were prepared as follows: prophages were induced from exponentially growing 37°C L broth cultures at $2x10^8$ cells/ml by addition of 0.5 ug/ml Mitomycin C or 1.5 ug/ml carbodox and continued shaking for 5-6 hours. Cells were concentrated by centrifugation and lysed by shaking with chloroform. Virions were purified, after a cell debris-removal low speed centrifugation, by two successive CsCl step gradient centrifugations (Earnshaw et al., 1976) and dialysis against 1 mM $MgCl_2$, 10 mM TrisCl, pH 7.4 buffer.

References

Adhikari, P., Berget, P.B., 1993. Sequence of a DNA injection gene from *Salmonella typhimurium* phage P22. Nucleic Acids Res 21, 1499.

Botstein, D., Waddell, C.H., King, J., 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. I. Genes, proteins, structures and DNA maturation. J Mol Biol 80, 669-695.

Casjens, S., Hayden, M., 1988. Analysis *in vivo* of the bacteriophage P22 headful nuclease. J Mol Biol 199, 467-474.

Casjens, S., King, J., 1974. P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. J Supramol Struct 2, 202-224.

Casjens, S., Wyckoff, E., Hayden, M., Sampson, L., Eppler, K., Randall, S., Moreno, E.T., Serwer, P., 1992. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J Mol Biol 224, 1055-1074.

Casjens, S.R., Molineux, I.J., 2012. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol 726, 143-179.

Chang, J., Weigele, P., King, J., Chiu, W., Jiang, W., 2006. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. Structure 14, 1073-1082.

Cheng, N., Wu, W., Watts, N.R., Steven, A.C., 2014. Exploiting radiation damage to map proteins in nucleoprotein complexes: the internal structure of bacteriophage T7. J Struct Biol 185, 250-256.

Conlin, C.A., Miller, C.G., 1992. Cloning and nucleotide sequence of *opdA*, the gene encoding oligopeptidase A in *Salmonella typhimurium*. J Bacteriol 174, 1631-1640.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97, 6640-6645.

Earnshaw, W., Casjens, S., Harrison, S.C., 1976. Assembly of the head of bacteriophage P22: x-

ray diffraction from heads, proheads and related structures. J Mol Biol 104, 387-410.

Greene, B., King, J., 1994. Binding of scaffolding subunits within the P22 procapsid lattice. Virology 205, 188-197.

Hershey, A.D., Chase, M., 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 36, 39-56.

Hoffman, B., Levine, M., 1975a. Bacteriophage P22 virion protein which performs an essential early function. I. Analysis of 16-ts mutants. J Virol 16, 1536-1546.

Hoffman, B., Levine, M., 1975b. Bacteriophage P22 virion protein which performs an essential early function. II. Characterization of the gene 16 function. J Virol 16, 1547-1559.

Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339, 576-579.

Israel, V., 1977. E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. J Virol 23, 91-97.

Jackson, E.N., Jackson, D.A., Deans, R.J., 1978. EcoRI analysis of bacteriophage P22 DNA packaging. J Mol Biol 118, 365-388.

Jin, Y., Sdao, S.M., Dover, J.A., Porcek, N.B., Knobler, C.M., Gelbart, W.M., Parent, K.N., 2015. Bacteriophage P22 ejects all of its internal proteins before its genome. Virology 485, 128-134.

King, J., Botstein, D., Casjens, S., Earnshaw, W., Harrison, S., Lenk, E., 1976. Structure and assembly of the capsid of bacteriophage P22. Philos Trans R Soc Lond B Biol Sci 276, 37-49.

King, J., Lenk, E.V., Botstein, D., 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. J Mol Biol 80, 697-731.

Mullaney, J.M., Black, L.W., 1998. Activity of foreign proteins targeted within the bacteriophage T4 head and prohead: implications for packaged DNA structure. J Mol Biol 283, 913-929.

Olia, A.S., Prevelige, P.E., Jr., Johnson, J.E., Cingolani, G., 2011. Three-dimensional structure of a viral genome-delivery portal vertex. Nat Struct Mol Biol 18, 597-603.

Poteete, A.R., King, J., 1977. Functions of two new genes in *Salmonella* phage P22 assembly. Virology 76, 725-739.

Tang, J., Lander, G.C., Olia, A.S., Li, R., Casjens, S., Prevelige, P., Jr., Cingolani, G., Baker, T.S., Johnson, J.E., 2011. Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in p22. Structure 19, 496-502.

Thomas, D., Prevelige, P., Jr., 1991. A pilot protein participates in the initiation of P22 procapsid assembly. Virology 182, 673-681.

Tye, B.K., Chan, R.K., Botstein, D., 1974. Packaging of an oversize transducing genome by *Salmonella* phage P22. J Mol Biol 85, 485-500.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., Copeland, N.G., 2005. Simple and highly efficient BAC recombineering using *galK* selection. Nucleic Acids Res 33, e36.

Weigele, P.R., Sampson, L., Winn-Stapley, D., Casjens, S.R., 2005. Molecular genetics of bacteriophage P22 scaffolding protein's functional domains. J Mol Biol 348, 831-844.

Wu, H., Sampson, L., Parr, R., Casjens, S., 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol Microbiol 45, 1631-1646.

Table 5.1

Numbers of P22 ejection protein molecules per virion

Phage [a]	gp20 (472 aa)	gp16 (609 aa)	gp7 (209 aa) [c]	Total MDa [d]
Wild type	31.6 ± 3.0 [b]	11.9 ± 2.4 [b]	11.3 ± 2.7 [b]	2.52 (1.00)
Δ20	0	6.0 ± 0.8	9.9 ± 0.5	0.60 (0.24)
amber 20	0	3.7 ± 1.9	11.2 ± 2.8	0.49 (0.19)
Δ16	22.4 ± 2.2	0	9.0 ± 1.3	1.29 (0.51)
amber 16	29.8 ± 2.0	0	9.8 ± 0.8	1.71 (0.68)
Δ7	4.7 ± 1.3	5.6 ± 2.0	0	0.64 (0.25)
amber 7	15.5 ± 1.9	8.3 ± 0.5	0	1.27 (0.50)

a. Phages are all isogenic derivatives of P22 UC-937 (see Appendix and Methods).
b. Ranges of values determined for the numbers of protein subunits per P22 virion in several experiments are indicated. These values were determined as follows: SDS-PAGE gels of purified virions were stained with Coomassie brilliant blue (Figure 5.1), scanned, and the gp1, gp4, gp7, gp16, and gp20 band intensities were quantified with ImageJ 1.48v (http://imagej.nih.gov/ij). Peak areas were normalized to each subunit's molecular weight, and these values were in turn normalized relative to a protein that has a precisely known copy number and which migrates at a nearby position in the gel; gp16 and gp20 were normalized to the twelve portal (gp1) subunits and gp7 was normalized to the twelve gp4 subunits.
c. A host protease removes the N-terminal 20 amino acids of the 229 amino acid gp7 subunit before it assembles into virion (Conlin and Miller, 1992).
d. Total measured mass of ejection proteins in particles (fraction of wild type in parentheses).

Table 5.2

Lengths of DNAs packaged by ejection protein mutants

Phage	Genome length[a]	Location of right-end headful cleavage[b]	Length of DNA packaged[c]
WT (UC-937)	39521	3950	43500
Δ20	38161	5500	43700
amber 20	39521	4500	44000
Δ16	37751	6150	44000
amber 16	39521	4600	44100
Δ7	38890	6560	45200
amber 7	39521	5300	44800
triΔ	36867	8200	45100

a. Determined from the complete sequence of P22 wild type (accession number TPA:BK000583) and P22 UC-937 (our unpublished results).
b. Location on the sequence of the phage DNA (where bp numbering begins at the same point as in GenBank annotation accession number TPA:BK000583). The first headful cleavage site is calculated from the measured length of the first headful right-end fragment and the location of the restriction site whose cleavage generated that fragment. The range of several determinations of the size of the center of the right-end XhoI and SpeI restriction fragments was about ±100 bp in the 2-6 kbp range and increased to about ±250 bp above that.
c. Length of DNA packaged = genome sequence length + number of bp between bp 1 of the P22 genome and the center of the range of headful cleavage sites of the first headful's right-end fragment (see Figure 5.2B).

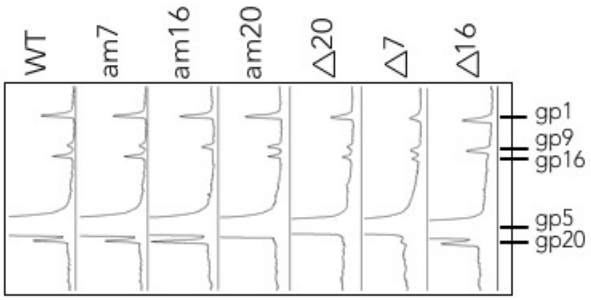


Figure 5.1 Quantitation of E-proteins in P22 virions.

A 10% acrylamide SDS-PAGE gel with purified virion-like particles of the genotypes indicated above was stained with Coomassie brilliant blue. Protein migration is downward. The gel was imaged with a BIORAD Gel Doc apparatus, and the image was converted to traces and the gp1, gp16, and gp20 peaks quantified with IMAGEJ. The scans shown integrated bands over the whole lanes (integrations of narrower lane sections down the center of the lanes gave indistinguishable results; not shown). The gp4 and gp7 bands were quantified in a parallel 12.5% acrylamide gel of the same samples (not shown). The resulting copy numbers for the ejection proteins are given in Table 5.1.

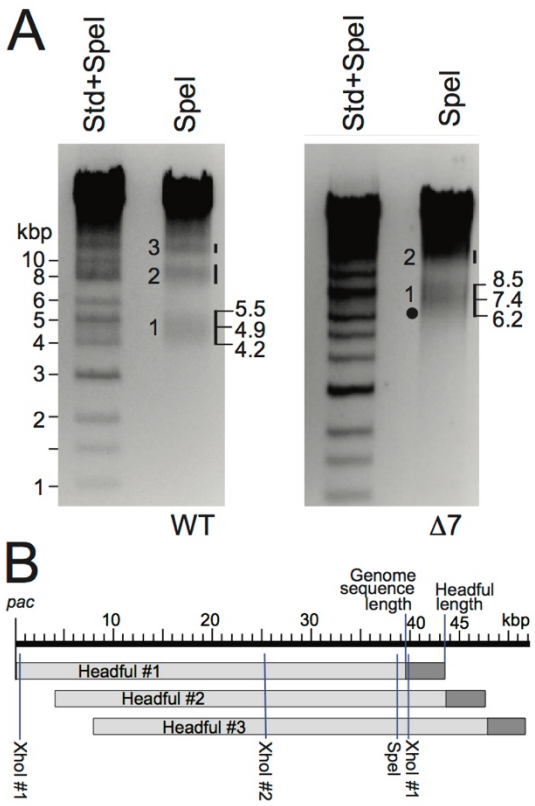


Figure 5.2 Particles without ejection protein gp7 contain shorter DNA molecules.

Panel A. DNA from the particles indicated below was cleaved with restriction enzyme SpeI and separated in a 0.6% agarose electrophoresis gel; molecular weight standard DNA fragments (Std) were mixed with an identical SpeI digest to avoid gel differences due to the large amount of phage DNA loaded. Sizes of the standards are shown on the right. To the left of the SpeI lanes, black vertical bars mark the "fuzzy" right-end fragment bands with the series headful numbers. The sizes in kbp of the upper and lower boundaries and average size of the first headful bands are indicated of the right of each panel. The black circle indicates the expected position of the center of first headful right-end band if the packaged DNA were wild type length (note the genome sequences of wild type and are not the same so their right end fragment lengths *per se* do not accurately reflect the amount of DNA packaged). **Panel B**. The first three packaged DNA molecules of a sequential headful series are shown as bars below the P22 UC-0937 genome kbp scale (its sequence is 39521 bp long); light gray denotes the first 39521 bp inserted into the procapsid and the dark gray denotes the terminal redundancy.

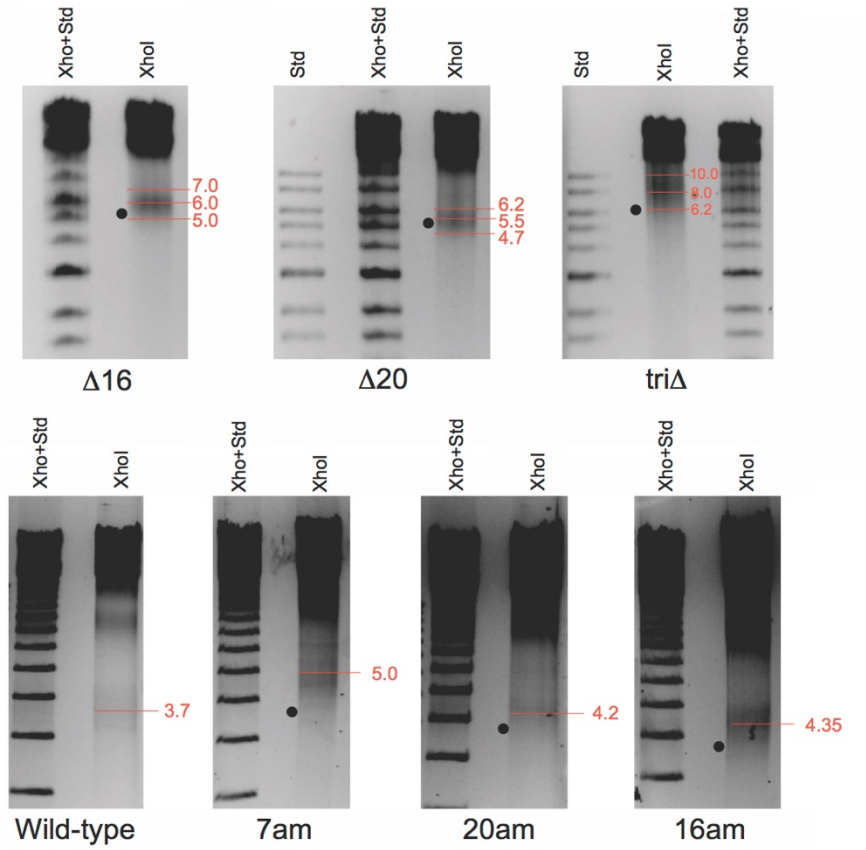


Figure 5.3 Particles without ejection proteins gp7, gp16, and/or gp20 contain shorter DNA molecules.

DNA from the particles indicated below was cleaved with restriction enzyme XhoI and separated in a 0.6% agarose electrophoresis gel; sizes of molecular weight standard DNA fragments are given in Figure 5.2A. The sizes in kbp of the upper and lower boundaries and average size of the first headful, right-end bands are indicated in red on the right of the XhoI lane. The black circles indicate the expected position of the center of first headful's right-end band if the packaged DNA were wild type length. Note that the genome sequences of Δ16, Δ20, and triΔ phages are not the same, since the deleted genes are different sizes; thus, their right end fragment lengths per se do not accurately reflect the amount of DNA packaged. The right end bands were unequivocally identified by showing that their right-end is diffuse (created by the headful packaging cut) and their left end is the precise restriction enzyme cut. The width of the right-end fragment bands appears to be somewhat larger than wild type in the 7*amber*, Δ7, and triΔ mutants. The reason for this is not yet known.

CHAPTER 6

GENERAL CONCLUSIONS

Despite the long history of the study of the dsDNA tailed bacteriophages, several aspects of their lifecycles are still quite poorly understood. In particular, many aspects of the "DNA transactions of virions" - DNA entry into and exit from virions - remain mysterious. A great deal of work remains to be done to obtain a basic understanding of DNA packaging into the virion and of DNA injection into target cells by virions. The work presented here provides insight into both of these processes and points out the unexpected complexity of these processes. The connection between bacteriophage DNA packaging and DNA ejection is a perfect example of the "simple complexity" of nature. Two complex interconnected functions, one of which is in a sense the reverse of the other, must take place using limited genome informational space and protein machinery. The DNA must be replicated, measured, and packaged. Then it must be delivered in a precise way to the correct host. All of this is done by a relatively small number of proteins. In the following paragraphs, a summary of the work on phage P22 and its relative Sf6 is presented. Our most important findings are discussed in the following sections.

DNA packaging

As described in the previous chapters, the dsDNA phages build virions by first assembling a procapsid shell and then using a DNA translocase to pump the DNA into this preformed container. The translocase complex has two subunits: TerL, which harbors the ATPase that powers packaging and the nuclease that creates the DNA ends or termini, and TerS, which is thought to recognize DNA for packaging by interacting with a specific sequence in the viral DNA. The closely related phages P22 and Sf6 were the subject of the studies presented here.

As was described in the Introduction chapter above, although a site called *pac* had been genetically identified *in vivo* for phage P22, *in vitro* recognition of this site by the purified TerS protein has never been achieved. Both P22 and Sf6 TerS proteins bind DNA without regard to any specific sequence i*n vitro*. We therefore chose to study TerS action *in vivo* using genetic approaches. The P22-like *pac* site is located within the TerS gene, and work described in Chapter 2 showed that site is both necessary and sufficient for successful DNA packaging. The availability

of many phage genome sequences has shown that the P22/Sf6-type phage virion assembly functions have undergone many past exchanges within the "P22-like" phage group such that any given late operon is mosaically related to the others. Examination of the exact locations of "mosaic boundaries" (points of past recombination between two divergent members of this phage group) has shown that these boundaries occur at protein domain boundaries rather than gene boundaries. Thus, protein domains define the "exchangeable units" in this system. Using what we now know about this modular nature of bacteriophage genomes, we were able to build "artificial" novel domain combinations and deduce interactive domains from the properties of such hybrid phages. In Chapters 2 and 4, we constructed P22 and Sf6 TerS hybrids of phages using their respective N-terminal domains of TerS. These experiments showed clearly that the N-terminal domains of both these TerS proteins harbor *pac* site specificity and that the C-terminal domains carry specificity determinants for interaction with the TerL subunit. We also carried out experiments that genetically identified the Sf6 *pac* site and found that it has partial similarity to the P22 *pac* site. We also demonstrated for the first time that the *pac* site, and therefore presumably its recognition by TerS protein, is essential for P22 growth in *Salmonella* cells.

It has been generally accepted that in the tailed phages, TerS is responsible for recognition of the DNA to be packaged, and the two most rational hypotheses for this action were either DNA wrapping around the outer rim or DNA threading through the central channel of the ring-shaped TerS oligomer. We therefore carried out an extensive program of mutagenesis of the P22 and Sf6 TerS proteins, with the goal of using *in vivo* phenotypic analysis to determine which part of TerS interacts with the *pac* site during the process of choosing a DNA for packaging (see Chapter 4). Our results indicate that changes in P22 TerS surface amino acids usually do not inactivate the protein's role as an essential component of the DNA packaging machinery, and many different changes in both the inner channel and the outer rim have affects on DNA recognition *in vivo*. It was previously shown (Zhao et al., 2012) that some outer rim mutations of Sf6 TerS block its ability to bind DNA nonspecifically *in vitro.* We recombineered these mutations into a P22-Sf6 hybrid phage that has the DNA recognition TerS domain of Sf6 and tested functionality of these

mutants. Two of these mutants are functional in terms of phage growth, even though their TerS proteins had been shown not to be able to bind DNA *in vitro*. It has been a long-standing central assumption in this field that this nonspecific DNA binding by TerS proteins (observed in a number of different headful packaging phages) is in some way a relevant reflection of the mechanism by which TerS proteins bind viral DNA and choose it for packaging; however, our finding that Sf6 TerS proteins that do not have this ability remain perfectly functional suggests that the observed nonspecific DNA binding may not be relevant to the real *in vivo* DNA packaging process.

Many drastic amino acid side chain charge reversal or charge removal alterations in both channel and rim locations had only relatively "minor" affects on packaging. The simplest interpretation of our findings suggests that direct DNA contact with these residues may not be responsible for *pac* recognition, and neither the outer rim nor the inner channel can be confidently identified as the region responsible for *pac* recognition. Nonetheless, our analysis of *pac* site utilization *in vivo* and generalized transduction by the panel of TerS mutant phages shows that a majority of the changes studied do affect DNA recognition in nonlethal ways. The fact that a large fraction of the TerS mutants transduce host DNA more frequently than wild type suggests that they have a lowered specificity for phage DNA, *i.e.*, recognition for packaging is more promiscuous in the mutant phages.

One of the most surprising findings was the isolation of a mutation in TerL (N116K) that allows terminase function and DNA packaging in the absence of a complete *pac* site. This indicates that TerS is not solely responsible for DNA recognition, and that TerL may play another role in addition to DNA cutting and packaging. This mutation alone (without the null-*pac* site) is a high transduction mutation, and the mutant phage does not use the defective mutant *pac* site, although it does use the native *pac* site when it is available.

DNA ejection

DNA delivery into target cells by tailed phage virions is a process that is much more complex than was originally envisioned. First the virion must bind specifically to cells of the right host

species; this is termed adsorption and is usually mediated by fibrous or spike-shaped proteins at the tip of the tail that bind to a specific host surface macromolecule. Simple binding to this host surface "receptor" is usually reversible and is not sufficient to program DNA release from the virion. After this initial binding to the cell, the distal tip of the tail must in some way "confirm" that the target is correct, presumably by irreversible binding to some cellular feature (sometimes called the "secondary receptor") near or at the surface of the cell's outer membrane. This then triggers DNA release from the virion in a way that directs it through the outer membrane, cell wall, and inner membrane so that it ends up in the cytoplasm of the target cell.

Phage P22 follows this delivery strategy. Its tailspike proteins first bind the *Salmonella* surface polysaccharide. The tailspike carries an endorhamnosidase enzymatic activity that cleaves the polysaccharide, and as the polysaccharide chains are shortened, the virion is brought, tail end first, down to the surface of the bacterium. The mechanism by which the DNA is triggered to release from the P22 virion is unknown; however, the position of the long tail needle suggests that its C-terminal distal tip should be the part of the virion that makes first contact with the surface of the outer membrane of the target cell. It is known that four "minor" virion proteins are then released from the virion. One of these proteins, the gene *26*-encoded tail needle, plugs the DNA passage channel. Its removal allows DNA release (Bauer et al., 2015), and the three "ejection" proteins are required for successful delivery of the virion DNA into the cell's cytoplasm (summarized in the Introduction Chapter 1).

A portion of the work described here was focused on understanding the release of the phage P22 tail needle and the triggering of the ejection process. In work not included in this dissertation, we collaborated with the laboratory of Dr. Alex Evilevitch to show that single amino acid change mutants in the N-terminal base (virion proximal) portion of the needle affect its release upon treatment of the virions with high temperatures (Bauer et al., 2015). In further work, we showed that the C-terminal distal tip of the tail needle also plays a role in the ejection process (Chapter 3). The replacement of the P22 C-terminal domains with parallel domains from related phages Sf6 or HS1 showed that this domain is interchangeable and not species-specific.

In addition, we replaced the needle's C-terminal domain with a completely foreign trimerization domain (called a "foldon"), and this phage was functional although it makes small plaques. The functionality (albeit limited) of the foldon swap phage was a surprise. If, as was predicted, the C-terminal domain is a switch that triggers the DNA release from the virion when it contacts some outer cell surface feature, replacing this domain with a completely different domain was expected to render the particles noninfectious. This, however, was not the case, although infection by this phage was less efficient than wild type. By measuring the rate of K^+ ions from the infected cell, we were able to indirectly measure the rate of DNA entry into the cell, and we found that the rate of DNA entry into the host cell was notably slower than that of wild type phage, or the Sf6 hybrid tail-needle C-terminal domain swap.

Because of its strategic location in the virion and role as the plug that keeps the DNA inside, we continue to entertain the idea that the tail needle contacts a specific secondary receptor that triggers DNA delivery into the cell. The mechanism by which this might occur remains unknown; needle conformational changes could be induced by its tip being submerged into the membrane, the mechanical force generated by the tail needle pushing against the membrane, some specific surface contact in addition to any the C-terminal domain might make, or a combination of these scenarios could participate in the process. The tail needle should therefore be just stable enough to allow efficient release when a proper stimulus is received, yet not be too unstable, or the DNA might be released at the wrong place or the wrong time, an event that would be lethal to the virion. Triggering the release of the tail-needle therefore should be a very tightly controlled process and may be controlled by multiple safeguard switches. We conclude at this time that the C-terminal domain of the tail needle protein contributes to this control, but is not the only factor that regulates the DNA release process. It remains possible that final triggering of DNA release is mediated by specific contact between the needle's middle shaft domain and a cell surface feature.

The subsequent events after the virion is “uncorked” by release of the tail needle are even more mysterious. The current working model in this field is that after the tail plug is released, the ejection proteins leave the virus particle and somehow allow the viral DNA to cross both

membranes and the cell wall. Again, we have studied this with phage P22 and its "ejection protein" products of genes *7, 16,* and *20* (Chapter 5 above). We first achieved a much more precise quantification of the number of ejection proteins that are present in purified virion preparations, 12, 12, and 30 molecules/virion, respectively. However, there was no evidence as to the locations of these three proteins in the virion; it was not even known if they are inside the capsid cavity on the surface of the tail. We therefore showed for the first time that these proteins occupy *internal* space that would normally be used by DNA. The latter was possible because P22 is a headful packaging phage, so if internal proteins are removed, the packaged DNA molecule should be commensurately longer. We constructed deletion mutants missing these proteins singly and missing all three of them, and showed that virion-like particles containing packaged DNA are made by all such mutants. The lengths of the DNAs in these particles were accurately measured and shown to be longer than wild type intra-virion DNA, indicating directly for the first time that the P22 ejection proteins are inside the DNA-containing part of the virion head. In addition, collaborative work is currently underway with the laboratory of Dr. Alasdair Steven that will bring further insight into the location of the ejection proteins inside the virion. We have found that the ejection proteins have a very strong propensity to stimulate hydrogen gas bubble formation (much more than any other P22 virion protein) when bombarded by electrons in the electron microscope, and that bubble formation in P22 virions begins along the tail axis at 25-30% the distance from the portal vertex to the opposite vertex. Thus, at least the bubble stimulating portion of the ejection proteins likely reside at this location in the virion. The dynamic nature of these proteins, and the number of functions and environments in which they complete their functions, has exciting potential for future work.

Broader impacts

The genetic, biochemical, and biophysical research done with model phages such as P22 gives a solid basis understanding the tailed phages in general (the most abundant "organisms" on Earth) as well as for large eukaryotic viruses such as herpesviruses and adenoviruses that

share assembly strategies with the tailed phages. The advantages that a reductionist view of phage provide to biology as a whole cannot be overlooked, and there may be *direct* applications to human well-being that the long and patient work of phage biologists may finally make possible. Since almost the moment of their discovery, phage therapy was viewed as a potential therapeutic for bacterial infections. With a more refined understanding of phage structure and function, and the genetic tools now available to us, this might finally (after more than 100 years since phage discovery) be viewed as a very real possibility.

References

Bauer, D.W., Li, D., Huffman, J., Homa, F.L., Wilson, K., Leavitt, J.C., Casjens, S.R., Baines, J., Evilevitch, A., 2015. Exploring the Balance between DNA Pressure and Capsid Stability in Herpesviruses and Phages. J Virol 89, 9288-9298.

Zhao, H., Kamau, Y.N., Christensen, T.E., Tang, L., 2012. Structural and functional studies of the phage Sf6 terminase small subunit reveal a DNA-spooling device facilitated by structural plasticity. J Mol Biol 423, 413-426.

APPENDIX

CONSTRUCTION OF P22 STRAINS

Recombineering of the *Salmonella* P22 prophage

In order to facilitate genetic study of the temperate bacteriophage P22, a strategy was needed that could introduce mutations into the prophage with as little collateral effect as possible and which allowed introduction of mutations that are lethal to lytic phage growth. In bacteria, markers that are both selectable and counter-selectable provide a way to genetically engineer a prophage. Thus, sections of prophage nucleotide sequence can be replaced with the selectable marker, and these can be subsequently replaced with any desired nucleotide sequence (Reyrat et al., 1998). The lytic growth phenotype of phage mutations introduced in this way can be determined by induction of the prophage.

Expression of the phage lambda *red* (exonuclease III) gene in bacteria has been shown to greatly increase the *in vivo* replacement of endogenous DNA sequence by exogenously added homologous (but different) DNA sequences (Datsenko and Wanner, 2000; Ellis et al., 2001). This lambda Red-mediated selection/counter-selection strategy, often referred to as "recombineering", provides an efficient way to engineer bacterial genome sequences. Since prophages are part of the host genome and since prophages do not require viral replication, recombineering of a P22 prophage should allow introduction of mutations into the phage genome regardless of their affect on lytic phage growth. A number of counter-selectable marker strategies have been developed in bacteria. For example, the fusaric acid sensitivity system (Li et al., 2013; Maloy and Nunn, 1981), the streptomycin sensitivity system (Dean, 1981; Lederberg, 1951), the *sacB* or sucrose sensitivity system (Li et al., 2013; Steinmetz et al., 1983)}, and the *galK* system (Warming et al., 2005) are all well-developed tools that are available. When recombineering a P22 prophage, specific mutations can be generated in plasmids containing cloned P22 DNA fragments by PCR amplification of the cloned plasmid with primers containing the desired mutation in their nucleotide sequence. The plasmid region containing such a mutation can then be amplified by PCR, and the resulting amplified DNA can be used as the exogenously added "replacement DNA" in such a recombineering strategy.

We have tried the fusaric acid and *galK* systems in *Salmonella* (the host for the P22

prophage), and the *galK* system provides the most robust selectivity and ease of use. In this system, lambda Red function is supplied by plasmid pKD46 (Datsenko and Wanner, 2000), and *galK* selection and counter-selection are carried out essentially as described by Warming *et al*. (2005). This strategy consists of lambda Red-mediated replacement of a target P22 prophage gene or gene section by an *E. coli galK* (which encodes galactokinase) expression cassette in a $galK^-$ host bacterium and selection for the ability to utilize galactose as a carbon source; this $galK^+$ DNA is introduced into the lysogen by electroporation (Padilla-Meier et al., 2012). Such a *galK* gene can be PCR amplified from plasmid pGalK (Warming et al., 2005) with primers whose 3'-tails each have ≥40 bp of identity to the targets that act as sites for Red-mediated homologous recombination and insertion of the *galK* gene that results in conversion of the bacterium from $galK^-$ to $galK^+$. The resulting $galK^+$ strain can then be supplied with a mutant homologous DNA fragment and counter-selected to $galK^-$ with 2-deoxy-D-galactose. Galactokinase (GalK) phosphorylates galactose to galactose-1-phosphate. GalK also phosphorylates 2-deoxy-D-galactose; however, 2-deoxy-D-galactose-1-phosphate cannot be utilized or altered further by the cell's metabolic machinery, and its build up is lethal to the cell. The replacement DNA fragment is typically generated by site-directed mutation of P22 DNA that has been cloned into a plasmid; the nucleotide sequence modification is accomplished by PCR amplification of the entire plasmid using overlapping and oppositely oriented mutation-containing oligonucleotide primers. Thus, the *galK* gene provides for both selection and counter-selection in this recombineering method.

Recombineering of a prophage presents some special problems that must be considered. A prophage can be excised and lost under laboratory conditions without a cost to the host (indeed its loss may give the host an advantage); therefore, care must be taken in the counter-selection to select for and/or screen out host cells whose prophage is absent. Our strategy accomplishes this by including a selectable drug resistance gene in the prophage. Spontaneous prophage induction also occurs at some level in nearly all temperate phages, so care must be taken to avoid wild type or parental phage contamination at all times.

We have made a number of changes in the phage P22 prophage genome (see Figure A.1) in order to facilitate genetic experiments and phage handling in the laboratory. We started with a P22 strain that carries a nonsense (*amber*) mutation in gene *13* that prevents lysis in a non-suppressing host, but allows plaque formation on an *amber* suppressing host. This allows strict control of lysis after prophage induction, since lysis can be induced by shaking the culture with a small amount of chloroform. Thus, the gene *13* defect allows easy purification of phage virions or phage-related particles, avoiding the long overnight centrifugations normally needed to concentrate phage particles for subsequent purification. A kanamycin resistance marker was inserted that replaces part gene *15* (Casjens et al., 1989) and all of *rha/orf25.* The gene *15* defect causes P22 to require citrate ions (which chelate divalent metal ions that inhibit cell lysis) in plates to make easily visible plaques. However, P22 plaques have very heterogeneous size on citrate plates and the presence of very tiny plaques among these can lead to under-estimation of plaque numbers. The kanamycin resistance marker allows easy selection for and monitoring of prophage presence. This *Salmonella* strain UB-1757 with prophage P22 *15*⁻Δsc302::KanR, *13*⁻ *am*H101 was described by Cortines *et al.* (2011).

Because of the actions of the phage-encoded Mnt and Arc repressors on their operators just upstream of the tailspike gene, P22 tailspike expression after induction is substantially lower than after infection (Adams et al., 1985; Israel, 1967). The result of this low expression is that a fraction of the virions produced by induction are not infectious because they lack tailspikes. In order to avoid this problem, we attempted to raise the expression of the tailspike gene by removal of the *sieA, arc* and*, ant* gene (*Immunity I*) region between genes *16* and *9* (tailspike) (Padilla-Meier et al., 2012) to create a P22 prophage-carrying strain UB-1790, whose plaque-forming prophage is called P22 UC-911 (genotype *15*⁻Δsc302::KanR, *13*⁻*am*H101, Δ*sieA*-1) in which tailspike is more highly expressed after induction. This *Salmonella* strain was also engineered to have a *galK* deletion in order to allow subsequent *galK* recombineering.

The well-characterized *Salmonella* laboratory strain LT2 that is normally used to propagate P22 also carries several other poorly inducible but fully functional lambdoid prophages Fels-1,

Gifsy-1 and Gifsy-2 (McClelland et al., 2001). In order to prevent low levels of spontaneous induction of these other prophages and possible recombination events with P22 (which have been observed; Droffner and Yamamoto (1982) and our unpublished results) from interfering with experiments, we obtained a prophage-free *Salmonella* LT2 from Dr. Anca Segall (UB-2232). This strain was further modified with a deletion of its native *galK* gene to enable *galK* recombineering, and a new P22 prophage variant was constructed. This new prophage has an intact gene *15*, so it does not require the addition of citrate for healthy plaque formation, and for prophage selection and monitoring, a chloramphenicol resistance cassette (Δorf25::CamR-EG1) was used to replace the small and completely dispensable *orf25* gene to the right (downstream) of the *rha* gene. In addition, a *c1*-7 clear plaque mutation (Botstein et al., 1973; Levine and Curtiss, 1961) was engineered into the prophage that lowers its ability to form a lysogen by several orders of magnitude but does not affect prophage stability or the ability to be induced. The CI protein is required for prophage establishment but not maintenance. This phage gives clear plaques, but the rare lysogens it forms can easily be selected by virtue of its CamR gene. Since a fraction of P22 $c1^{+}$ phages (such as UC-911 the prophage in UB-1790, above) form lysogens rather than infect lytically, the clear plaque morphology conferred by *c1*-7 is useful for more accurate titers and higher phage yields as well as for phenotype analysis of functional mutant phages during lytic infection. This P22 phage strain UC-937 (genotype Δorf25::CamR-EG1, *c1*-7, 13^{-}*am*H101, Δ*sieA*-1) was used to lysogenize the prophage-free strain UB-2232 to make UB-2235. These engineered, research-optimized versions of "wild type" (with regard to lytic growth) P22 strains greatly increase our ability to make genetic manipulations of all sorts as well as to examine the phenotypes of any mutant phages created.

References

Adams, M.B., Brown, H.R., Casjens, S., 1985. Bacteriophage P22 tail protein gene expression. J Virol 53, 180-184.

Botstein, D., Waddell, C.H., King, J., 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. I. Genes, proteins, structures and DNA maturation. J Mol Biol 80, 669-695.

Casjens, S., Eppler, K., Parr, R., Poteete, A.R., 1989. Nucleotide sequence of the bacteriophage P22 gene 19 to 3 region: identification of a new gene required for lysis. Virology 171, 588-598.

Cortines, J.R., Weigele, P.R., Gilcrease, E.B., Casjens, S.R., Teschke, C.M., 2011. Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. Virology 421, 1–11.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97, 6640-6645.

Dean, D., 1981. A plasmid cloning vector for the direct selection of strains carrying recombinant plasmids. Gene 15, 99-102.

Droffner, M.L., Yamamoto, N., 1982. Analysis of proteins induced by the *Salmonella typhimurium* Phage P221, a hybrid between serologically and morphologically unrelated phages P22 and Fels 1. J Gen Virol 59, 377-385.

Ellis, H.M., Yu, D., DiTizio, T., Court, D.L., 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98, 6742-6746.

Israel, V., 1967. The production of inactive phage P22 particles following induction. Virology 33, 317-322.

Lederberg, J., 1951. Streptomycin resistance; a genetically recessive mutation. J Bacteriol 61, 549-550.

Levine, M., Curtiss, R., 1961. Genetic fine structure of the C region and the linkage map of phage P22. Genetics 46, 1573-1580.

Li, X.T., Thomason, L.C., Sawitzke, J.A., Costantino, N., Court, D.L., 2013. Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in *Escherichia coli*. Nucleic Acids Res 41, e204.

Maloy, S.R., Nunn, W.D., 1981. Selection for loss of tetracycline resistance by *Escherichia coli*. J Bacteriol 145, 1110-1111.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., Wilson, R.K., 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature 413, 852-856.

Padilla-Meier, G.P., Gilcrease, E.B., Weigele, P.R., Cortines, J.R., Siegel, M., Leavitt, J.C., Teschke, C.M., Casjens, S.R., 2012. Unraveling the role of the C-terminal helix turn helix of the coat-binding domain of bacteriophage P22 scaffolding protein. J Biol Chem 287, 33766-33780.

Reyrat, J.M., Pelicic, V., Gicquel, B., Rappuoli, R., 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. Infect Immun 66, 4011-4017.

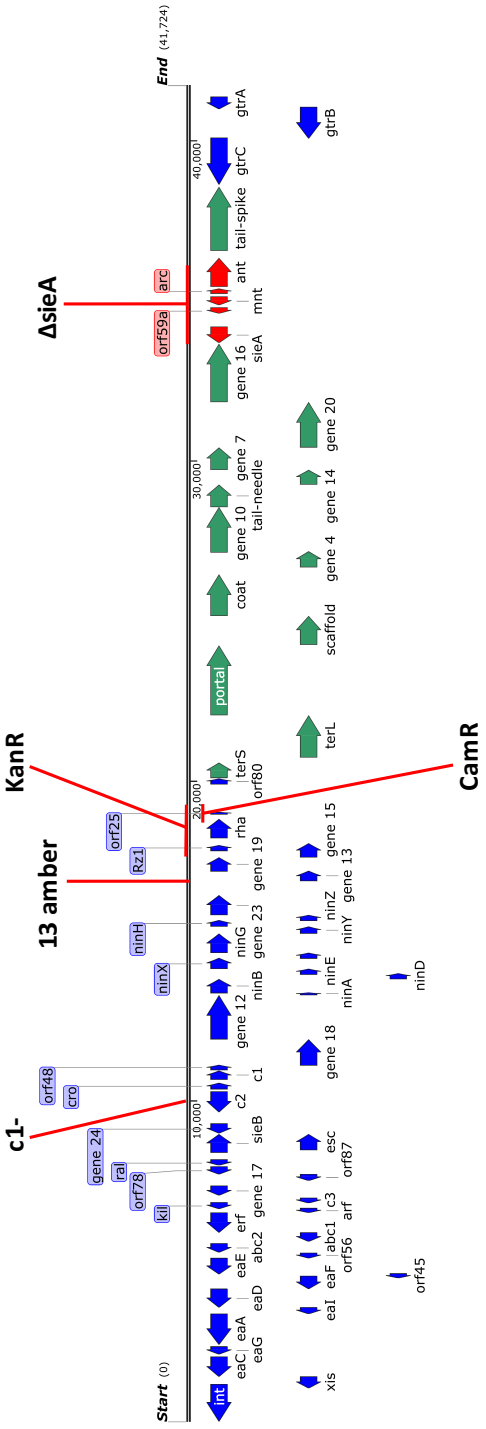
Steinmetz, M., Le Coq, D., Djemia, H.B., Gay, P., 1983. Genetic analysis of *sacB*, the structural gene of a secreted enzyme, levansucrase of *Bacillus subtilis* Marburg. Mol Gen Genet 191, 138-144.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., Copeland, N.G., 2005. Simple and highly efficient BAC recombineering using *galK* selection. Nucleic Acids Res 33, e36.

Figure A.1 Modifications made to the P22 prophage genome that increase recombineering efficacy.

The P22 genome is shown in its prophage orientation (small numbers indicate locations in bp). Colored arrows denote the phage P22 genes. Green genes are required for virion assembly, *Immunity I* region genes are red, and all other genes are blue. The image was generated with SnapGene (GSL Biotech; available at snapgene.com) and modified with Microsoft PowerPoint. The engineered mutations discussed in the text are indicated in large black text with red lines showing their position in the genome as follows:

1. *c1*$^{-}$, the clear plaque mutation; allele name is " *c1*-7"
2. 1*3*$^{-}$ *amber* , the nonsense mutation in gene *13*; allele name is "*13*$^{-}$*am*H101"
3. KanR, the kanamycin resistance marker that replaces genes *15, Rz1, rha*, and *orf25*; allele name is "*15*$^{-}$Δsc302::KanR"
4. CamR, the chloramphenicol resistance marker that replaces *orf25*; allele name is "Δorf25::CamR-EG1"
5. ΔsieA, the deletion of the *Immunity I* region and *sieA* gene; allele name is "Δ*sieA*-1"


Start (0)
End (41,724)
10,000
20,000
30,000
40,000
int
xis
eaC
eaA
eaG
eaI
eaD
eaF
eaE
orf56
abc2
abc1
erf
c3
arf
gene 17
esc
orf87
sieB
orf45
c2
c1
gene 18
gene 12
ninB
ninA
ninE
ninG
gene 23
ninZ
ninY
gene 13
ninD
rha
gene 19
gene 15
terS
orf80
portal
terL
coat
scaffold
gene 10
tail-needle
gene 4
gene 14
gene 7
gene 20
gene 16
sieA
ant
mnt
tail-spike
grfC
grfA
grfB
orf48
cro
gene 24
ral
orf78
kil
ninX
ninH
orf25
Rz1
orf59a
arc
c1-
KanR
13 amber
ΔsieA
CamR