Journal of Virology

A Multivalent Adsorption Apparatus Explains the Broad Host Range of Phage phi92: a Comprehensive Genomic and Structural Analysis

David Schwarzer, Falk F. R. Buettner, Christopher Browning, Sergey Nazarov, Wolfgang Rabsch, Andrea Bethe, Astrid Oberbeck, Valorie D. Bowman, Katharina Stummeyer, Martina Mühlenhoff, Petr G. Leiman and Rita Gerardy-Schahn

J. Virol. **2012, 86(19):10384. DOI: 10.1128/JVI.00801-12. Published Ahead of Print 11 July 2012.**

JVI Journals.ASM.org

A Multivalent Adsorption Apparatus Explains the Broad Host Range of Phage phi92: a Comprehensive Genomic and Structural Analysis

David Schwarzer,[a] Falk F. R. Buettner,[a] Christopher Browning,[b] Sergey Nazarov,[b] Wolfgang Rabsch,[c] Andrea Bethe,[a] Astrid Oberbeck,[a] Valorie D. Bowman,[d] Katharina Stummeyer,[a]* Martina Mühlenhoff,[a] Petr G. Leiman,[b] and Rita Gerardy-Schahn[a]

Hannover Medical School (MHH), Institute for Cellular Chemistry (OE 4330), Hannover, Germany[a]; Ecole Polytechnique Fédérale de Lausanne (EPFL), Institut de physique des systèmes biologiques, LBBS, Lausanne, Switzerland[b]; Robert Koch Institute, Wernigerode Branch, National Reference Centre for Salmonellae and Other Enterics, Wernigerode, Germany[c]; and Department of Biology, Purdue University, West Lafayette, Indiana, USA[d]

Bacteriophage phi92 is a large, lytic myovirus isolated in 1983 from pathogenic *Escherichia coli* strains that carry a polysialic acid capsule. Here we report the genome organization of phi92, the cryoelectron microscopy reconstruction of its virion, and the reinvestigation of its host specificity. The genome consists of a linear, double-stranded 148,612-bp DNA sequence containing 248 potential open reading frames and 11 putative tRNA genes. Orthologs were found for 130 of the predicted proteins. Most of the virion proteins showed significant sequence similarities to proteins of myoviruses rv5 and PVP-SE1, indicating that phi92 is a new member of the novel genus of rv5-like phages. Reinvestigation of phi92 host specificity showed that the host range is not limited to polysialic acid-encapsulated *Escherichia coli* but includes most laboratory strains of *Escherichia coli* and many *Salmonella* strains. Structure analysis of the phi92 virion demonstrated the presence of four different types of tail fibers and/or tailspikes, which enable the phage to use attachment sites on encapsulated and nonencapsulated bacteria. With this report, we provide the first detailed description of a multivalent, multispecies phage armed with a host cell adsorption apparatus resembling a nanosized Swiss army knife. The genome, structure, and, in particular, the organization of the baseplate of phi92 demonstrate how a bacteriophage can evolve into a multi-pathogen-killing agent.

Among Gram-negative bacteria, a huge variety of cell-surface polysaccharides is found (79). These sugar polymers can form thick coats, which mask underlying surface structures, protect the bacterium from the host immune system, and create an efficient barrier against bacteriophage infection (15, 74). However, the protective sugar coats can also serve as attachment sites for specialized phages that possess polysaccharide depolymerases as tailspike proteins, enabling these phages to penetrate the coat and to gain access to cell surface receptors (71).

Well-studied examples of tailspike proteins with polysaccharide depolymerase activity are endosialidases or endo-N-acylneuraminidases (endoN) (26, 28, 32, 52, 67, 68, 72). Endosialidases specifically degrade α2,8-linked polysialic acid (polySia), the capsular polysaccharide of *Escherichia coli* K1 (*E. coli* K1) (4). K1-specific phages are mainly small bacteriophages of the *Podoviridae* family (podoviruses) with short noncontractile tails (42; for an overview, see reference 32). The size of podoviruses allows the incorporation of only one or two different tailspike proteins (42), resulting in a very limited host range. K1-specific podoviruses evolved from different progenitor types and gained their new host specificity by horizontal uptake of an endosialidase gene (73). For example, the genomes and virion structures of K1F and T7 are similar except for the part responsible for host cell binding. The T7 fiber is replaced by an endoN tailspike in K1F, and this fully commits K1F to encapsulated K1 strains (73). Whereas the K1 capsule is absolutely required for K1F infection (65), it blocks infection by T7, which infects only nonencapsulated *E. coli* (64).

The larger genomes of *Myoviridae*—phages with long, contractile tails—give rise to higher complexity in structural organization and infection, mechanism as extensively studied for coliphage T4 (for an overview, see reference 44). The only anti-K1 myovirus described so far is phi92 (or φ92). It was isolated in 1983 (38) from *E. coli* K92 (Bos12 [serogroup O16:K92:H$^-$]), which is encapsulated by polySia with alternating α2,8/α2,9 linkages (19, 25). It has been further demonstrated that phi92 also degrades the α2,8-linked polySia of *E. coli* K1 but not the α2,9-linked polySia capsules of *Neisseria meningitidis* serogroup C (38, 39), suggesting the presence of an α2,8-specific endosialidase tailspike. To date, no further studies of phi92, including the genome, structure, or host range, have been described.

In the present report, we demonstrate that phi92 infects not only encapsulated but also nonencapsulated *E. coli* as well as clinically important *Salmonella* strains. Sequencing of the 148-kb genome of phi92 revealed several putative tail fiber and tailspike genes. Cryoelectron microscopy (cryoEM) reconstruction of phi92 phage particles highlighted the presence of a multivalent host adsorption apparatus providing the phage with broad host specificity.

MATERIALS AND METHODS

Bacteria and bacteriophages. *E. coli* Bos12 (O16:K92:H$^-$) (ATCC 35860) and coliphage phi92 ("phi 92"; ATCC 35860-B1) were obtained from the American Type Culture Collection. *Bacillus subtilis* DSM 5545, *Bacillus pumilus* G1R (DSM 5549), and *Pseudomonas putida* sp. DSM 50906 as

Received 30 March 2012 Accepted 6 July 2012
Published ahead of print 11 July 2012
Address correspondence to David Schwarzer, Schwarzer.David@mh-hannover.de, or Petr G. Leiman, petr.leiman@epfl.ch.
* Present address: Katharina Stummeyer, GRS—Gesellschaft für Anlagen- und Reaktorsicherheit, Cologne, Germany.
P. G. Leiman and R. Gerardy-Schahn contributed equally to this article.
Supplemental material for this article may be found at http://jvi.asm.org/.

doi:10.1128/JVI.00801-12

well as *Enterobacteria* phage T7 (DSM4623) were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). *E. coli* JM105 and YZ2000 were purchased from Pharmacia and Gene Bridges, respectively. *E. coli* BL21(DE3) and BL21-Gold(DE3) were obtained from Stratagene.

The *E. coli* K5 strain has been described previously (68). *E. coli* strains K1 B2032/82 (24), K-235 (serogroup O1:K1:H$^-$) (37), K1 U9/41 (O2:K1:H4) (28), EV5 (K1$^-$) (76), and Nissle 1917 (O6:K5:H) (54) as well as *Enterobacteria* phage K1F were obtained from the strain collection of the Department of Medical Microbiology of the Hannover Medical School (MHH, Hannover, Germany). *Salmonella* strains used in this study were obtained from the strain culture collection of the Robert Koch Institute, Wernigerode Branch (Wernigerode, Germany).

Phage propagation. *E. coli* strain Bos12 (O16:K92:H$^-$) was used for propagation of coliphage phi92 (38). *Enterobacteria* phage K1F was grown on *E. coli* U9/41 (O2:K1:H4) (28). Purification of phage particles and isolation of phage DNA were performed as described previously (26). The protein content of phage particles was estimated using the Bio-Rad protein assay following the manufacturer's guidelines.

Determination of phage host range. Infection studies of *Salmonella* strains were performed on blue plates as described previously (60). Essentially, 10 μl of 10-fold serial dilutions (1:100 to 1:10^{-10}) of phage suspension was spotted onto bacterial lawns. The plates were incubated overnight at 37°C, and the lytic activity—expressed as the maximal dilution at which no lysis was observed—was determined by monitoring lysis areas on the bacterial lawns. The routine phage typing of *Salmonella enterica* serovar Enteritidis was performed according to the typing system of Ward et al. and László et al., e.g., PT21/1b (phage type 21) according to Ward et al. and 1b according to László et al. (40, 77). The lytic activity of phi92 on *S.* Enteritidis was set to 100%.

Infection studies of other hosts were performed by double-layer plaque assays as described previously (61). In brief, 10 μl of 10-fold serial dilutions (1:100 to 1:10^{-10}) in SM buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 8 mM $MgSO_4$, 0.01% [wt/vol] gelatin) of phi92 and K1F was incubated with 400 μl of prepared host cells (A_{600} = 0.5 in 10 mM $MgSO_4$) for 20 min at room temperature followed by a step for 10 min at 37°C. The suspension was thoroughly mixed with 4 ml of melted top agarose (1% [wt/vol] Bacto Tryptone, 43 mM NaCl, 0.7% [wt/vol] agarose at 48°C) and immediately plated on a prewarmed (37°C) agar plate using LB agar (for *E. coli* strains), tryptic soy agar (for *Bacillus* strains), or nutrient agar (for *Pseudomonas putida*). After incubation for 6 h at 37°C, plaques were counted and titers were calculated (in PFU per milliliter). To calculate relative titers on *E. coli* strains, the titer of the phi92 standard host *E. coli* K92 was set to 100%.

phi92 genome sequencing and annotation. The genome of phi92 was *de novo* sequenced by a combination of shotgun sequencing and primer walking at GATC Biotech AG (Konstanz, Germany), with an average coverage of 3×. Resequencing of the genome was performed with Next Generation Sequencing at GATC Biotech AG (Konstanz, Germany), with an average coverage of 53.7×. The procedure was essentially as follows. The double-stranded DNA (dsDNA) was sheared on a Covaris shearing system to fragments of a size of approximately 700 bp and ligated with the A and B adapters (Roche) for sequencing. The samples were separated on a 2% agarose gel, and the band in a size range of 700 to 900 bp was excised and column purified. The resulting library was immobilized onto DNA capture beads and amplified through emulsion-based PCR (emPCR) according to the manufacturer's recommendations. The emulsion was chemically broken, and the beads carrying the amplified DNA library were recovered and washed by filtration. The sample was sequenced on a 1/16 GS FLX Pico-Titer plate device with GS FLX Titanium XLR70 chemistry. The GS FLX system produced the sequence data as a Standard Flowgram Format (SFF) file containing flowgrams for each read with base calls and per-base quality scores. The analysis was performed with a GS FLX System Software GS De Novo Assembler (Newbler). In total, 27,496 reads were assembled into one contig of 148,269 bp lacking one direct terminal repeat (DTR), which was determined by Sanger sequencing.

For the prediction of open reading frames (ORFs), three bioinformatics tools were consecutively applied. Using *ab initio* ORF searching and GenemarkS (GeneMark.hmm-PS; PROKARYOTIC [Version 2.6p] [7, 47]), 247 potential ORFs were identified in the phi92 genome (see column *H*: 'G' in Dataset S1 in the supplemental material). The data were revised with NCBI ORF Finder (*N*) (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and with Prodigal (*P*) (30). The latter novel prediction server returned extensive statistics for each ORF (see Dataset S1 in the supplemental material). In total, 4,093 possible ORFs were identified (data not shown), and the 239 ORFs with the best statistics were predicted as potential genes. The full list of 4,093 possible ORFs was used to create a Microsoft Excel macro, which automatically compared the results of the three prediction programs. The NCBI ORF Finder (*N*) returned 631 putative ORFs, of which 238 matched the genes predicted by GenemarkS and Prodigal. The remaining ORFs had poor Prodigal statistics and were disregarded. Due to poor statistics, 9 (ORFs 9, 40, 42, 90, 113, 117, 133, 215, and 237) of 248 ORFs found by GenemarkS were not confirmed by Prodigal. For another 11 ORFs, Prodigal predicted alternative start codons leading to increased ORF sizes (*G_alt*). In contrast, ORF 228 was not identified by GenemarkS. For ORF 106, two alternative starts (referred to as genes 106A and 106B, respectively) were predicted by GenemarkS and Prodigal with similar statistics.

Homologous protein searches were performed using NCBI BLAST against the database of nonredundant protein sequences (nr) and the following parameters and values: search algorithm, BLASTP 2.2.25+ (1, 2); matrix, BLOSUM62; gap existence, 11; gap extension, 1; allow low complexity regions; method, compositional matrix adjustment. Analysis using HHpred/HHsearch (homology detection and structure prediction by hidden Markov model [HMM]-HMM comparison) was performed online at http://toolkit.tuebingen.mpg.de/hhpred (69, 70). Secondary structure prediction was performed with Psipred (10, 33). Potential tRNA genes were predicted using tRNAscan-SE v.1.21 (45), and potential signal peptides were searched using SignalP 3.0 HMM (20). The genome map was drawn using DNAPlotter (12).

1-DE. Whole-phage particles were incubated in the presence of 1.5% (wt/vol) sodium dodecyl sulfate (SDS) for 5 min at 95°C and loaded onto SDS gels. One-dimensional gel electrophoresis (1-DE) was performed at 85 V. Proteins were stained using RotiBlue (Carl Roth GmbH) according to the manufacturer's guidelines. Scanning of RotiBlue-stained gels was performed in an Odyssey Infrared imaging system (Li-COR).

2-DE. Whole phi92 phage particles were denatured for 5 min at 95°C in the presence of 2.5% (wt/vol) SDS. Proteins were precipitated with 6 volumes of acetone. The protein pellet was washed twice with 80% acetone, dried by aeration with nitrogen, and dissolved in 30 mM Tris-HCl (pH 8.5)–7 M urea–2 M thiourea–4% (wt/vol) CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}. Electrophoresis was performed on Immobiline DryStrips (GE Healthcare) (24 cm in length) with a pI gradient of from 3 to 11 for the first-dimension focusing and 10% polyacrylamide gels for the second dimension according to a procedure described previously (11).

Protein identification by UPLC-coupled Q-TOF MS/MS. Protein identification was performed by ultraperformance liquid chromatography (UPLC)-coupled quadrupole time of flight (Q-TOF) tandem mass spectroscopy (MS/MS) (nanoACQUITY UPLC system coupled to Q-TOF Ultima; Waters) as described previously (11). Processing of spectra and searching of the phi92 protein database generated from its genome (accession number FR775895 [this study]) were performed using the program ProteinLynxTM Global Server (Version 2.1; Waters). Additionally, mass spectra were analyzed searching an *E. coli* genome database (*Escherichia coli* [562]) downloaded at http://www.uniprot.org/ from the Protein Knowledgebase (UniprotKB) on 31 January 2011. The results are shown in Dataset S2 in the supplemental material and summarized here (see Fig. 3 and Table 2, respectively).

Cryoelectron microscopy and reconstruction of phi92. Low-dose cryoEM was performed as described previously (3). The images were recorded on Kodak SO-163 film using a Philips CM300 FEG microscope at a magnification of ×33,000 with a radiation dose of ~20 e−/Å^2 and a defocus of −3.5 to −2.0 μm. The images were scanned using a Zeiss SCAI scanner with a 7-μm step size and were binned to obtain a pixel size of 2.108 Å. The tail and baseplate images were binned further to give a pixel size of 4.216 Å, whereas the final capsid reconstruction was calculated with a pixel size of 2.811 Å. The capsid and baseplate/tail datasets contained 1,137 and 985 boxed images, respectively, which were picked from 38 micrographs. The contrast transfer function (CTF) was calculated and then corrected with either EMAN (46) or CTFFIND3 (50) software.

The icosahedral reconstruction of the phi92 capsid was calculated with EMAN from scratch. The initial model was constructed with a small subset of initial data using EMAN routines. This model was then subjected to several rounds of refinement until convergence. The reconstruction of the phi92 tail was obtained in two steps. Initially, the structure of the distal part of the tail containing the baseplate and about one-third of the sheath was calculated. Then, the reconstructed volume was increased to include the entire sheath, excluding the fibers emanating from the baseplate.

The T4 baseplate-tail tube complex (EMDB-1048) was used as the initial model for the baseplate reconstruction. The SPIDER (23) and EMAN software suites were employed in image processing. The reconstruction was initiated with EMAN, but due to the large discrepancy between the initial model and the phi92 images, the refinement procedure was quickly trapped in a local minimum, with tail fibers being aligned to the central spike and the tail sheath layers incorrectly cross-correlated with each other. This problem was overcome by modifying this "ensemble" structure with a threshold-based Gaussian-filtered mask and limiting the search angles to a region of in-plane orientations. The search angles for the tilt were restricted to between 40 and 90 degrees, as most phage particles had their long axes roughly parallel to the imaging plane. These operations and subsequent refinement were performed with SPIDER. The resolution of the reconstruction was found to be 26 Å using the Fourier shell correlation coefficient method, assuming the limit of resolution to be at the point where the correlation coefficient drops to below 0.5.

The tail reconstruction was calculated by increasing the box size of the baseplate map to include the sheath in its entirety. The center of this box was then shifted to roughly coincide with the sheath's center. The reconstruction was refined with FREEALIGN (27). Six-fold symmetry was used for all orientation searches and reconstructions of the baseplate and the tail. As the final maps had shown the features expected to be seen at their medium resolutions, the 6-fold symmetry assumption was correct. The left-right structural orientation ("hand") of the reconstruction was confirmed with atomic force microscopy (AFM) analysis as described in reference 9. Fitting of the HK97 capsid structure into the phi92 cryoEM map was performed with UCSF, Chimera (57).

Nucleotide sequence accession numbers. Genome sequence data have been submitted to the DDBJ/EMBL/GenBank databases under accession number FR775895 (see Dataset S1 in the supplemental material for protein accession numbers). The cryoEM maps of the phi92 capsid, tail, and baseplate were deposited into the EMDataBank under accession numbers EMD-2063, EMD-5409, and EMD-2064, respectively.

RESULTS AND DISCUSSION

Genome structure, protein homologs, and related phages. The 148,612-bp genome of double-stranded DNA phage phi92 was sequenced to a total average coverage of 56.7×. Sudden dropoffs in the Sanger-sequencing signals were consistent with the ends of the phi92 genome comprising invariant DTRs of 343 bp in length. The genome size is in line with the previous grouping of phi92 into the family of large *Myoviridae* phages (29, 38). The average moles percent GC content was determined to 37.4%, which is comparable to that of other lytic phages such as T4 and Felix-O1 (35.4% and 39.0%, respectively) but significantly lower than the GC content of the host *E. coli* (~50%) (49, 78).

By consecutive application of three different bioinformatics tools (see Materials and Methods), a total of 248 ORFs were annotated (see Dataset S1 in the supplemental material). Moreover, 11 tRNA genes, 2 pseudo-tRNA genes, and 5 hypothetical tRNA genes were identified (see Dataset S1 in the supplemental material and tRNA cluster in Fig. 1A). High sequence similarities were found for 130 putative phi92 proteins (dark blue or dark green in Fig. 1A), and a potential function could be assigned for 80 of these.

The largest numbers of protein homologs in single complete GenBank entries (56 [23%] and 63 [25%]) were found for myoviruses PVP-SE1 and rv5 (GenBank accession numbers NC_011041 and GU070616), respectively (Fig. 1B [see also Tables S1 and S2 in the supplemental material]; cf. Fig. 2). Further common features of the genomes of phi92 PVP-SE1 and rv5 are their sizes (138 and 146 kb) and the numbers of genes (233 and 244) and proportions of GC content (43% and 46%). The orthologous proteins are encoded in a collinear order in all three phage genomes, suggesting a common ancestry (Fig. 1B). Moreover, PVP-SE1 and rv5 have recently been proposed to belong to a novel myovirus genus, the rv5-like phages (63).

Genome organization. As illustrated in Fig. 1A, the predicted genes in phi92 are clustered in at least five transcriptional units, which lie alternately on the direct and complementary strands. The first operon (genes 1 to 36) is located at the left end of the genome. Six gps (gp1, gp5, gp25, gp26, gp27, and gp33) contain DNA-binding motifs and might encode putative transcription factors that act in the early phase of the infection cycle. The last (fifth) operon on the right end is also positioned on the direct strand. It is possible that the first and the fifth operons were part of a "superoperon" which was split by the DTRs at an earlier evolutionary time point. Proteins that are encoded in the second and the fourth operons are known to function in the middle phase of the phage infection cycle and are involved in nucleotide metabolism (Table 1 and Fig. 1A), polynucleotide synthesis, modification, recombination, and repair (Table 1).

TBLASTN search analysis of genes 192 and 193 suggested that they are likely to have originated from a single gene (referred to as 192A) that was interrupted by a 279-bp group I intron. In contrast, the gp192A orthologs rv5 gp108 and PVP-SE1 gp144 are encoded by continuous genes (see Tables S1 and S2 in the supplemental material and cf. Figure 1A). The gp192A product is only 8 amino acids (aa) shorter than the combined lengths of the gp192 and gp193 products and shows significantly better statistics compared to the nrdD-like proteins of the ribonucleotide reductase family than those of the two separate gps (see Dataset S1 in the supplemental material). Since the length of the phi92 intron is greater than the 250 nucleotides (nt) required for efficient splicing of group I introns (6), the insertion between genes gp192 and gp193 is most likely a group I self-splicing intron.

Bioinformatic analysis of the structural operon. Using the HHpred analysis combined with the genome location for any given gene, it was possible to assign functions to genes 114 to 153 from the late or structural operon, which is responsible for the assembly of the phage particle (see Dataset S1 in the supplemental material and Fig. 1). The structure of proteins comprising the phage particle is highly conserved (44), and, in agreement with this, 85% of the proteins from the phi92 structural operon were found to have orthologs in other phages, especially from the rv5-

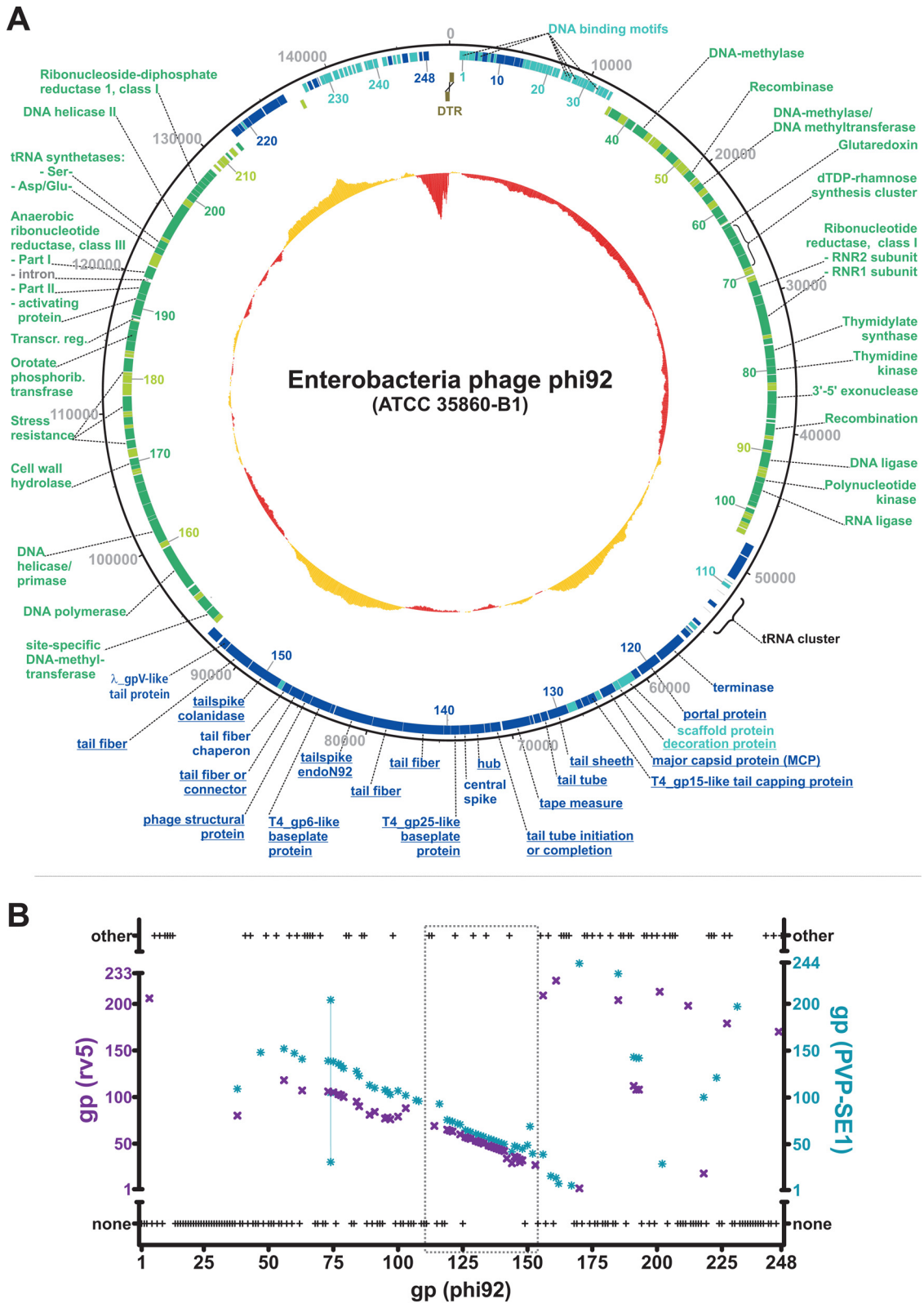
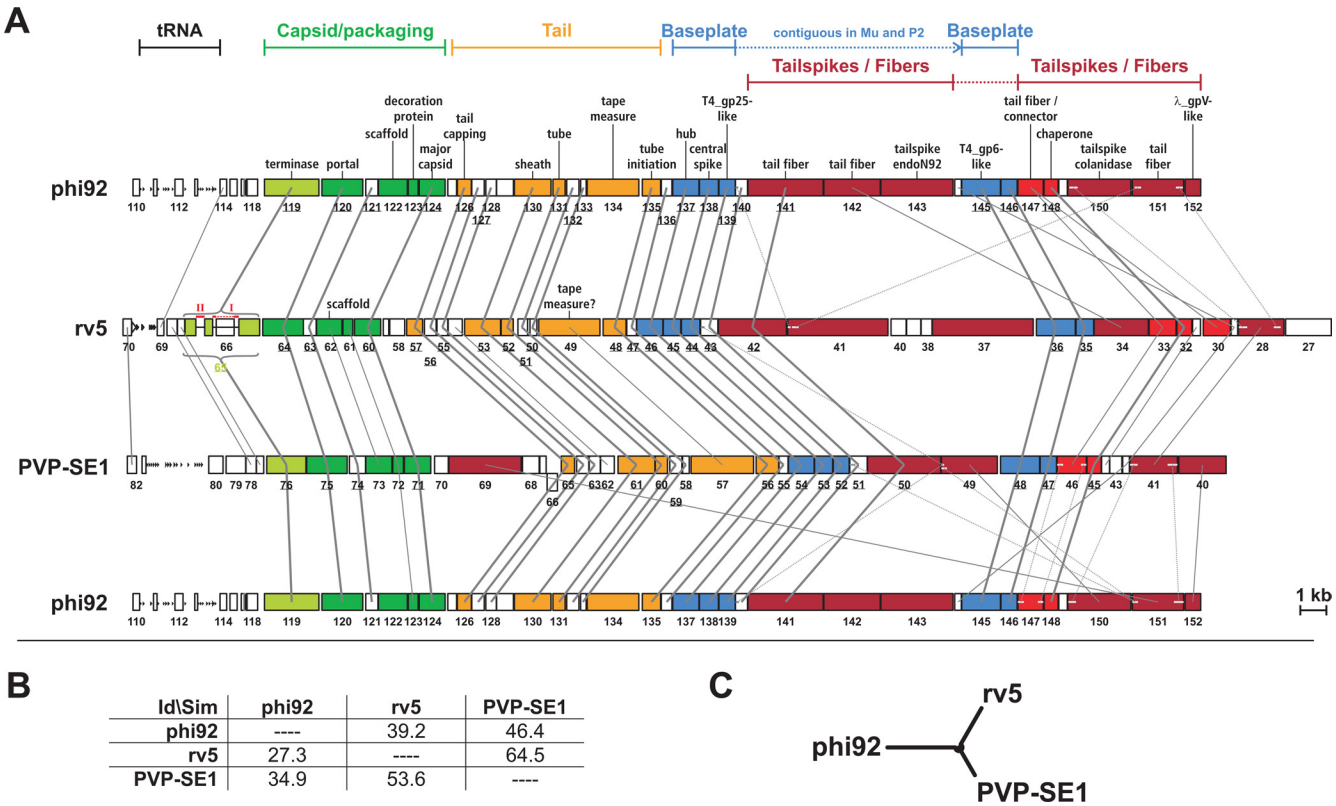


FIG 1 Schematic overview of the phi92 genome. (A) Genome map of phi92. The circular map represents the linear phi92 genome, including both DTRs (drawn as brown boxes). Predicted proteins encoded on the direct or complementary strand are shown in blue or green, respectively. Hypothetical proteins without homologs in other organisms are depicted in pale blue and pale green. The predicted function of selected phi92 proteins is indicated outside and the gene numbers are inside the circle, each in the corresponding color. Underlined names indicate proteins identified by MS (cf. Fig. 3). The cluster of predicted tRNA is shown in black characters. The GC content is depicted in the center of the map, whereas values above or below the average GC content value (37.4%) are shown in orange or red, respectively (see also Datasets S1 and S2 in the supplemental material). (B) Proteome plot of phi92 [gp (phi92)] (*x* axis) versus protein orthologs from rv5 [magenta, left axis; gp (rv5)] and PVP-SE1 [turquoise, right axis; gp (PVP-SE1)] with more than 35% sequence similarity over the whole protein sequence (see Tables S1 and S2 in the supplemental material) or from other organisms ("other") or when no homologs were found in the databases ("none"). The three PVP-SE1 proteins (gp030, gp104, and gp204) that are homologous to phi92 gp74 are connected by a turquoise line. The dashed box borders the virion-associated proteins encoded in the structural operon.



Id\Sim	phi92	rv5	PVP-SE1
phi92	----	39.2	46.4
rv5	27.3	----	64.5
PVP-SE1	34.9	53.6	----

FIG 2 Comparison of structural operons of rv5-like phages. (A) Comparison of the structural operons of phi92, rv5, and PVP-SE1 drawn at the same scale (bar, 1 kb). Genes encoding a protein are represented by boxes filled with the respective colors according to the predicted function. Gene numbers are given below the boxes. The PVP-SE1 and rv5 gp numbers are according to a recent publication (63). In contrast, in the genome Genbank file (NC_016071.1) of PVP-SE1, gp015 to gp144 are shifted by one (i.e., gene 16 encodes gp015, etc.). tRNA genes are given as black arrows. Note that gene 66 of phage PVP-SE1 is encoded on the complementary strand. The genes can be grouped according to the locations and functions of their products in the phage virion (bars above) as represented in different colors. Functions assigned to phi92 proteins are given above the phi92 genome data. Gray solid lines and dashed lines between two organisms represent homologous proteins and protein domains, correspondingly, based on BLASTP search analysis (phi92 versus rv5 and PVP-SE1; E-value $< 10^{-5}$ [cf. Tables S1 and S2 in the supplemental material]) or on recently published data (rv5 versus PVP-SE1) (63), respectively. Proteins conserved in all three phages (underscored gene numbers) are connected by thick gray lines. Terminase gene 65 of phage rv5 contains two introns (red I and II). (B) Identity/similarity matrix based on a CLUSTAL W alignment of 21 homolog proteins of rv5-like phages (gp119, gp120, gp121, gp124, gp126, gp127, gp128, gp130, gp131, gp132, gp133, gp135, gp136, gp137, gp138, gp139, gp140, gp141, gp145, gp146, and gp148 of phi92 and the respective orthologs of rv5 and PVP-SE1 [underscored numbers in panel A]) concatenated into one polypeptide sequence for each phage. (C) Phylogenetic tree based on the CLUSTAL W alignment calculated as described for panel B.

like genus. Genes corresponding to the conserved structure of a contractile tail could be easily identified. The tail sheath and tube are products of genes 130 and 131; the tape measure protein and its putative chaperone (or tube initiator) are products of genes 134 and 135. Genes 126, 139, and 145 encode orthologs of the T4 baseplate proteins gp15, gp25 and gp6, respectively.

Twenty-one proteins from the phi92 structural operon (underlined numbers in Fig. 2A) display a high sequence similarity to those of phages rv5 and PVP-SE1 (on average, 39.2% and 46.4%, respectively; Fig. 2B), showing that the rv5-like phage genus is characterized by a common structural operon. The order of the genes in this operon is very similar to that of phage Mu (GenBank accession number NC_000929; see Fig. S1 in the supplemental material) but shows an increased complexity (Fig. 2A). Compared to that of Mu, the phi92 baseplate cluster contains an insertion of five genes (three of which are tail fiber/tailspikes) between genes 139 and 145, which correspond to Mu genes 46 and 47 (see Fig. S1 in the supplemental material). A phylogenetic analysis of the virion proteins illustrates that phi92 is more closely related to PVP-SE1 than to rv5 in the genus (Fig. 2C).

MS study of protein composition of the phi92 phage particle. Bioinformatic analysis was further validated with 1-DE and 2-DE and subsequent MS (Fig. 3; see also Dataset S2 in the supplemental material). All phi92 proteins identified in this procedure are exclusively encoded in the structural operon (Table 2; underlined in Fig. 1A), with the gp124 major capsid protein (MCP) forming the most prominent signal in both the 1-DE and 2-DE gels.

Several prominent spots observed in the 2-DE gel (Fig. 3C) contained *E. coli* proteins (in repeated experiments). In spot 9 of the 2-DE gel, the phi92 proteins gp123 and gp135 were identified together with LamB (maltoporin) from *E. coli*. Spots 6, 12, 14, 15, 16, 17, 21, and 25 (underscored and italicized in Fig. 3C) contained the outer membrane proteins OmpA, OmpC, and OmpW. For phage Sf6, OmpA and OmpC have been found as contaminants during purification (55), suggesting that the host outer membrane proteins were copurified with phi92 as well.

CryoEM reconstruction of the phi92 virion. Due to the limited size of the cryoEM dataset, we used the divide-and-conquer image-processing approach and the reconstructions of the phi92 capsid, tail, and baseplate were calculated separately, assuming the

TABLE 1 phi92 proteins of the middle phase assigned to a potential functional category

Protein and category	Predicted function or description	Category
Nucleotide metabolism		
gp085	5′-Exodeoxyribonuclease	Scavenging
gp078	Thymidylate synthase	dTTP synthesis
gp080	Thymidine kinase	
gp063	Glutaredoxin 1 (nrdC-like)	dNTP synthesis
gp073	Class I ribonucleotide reductase (RNR2 subunit; nrdB-like)	
gp075	Class I ribonucleotide reductase (RNR1 subunit; nrdA-like)	
gp186	Orotate phosphoribosyltransferase	
gp202	Class I ribonucleoside-diphosphate reductase 1 (alpha subunit, nrdA.1-like)	
gp191	Class III anaerobic ribonucleotide reductase; activating protein (nrdG-like)	
gp192	Class III anaerobic ribunucleotide reductase; part II (nrdD-like)	
gp193	Class III anaerobic ribunucleotide reductase; part I (nrdD-like)	
gp192A	Class III anaerobic ribunucleotide reductase; spliced variant (nrdD-like)	
Polynucleotide metabolism		
gp159	DNA polymerase	Synthesis
gp161	Primase/helicase (T7-like)	
gp053	Recombinase, resolvase family protein/DNA invertase	Recombination
gp087	Recombination-associated protein	
gp043	Site-specific DNA-methylase	Modification
gp056	DNA-methylase/DNA-methyltransferase	
gp155	Site-specific DNA-methyltransferase	
gp084	DNA polymerase-type 3′–5′-exonuclease	Repair
gp091	ATP dependent DNA ligase	
gp095	Polynucleotide kinase	
gp096	RNA ligase/RNA repair	
gp198	DNA helicase II/UvrD superfamily I DNA and RNA helicase	

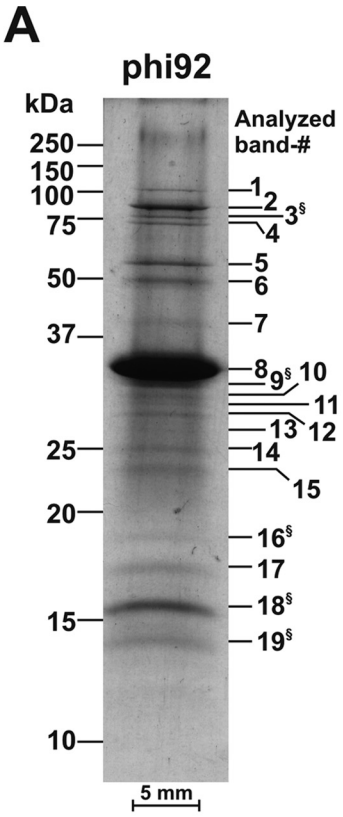
most appropriate symmetry for each part. Most particles were imaged with their long axis parallel to the imaging plane, thus aiding the image-processing procedure (Fig. 4A).

Structure of the phi92 capsid. Neglecting the unique portal vertex of the capsid (the tail attachment vertex [gp120]), the phi92 capsid shell obeyed icosahedral symmetry, and its structure, reconstructed as such, was obtained at a resolution of 19 Å (Fig. 4B). The capsid dimensions were found to be ~930 Å, ~795 Å, and ~810 Å along one 5-fold axis, one 3-fold axis, and one 2-fold axis, respectively. The thickness of the capsid shell ranged between 23 and 43 Å in a map contoured at a level of 1.5 standard deviations (σ) from the mean. Interestingly, even at this low contour level, the shell is permeated with holes of 13 by 21 Å (see Fig. S2 in the supplemental material).

The large body of existing structural data shows that the major capsid proteins (MCPs) of all known dsDNA tailed phages have a HK97 fold and that the capsid shell is composed of flat hexamers and wide-based cone-shaped pentamers formed by the MCP (80). In agreement with this, the phi92 capsid electron density could be segmented into hexamers and pentamers (Fig. 4B), easily accommodating the respective oligomers of the HK97 capsid protein (PDB ID 1OHG).

The hand of the cryoEM map was determined by comparing the original and mirrored maps with the structure of the HK97 capsid (36). Due to the limited resolution of the map, we used an automated fitting procedure that maximizes the number of atoms fitted into the density of the data as implemented in UCSF Chimera (57). The HK97 asymmetric unit comprising one full hexamer plus one subunit of the pentamer was fitted into the original and mirrored maps (see Fig. S2 in the supplemental material). Of 15,070 Cα atoms, the original and mirrored maps contoured at 1.5 σ do not account for 2,891 and 3,767 atoms, respectively. The positions of the hexamer autofitted with or without the pentamer subunit into the original hand map were virtually identical. The results of the autofit procedure involving the hexamer without the pentamer subunit fitted into the mirrored map did not converge to a unique solution. As a result of these calculations, the triangulation number of the phi92 icosahedron was found to be $T = 13$ dextro ($h = 3$, $k = 1$), resulting in 775 copies of the MCP per capsid. Interestingly, the capsids of the isometric T4 phage mutant and phage T5 were reported to have a mirrored icosahedral lattice with $T = 13$ laevo (18). Apparently, the HK97 fold has enough plasticity to accommodate both arrangements of the capsid subunits, which have the same T number.

The cryoEM map of the phi92 capsid displays prominent protrusions on the periphery of the capsomers near the local 3-fold axes (Fig. 4C). These protrusions represent either a domain which is not present in the HK97 capsid protein or an extra capsid decoration protein (Fig. 4D). The volume of the electron density suggests that each protrusion is formed by a polypeptide chain with a length of ~110 residues. As the phi92 MCP (gp124; Table 2) is only ~50 residues longer than the HK97 capsid protein, the num-



B

Band-#	Identified protein	Predicted function	Calc. MM [kDa]	App. MM [kDa]	Peptide matches	Sequence coverage [%]
1	gp123	Potential decoration protein	13.8	100	2	25.4
	gp141	Potential tail fiber protein	108	100	4	4.91
	gp143	Endosialidase 92 precursor (endoN92)	102	100	4	4.03
2	gp143*	Maturated endosialidase 92* (endoN92*)	83.9	85	8	8.93
	gp150	Potential colanidase	86.8	85	4	5.22
3§	gp142	Potential tail fiber protein	81.7	75	1	1.40
4	gp134	Potential tape measure protein	72.8	72	3	4.40
5	gp120	Potential portal protein	58.0	55	4	9.20
	gp145	Potential T4 baseplate gp6 ortholog	54.0	55	2	4.85
6	gp123	Potential decoration protein	13.8	50	2	25.4
	gp124	Major capsid protein (MCP)	37.1	50	2	4.81
	gp130	Potential tail sheath	51.1	50	10	21.6
7	gp123	Potential decoration protein	13.8	39	2	25.4
	gp124	Major capsid protein (MCP)	37.1	39	7	19.5
	gp130	Potential tail sheath	51.1	39	2	5.77
	gp137	Potential baseplate hub	37.0	39	3	8.33
8	gp124	Major capsid protein (MCP)	37.1	32	11	32.7
9§	gp123	Potential decoration protein	13.8	32	1	14.6
10	gp124	Major capsid protein (MCP)	37.1	30	11	36.6
	gp135	Potential tail tube initiator or completion protein	25.9	30	7	25.0
11	gp124	Major capsid protein (MCP)	37.1	30	10	30.9
	gp135	Potential tail tube initiator or completion protein	26.1	30	3	14.8
12	gp123	Potential decoration protein	13.8	28	3	30.0
	gp124	Major capsid protein (MCP)	37.1	28	10	31.5
13	gp123	Potential decoration protein	13.8	27	2	25.4
	gp124	Major capsid protein (MCP)	37.1	27	3	8.41
14	gp123	Potential decoration protein	13.8	25	2	25.4
15	gp124	Major capsid protein (MCP)	37.1	23	3	8.41
	gp146	Conserved phage structural protein	23.3	23	3	15.8
16§	gp123	Potential decoration protein	13.8	18.5	1	14.6
16§	gp126	Potential T4 baseplate gp15 ortholog	20.1	18.5	1	4.90
17	gp123	Potential decoration protein	13.8	17	2	25.4
18§	gp131	Potential tail tube	17.3	16	1	13.2
19§	gp123	Potential decoration protein	13.8	14	1	10.8

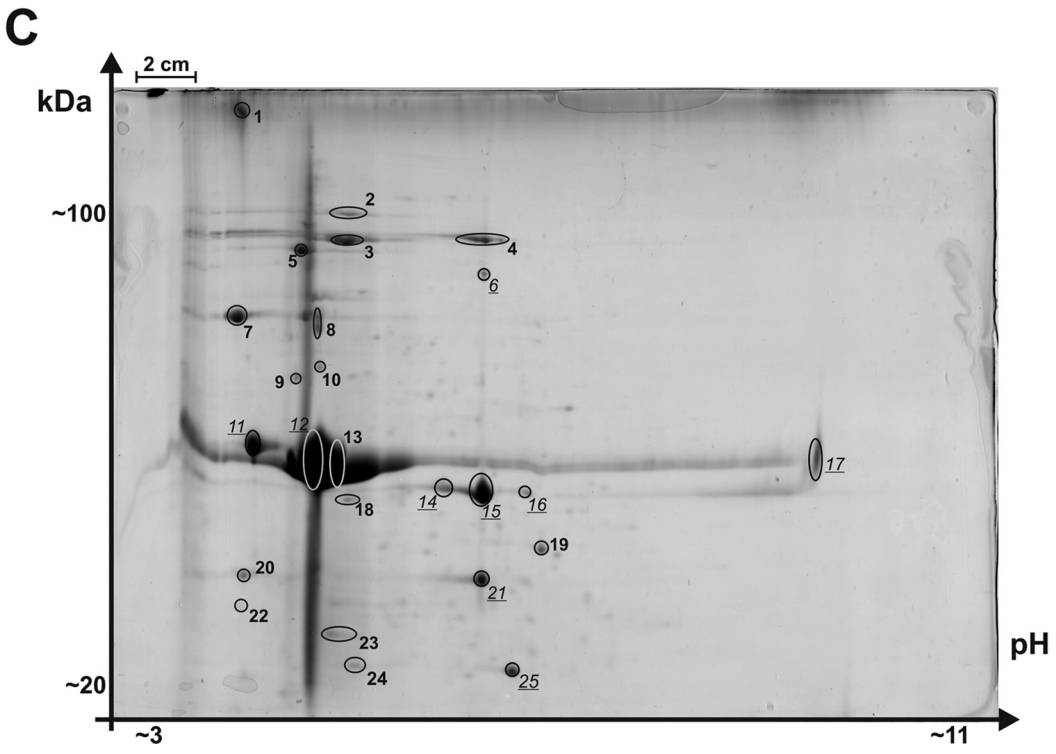


FIG 3 Identification of phi92 virion proteins. (A) 1-DE and MS analysis. Approximately 4 μg of phi92 protein was separated using 14% 1-DE. The marker bands (Precision Plus) on the left indicate the molecular masses. The bar represents 5 mm in the original scale of the gel. (B) The table summarizes the phi92 proteins described for panel A identified by MS. For proteins highlighted by a gray background, the calculated (Calc.) molecular mass (MM) values deviate less than 20% from the apparent (App.) molecular mass value determined using the 1-DE gel as described for panel A. §, proteins identified by single peptide matches. (C) 2-DE and MS analysis. Approximately 800 μg of phi92 protein was separated by isoelectric focusing (pI 3 to 11) followed by 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using 2-DE analysis. The original size of the gel was ~20 by 30 cm^2. The bar represents 2 cm in the original scale. Spots containing *E. coli* proteins are underscored and italicized. See also Table 2 and Dataset S2 in the supplemental material.

TABLE 2 Proteins identified by MS after isolation from 2-DE[a]

Spot	Identified protein	Predicted function	Calculated molecular mass [kDa]	Theoretical pI	No. of peptide matches	Sequence Coverage [%]
1	gp151	Potential tail fiber protein	69.0	4.4	5	9.0
2	gp141	Potential phage tail protein	108	5.1	12	13.2
3	gp150	Potential colanidase	86.8	5.0	14	19.5
4	gp143*	Maturated endosialidase 92* (endoN92*)	83.9	5.1	14	17.3
5	gp123	Phage structural protein	13.8	4.9	2	25.4
	gp124	Major capsid protein (MCP)	37.1	5.0	6	12.3
	gp142	Potential tail fiber protein	81.6	4.8	16	20.4
6	**OmpA**	**Outer membrane protein A**	37.2	6.3	8	24.0
7	gp145	Potential T4 baseplate gp6 ortholog	54.0	4.3	9	21.2
8	gp120	Potential portal protein	58.0	5.1	3	7.3
	gp123	Phage structural protein	13.8	4.9	2	25.4
	gp124	Major capsid protein (MCP)	37.1	5.0	12	24.0
9	gp123	Phage structural protein	13.8	4.9	2	25.4
	gp135	Potential tail tube initiator or completion protein	26.1	6.0	2	9.7
	LamB	**Maltoporin**	50.0	4.9	3	8.7
10	gp124	Major capsid protein (MCP)	37.1	5.0	10	21.6
11	gp124	Major capsid protein (MCP)	37.1	5.0	2	5.1
	OmpC	**Outer membrane protein C**	40.0	4.5	8	17.1
12	gp124	Major capsid protein (MCP)	37.1	5.0	10	36.3
	nmpC	**Outer membrane porin protein C**	39.6	4.9	5	13.9
13	gp124	Major capsid protein (MCP)	37.1	5.0	5	14.7
14	**OmpA**	**Outer membrane protein A**	37.2	6.3	10	26.6
15	gp124	Major capsid protein (MCP)	37.1	5.0	4	12.6
	OmpA	**Outer membrane protein A**	37.2	6.3	12	33.5
16	gp146	Conserved phage structural protein	23.3	4.9	2	9.6
	gp147	Potential tail fiber or connector protein	35.6	8.5	2	6.7
	OmpA	**Outer membrane protein A**	37.2	6.3	8	25.4
17	gp124	Major capsid protein (MCP)	37.1	5.0	2	5.1
	gp147	Potential tail fiber or connector protein	35.6	8.5	4	15.3
	OmpC	**Outer membrane porin protein C**	34.4	4.6	2	7.4
18	gp123	Phage structural protein	13.8	4.9	2	25.4
	gp124	Major capsid protein (MCP)	37.1	5.0	10	20.1
19	gp135	Potential tail tube initiator or completion protein	26.1	6.0	4	15.3
20	gp139	Potential T4 baseplate gp25 ortholog	23.6	4.3	2	10.4
	gp145	Potential T4 baseplate gp6 ortholog	54.0	4.3	4	9.1
21	**OmpA**	**Outer membrane protein A**	37.2	6.3	4	10.4
22	gp120	Potential portal protein	58.0	5.1	2	3.3
23	gp146	Conserved phage structural protein	23.3	4.9	8	26.8
24	gp126	Potential T4 gp15 ortholog	20.1	5.0	4	20.3
	gp145	Potential T4 baseplate gp6 ortholog	54.0	4.3	3	6.7
25	**OmpW**	**Outer membrane protein W**	22.9	6.4	3	18.9

[a] Only proteins identified by two or more peptide matches have been taken into account (Fig. 3C; see also Dataset S2 in the supplemental material). Peptides assigned to *E. coli* proteins are indicated in bold.

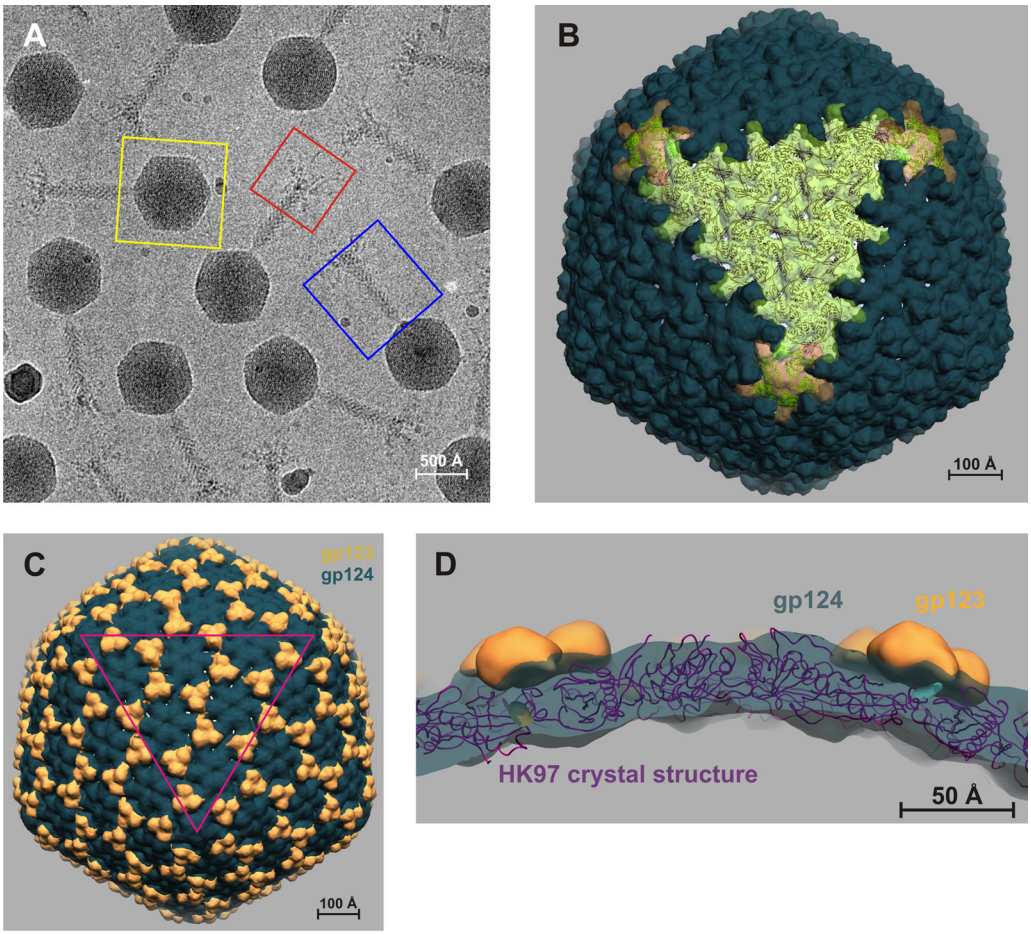


FIG 4 CryoEM structure of phi92. (A) A raw cryoEM image of phi92. Three separate reconstructions were calculated for areas highlighted with boxes colored in yellow, blue, and red. (B) The structure of phi92 capsid with the imposed icosahedral symmetry. On one face of the icosahedron, the HK97 capsid protein structure is fitted into the cryoEM density. (C) The location of the decoration protein gp123 (gold) on the capsid shell formed by gp124 (moss green). The magenta triangle highlights one face of the icosahedron. (D) Interpretation of the capsid cryoEM density with the help of the crystal structure of the HK97 capsid protein. In panels B, C, and D, the cryoEM map is contoured at 1.5 σ from the mean.

ber of additional amino acids is not sufficient to account for the density of all these protrusions. Furthermore, as identified with HHpred, the closest ortholog of the phi92 MCP with known structure is the MCP of a prophage from *E. coli* CTF073 (PDB ID 3BQW). This capsid protein attains an HK97 fold with a short insertion formed by residues 230 to 250. Hexamers and pentamers created by superimposing the 3BQW structure onto the HK97 capsomers showed that these extra residues were at the centers of the capsomers and faced inward. The phi92 cryoEM density can accommodate this insertion. Therefore, a separate decoration protein that is not part of the MCP must form the phi92 capsid protrusions (Fig. 4D).

The density corresponding to the decoration protein in the cryoEM map is as great as that of the MCP, suggesting that the ratio of the two proteins in the virion is close to 1.0. As there are 775 copies of the MCP in each virus particle, the decoration protein must form a prominent band on the 1-DE gel similar to those seen with other phages carrying capsid decoration proteins (42). The 1-DE analysis of the phi92 particle is, however, somewhat less clear in this respect (Fig. 3A). The protein corresponding best to the estimated molecular mass of 13.8 kDa was identified as gp123 of the 1-DE gel in band 19 (see Fig. 3B) but also in eight other bands (bands 1, 6, 7, 9, 12, 13, 14, and 17; see Fig. 3A and B). The apparent molecular mass of gp123 is 13.8 kDa, which is 12% greater than the molecular mass calculated from the cryoEM density. The observed difference can be attributed to the disorder and "averaging-out" effects, which are common to surface-exposed features in cryoEM reconstructions (21). The determination that the predicted secondary structure of gp123 is dominated by β-strands can be asserted with a high degree of confidence. Although the evidence is indirect, this observation supports the assignment of gp123 as the capsid decoration protein, since all known capsid decoration proteins and additional domains of MCPs are β-structural elements (5, 22, 51, 56, 59, 82).

Orthologs of phi92 decoration protein gp123 are found in phages rv5 and PVP-SE1 and are encoded by genes 61 and 72, respectively (Fig. 2A). In the absence of any additional evidence, gp72 of PVP-SE1 was annotated as a capsid decoration protein solely due to its abundance on the gel (63). The structure of the phi92 capsid obtained in this study confirms the function of the phi92_gp123 homolog PVP-SE1_gp72 and in addition strongly suggests that gp61 of phage rv5 is a capsid decoration protein. As

the capsids of rv5-like phages are fairly large, the conservation of the decoration protein must represent a functional requirement, because decoration proteins are known to increase the stability of virus capsids (31).

Capsid scaffolding proteins of most phages have two common features: they are rich in α-helical content and the genes encoding these proteins are found immediately upstream of the MCP genes in the genome (58). Both criteria apply to phi92 gp122, suggesting that it is the scaffolding protein. In phages rv5 and PVP-SE1, the equivalent genome positions are occupied by genes 62 and 73, respectively. These genes are annotated as scaffolding protein genes, albeit the reasons behind this annotation are unpublished. Neither displayed significant sequence similarity to phi92 gp122, but they are of similar sizes and have predominantly α-helical structures.

The structure of the head-to-tail region was investigated by calculating the 5-fold symmetric reconstruction of the capsid and at the same time increasing the reconstructed volume to include a proximal part of the tail. Compared to the icosahedrally averaged reconstruction, the 5-fold symmetric map of the capsid did not show any new features and was of lower quality. Nevertheless, the 5-fold averaged reconstruction demonstrated that, unlike the extensive neck and whiskers structures found in bacteriophage T4, the phi92 particle carries no collar (41).

Genes located between the capsid and tail gene clusters are likely to participate in the attachment of the tail to the capsid by expressing either a structural protein of the neck or a chaperone (16). One such gene is gene 126 (Fig. 2A), and gp126 was found in the phage particle by MS (Fig. 3, Table 2). The genome location, secondary structure, and molecular mass of gp126 suggest that it is a tail-capping protein (Fig. 5) and most likely an ortholog of T4 gp15 (41).

Structure of the tail. The contractile sheath and the central tube are two major components of the phage tail. The sheath has an external diameter of 220 Å and is 840 Å long (Fig. 5). The sheath is a six-start superhelix, with each helical strand containing 24 copies of the gp130 sheath protein, resulting in a total of 144 gp130 subunits per tail. The helical rise is 38.9 Å, and the twist is 26.0 degrees. In comparison, the T4 sheath is also a six-start superhelix with 23 subunits per strand, but the parameters of the helix are slightly different: the rise is 40.6 Å and the twist is 17.2 degrees (35). This gives the two tails almost the same length and most probably results in the two sheaths contracting and extruding the tail tube from the plane of the baseplate upon attachment to the host cell surface by the same amount. Again, similar to the observations of the T4 tail, the symmetry of the tail tube is not apparent in the reconstruction and its density partially overlaps that of the sheath (35). Nevertheless, the phi92 tail tube is in all likelihood also a six-start helix assembled of hexamers of the gp131 tube protein, as is found in all related contractile tail-like systems (44, 75).

Similar to phage capsid proteins, contractile sheath proteins have a common ancestor and appear to be built using a common conserved structure consisting of three domains with a total molecular mass of ~45 kDa (44). Some sheath proteins such as those of T4 and its close relatives contain an additional surface-exposed domain and have a greater molecular mass. The 52-kDa phi92 sheath protein gp130 is likely to have the conserved three-domain organization, which also characterizes the sheath proteins of phages Mu and P2 (44). None of the available crystal structures of sheath proteins shows a satisfactory fit into the phi92 tail cryoEM map. The relative orientations of the three constituent domains are likely different from those found in the crystal structures. The resolution of the cryoEM map is insufficient to fit the domains separately.

The length of the tail is determined by the size of the "tape measure" protein, which is thought to form an alphahelical structure which extends through the entire central channel of the tail tube (16). The tape measure proteins of phi92 (gp134) and T4 (gp29) are 659 and 590 residues long, respectively, which agrees with the similar lengths of the tails. In the two related phages rv5 and PVP-SE1, the proposed tape measure proteins (gp49 and gp57) are significantly longer (778 aa and 795 aa, respectively) than gp134 of phi92 (659 aa), suggesting that the tails of the rv5 and PVP-SE1 phages might be longer than that of phi92.

Structure of the phi92 baseplate. The phi92 baseplate with emanating fibers and tailspikes has a complex structure and resembles an open Swiss army knife (Fig. 5). The bioinformatic analysis outlined above shows that the genes comprising the core part of the phi92 baseplate—139, 145, and 146—are orthologous to those of phages Mu and P2, although slightly larger (e.g., phi92 gp145 is 54 kDa compared to P2 gpJ's 34 kDa) (Fig. 2A; see also Fig. S1 in the supplemental material).

phi92 gp138 and gp137 are orthologs of the gp5-gp27 cell-puncturing complex of phage T4 (34, 44) and thus must form the "cell-puncturing device" or the central hub of the baseplate (34). Although not identified by MS, gp138 is part of the phage particle and constitutes the baseplate central spike protein that is essential for breaching the host's outer membrane during infection (9). There are only 3 copies of gp138 polypeptides per particle, and they form a densely folded structure resistant to SDS and proteases, which might have prevented its identification in the MS analysis. gp137 forms a stable trimer in solution which is likely to have the same fold and donut-like shape as Mu gp44, T4 gp27, and many other central hub proteins, including those from noncontractile tail phages (16, 44). Therefore, the ring-shaped density immediately on top of gp138 in the cryoEM map can be attributed to gp137. Unlike T5 gp5 and gp27 (34), the trimers of phi92 gp137 and gp138 form a weak complex in solution.

RBPs. A combination of bioinformatics, MS, and cryoEM analyses showed that each phi92 particle carries at least five different tailspike and tail fiber proteins (or receptor binding proteins [RBPs]), gp141, gp142, gp143, gp150, and gp151, which are located at the 3′ end of the tail operon. Three sets of RBPs emanating from the baseplate can be distinguished in the cryoEM map. They are directed downward, sideward, and upward (fibers 1, 2, and 3 in Fig. 5). The downward-pointing fibers have the strongest cryoEM density, whereas those pointing upward have the weakest. The cryoEM map is a representation of all particle images that were included in the 3D reconstruction procedure. As a consequence, the RBPs pointing downward are likely to have the most similar orientations in all particles whereas those pointing upward are likely to have the most dissimilar and thus are "averaged out" in the final 3D maps. It is also possible that each particle carries a full complement of the downward-pointing RBPs but only a few of those pointing sideward and even fewer of those pointing upward. The latter hypothesis is less likely, as it has never been reported for any other virus or phage. The quality of the cryoEM map does not allow reliable identification of the gps comprising these RBPs. Nevertheless, the genome of phi92 contains enough gps to account for all the RBPs, with each having a full occupancy per particle (Fig. 2A).

The RBP regions of the phi92, PVP-SE1, and rv5 genomes

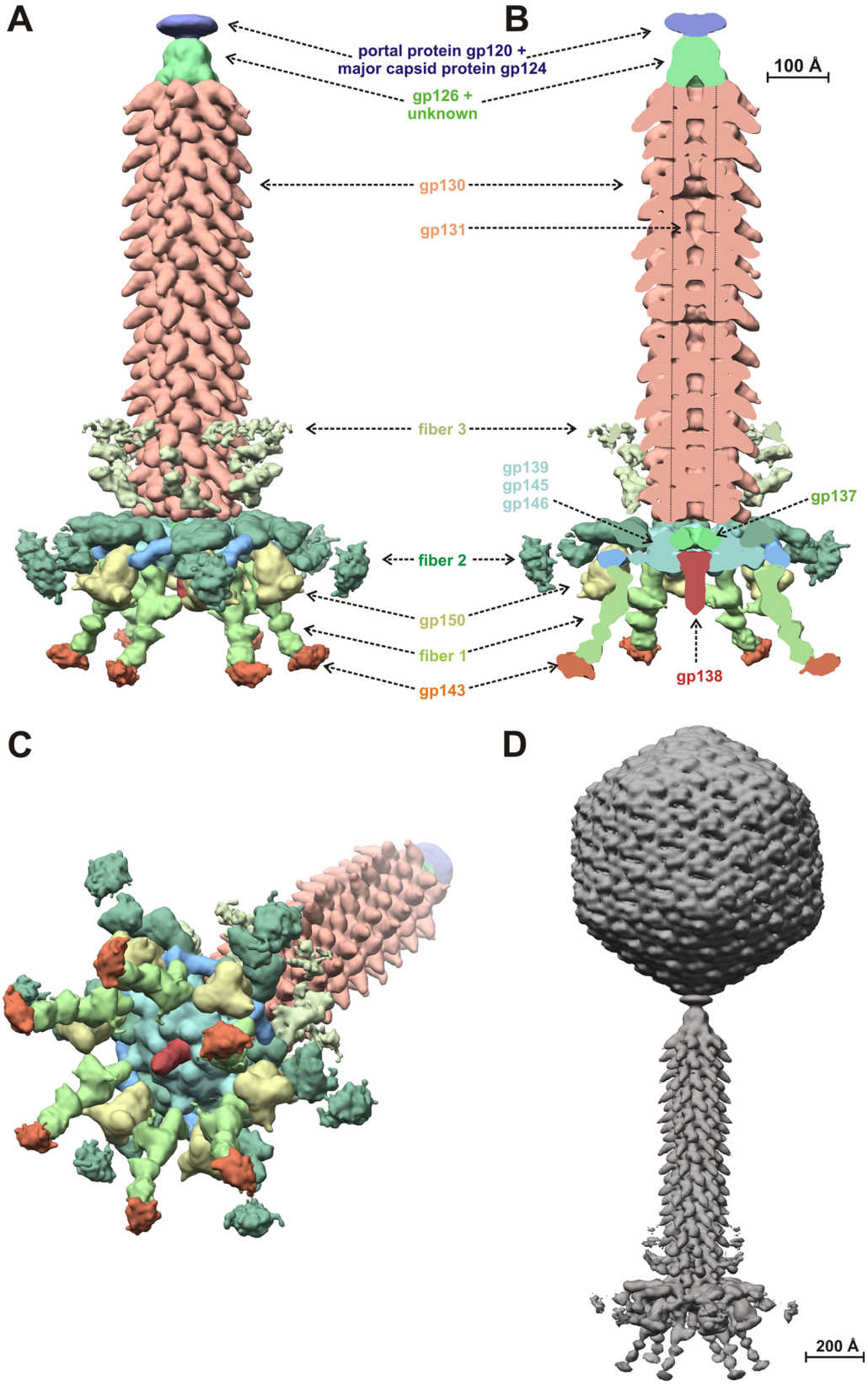


FIG 5 Structure of phi92 tail and baseplate. (A and B) Side and cutaway views, respectively, of the phi92 tail structure. A composite map, which includes two overlapping maps (tail sheath and the baseplate) produced separately, is shown. The map is interpreted in terms of the component proteins that were known. (C) A tilted view of the tail colored identically to the views in panels A and B. (D) The structure of phage phi92 produced by combining three reconstructions (the 5-fold-averaged capsid, the 6-fold-averaged tail, and the 6-fold-averaged baseplate). In panels A, B, and C, the cryoEM maps are contoured at 1.5 σ from the mean.

show the highest degree of mosaicism compared to the rest of the structural operon (Fig. 2A), a feature common to all groups or families of phages (13). All known RBPs have a modular architecture and consist of up to three different parts: (i) a small N-terminal domain or peptide for attachment of the RBP to the baseplate or tail region; (ii) a large central part that forms an extended trimeric structure with a coiled coil or beta-helical fold and is the key factor for host specificity; and (iii) a small C-terminal domain that usually functions as an intramolecular chaperone. The three parts of an RBP might be encoded by different ORFs or belong to a single polypeptide chain. In the three rv5-like phages, the N-terminal and C-terminal parts of several RBPs are significantly better conserved than their central parts (Fig. 2A). For example, the N-terminal regions of phi92 gp150 and gp151 show similarity to those of rv5 gp28 and gp41, respectively, and the C-terminal parts of rv5 gp28 and phi92 gp151 are also similar at the amino acid level (see Fig. 2A; see also Table S1 in the supplemental material). Phages are able to exchange the host-specific modules, while maintaining the N-terminal tail attachment domain unchanged, to increase their host range or to adapt to a new environment (73). The C-terminal intramolecular chaperones of some RBPs are known to facilitate folding of foreign RBPs (52, 66, 68, 81).

phi92 infects K1 and K92 strains of *E. coli* (38), which are encapsulated by polySia with α2,8- or alternating α2,8/α2,9-linkages, respectively (4, 25, 48). The polySia capsule-digesting tailspike of the phi92 particle is likely to be encoded by gene 143. It displays 54% sequence identity to the well-studied endosialidase of coliphage K1F (endoNF) through the whole length of the protein, suggesting that gp143 has the same structure as the mushroom-shaped endoNF (52, 72). By analogy with other endosialidases (endoNF, endoNE, etc.), we refer to it as endoN92. It is located at the distal end of the downward-pointing fibers in the cryoEM map (Fig. 5). This fiber ends in a bulge that corresponds to the sialidase domain of endoN92. The narrow stalk domain of endoNF forms the tip of the 650 Å-long fiber and thus is disordered (Fig. 5).

At present, the other putative fibrous proteins (e.g., gp141, gp142, and gp151) cannot be located in the cryoEM map without additional information regarding their structure or interactions with other proteins (fibers 1 to 3 in Fig. 5A). However, gene 150 of phi92 is likely to encode a tailspike protein and the cryoEM map contains a density that might correspond to gp150. HHpred analysis shows that gp150 is likely to have the trimeric β-helical fold that characterizes many other phage tailspikes (e.g., phi29, P22, Sf6, HK620, or K5A), and therefore is likely to have a similar compact and trimeric shape (43). The cryoEM map contains a tailspike-like density with a clear 3-fold symmetry and a volume sufficient to account for three polypeptide chains of gp150 (Fig. 5).

The sequence of gp150 shows similarity to that of a colanic acid-degrading enzyme (see Dataset S1 in the supplemental material) and, in analogy to other glycosidases, can be referred to as "colanidase." The phi92 host strain *E. coli* K92 produces colanic acid at lower temperatures than the K92 polySia (53), and one of the possible functions of gp150 is likely that of breaching the colanic acid layer. Interestingly, the catalytic domain of gp49 from phage PVP-SE1 has a significant sequence similarity to that of gp150, whereas its N-terminal region is related to the N-terminal part of phi92 gp151 (Fig. 2A), suggesting that the colanidase might be attached to the PVP-SE1 baseplate at a site different from that of phi92 gp150. It is possible that the phi92 colanidase attachment site in PVP-SE1 is occupied by gp41, since their N-terminal domains share high sequence similarity (64.2%; see Table S2 in the supplemental material).

phi92 host range. Considering the apparent polyvalence of the remarkable phi92 baseplate structure, we decided to reinvestigate the host range of phi92 using several bacterial strains, including laboratory strains of *E. coli* and a large set of *Salmonella* strains. We compared these results with the host range of the well-studied K1-specific phage K1F (Table 3). K1F contains only endoNF as a tailspike protein, and its replication is strictly limited to hosts that are encapsulated by polySia (Table 3) (65). phi92, in contrast, is able to infect nonencapsulated *E. coli* strains such as, for instance, EV5 (a capsule-defective derivative of a K1 strain [76]) and many laboratory K-12 and B strains with different plating efficiencies (Table 3). Thus, at least one of the several putative phi92 RBPs binds to a receptor exposed directly on the outer bacterial membrane.

Interestingly, phi92 infects *E. coli* BL21-Gold(DE3) but not BL21(DE3). One important difference in the two genotypes is the endonuclease gene (*endA1*), which is absent in the BL21-Gold(DE3) genome. The *endA1* gene is also absent from K-12 strain JM105 (also infected by phi92) but present in YZ2000 (not infected). Therefore, the phi92 genome is likely to be sensitive to the *endA1* endonuclease.

Besides many *E. coli* strains, phi92 infected different *Salmonella* serovars with distinct O-antigens. The 19 tested serovars were infected with similar efficiencies, strongly suggesting that phi92 infection does not depend on a particular O-antigen (Table 3). Whereas phi92 formed clear plaques on *E. coli*, phi92 infection of *Salmonella* strains resulted in opaque lysis (ol), a lysis pattern which has been previously described for several myoviruses (17). Phage PVP-SE1, the closest relative of phi92, has been shown to infect many *Salmonella* strains as well as several *E. coli* strains (62). The PVP-SE1 genome contains six different tail fiber and tailspike genes, four of which show a significant sequence similarity to phi92 tail genes, making it difficult to explicitly assign host specificities to elements of the phage adsorption apparatus (Fig. 2A).

The phi92 host range includes the nonpathogenic *E. coli* Nissle 1917—a K5 strain, which was isolated during World War I from a patient with resistance to severe diarrhea (54). K5 strains are encapsulated by poly(β1,4-GlcA-α1,4-GlcNAc) sugar polymers, and phages that infect such strains carry tailspike enzymes with K5 lyase activity (14). However, the phi92 genome does not contain a K5 lyase gene. In agreement with this, phi92 does not infect a pathogenic K5 strain (Table 3). The observed discrepancy may be explained by the density of the capsule, which in Nissle 1917 may be lower than in other K5 strains, thus allowing phi92 to access the membrane without the need for capsule degradation.

Concluding remarks. This report describes the genome, structure, and host range of the *Enterobacteria* phage phi92. The phi92 genome displays mosaicism similar to that of other phages and was found to be closely related to the genomes of phages rv5 and PVP-SE1, favoring suggestions to group them into the new genus of rv5-like phages. The structural operons are well conserved, strongly suggesting that the structures of the capsid (including the location of the decoration protein) and the tail and the organization of the major part of the baseplate are conserved. The baseplates of rv5 and PVP-SE1 are likely to have a similar Swiss army knife-like structure, although the numbers of tail fibers and tailspikes emanating from the baseplate and their exact attachment sites on the baseplate might differ.

TABLE 3 Plating efficiencies of isolated bacteriophages phi92 and K1F for different bacterial strains

Strain and description[a]	Titer (%)[b]		RKI strain	Source or reference(s)
	phi92	K1F		
Escherichia coli				
K92 (O16:K92:H$^-$)	100	100		ATCC 35860
K-235 (O1:K1:H$^-$)	40	400		37
K1 B2032/82	80	180		24
K1 U9/41 (O2:K1:H4)	4	40		28
EV5 (K1$^-$)—deficient in K1 capsule synthesis	150	—		76
BL21-Gold (DE3)—B F$^-$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA Hte*	310	—		Stratagene
BL21(DE3)—B F$^-$ *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3)	—	—		Stratagene
K-12 (JM105)—*endA*$^-$	280	—		Pharmacia
YZ2000	—	—		GeneBridges (83)
Nissle 1917 (O6:K5:H1)	100	—		8, 54
K5—heparosan capsule	—	—		68
Salmonella serovar				
S. subsp. I Enteritidis 9,12:g,m (PT21/1b)	100 (ol)	—	11-07263	
S. subsp. I 9,12:l,v:-	100 (ol)	—	11-07263	
	100 (ol)	—	11-07037	
S. subsp. I Chailey 6,8:z_4,z_{23}	100 (ol)	—	11-08049	
S. subsp. I Infantis 6,7:r:1,5 (PT29)	100 (ol)	—	05-02124	
S. subsp. I Kintambo 13,23:m,t:-	100 (ol)	—	11-06665	
	100 (ol)	—	11-01214	
S. subsp. I Kottbus 6,8:e,h:1,5	100 (ol)	—	11-01222	
S. subsp. I Newport 6,8:e,h:1,2	100 (ol)	—	04-00225	
S. subsp. I Ohio 6,7:b:l,w	100 (ol)	—	98-08252	
S. subsp. I Poano 1,6,14,25:z:l,z_{13},z_{28}	100 (ol)	—	11-07117	
S. subsp. I Pomona 28:y:1,7	100 (ol)	—	11-01216-2	
S. subsp. I Tennessee 6,7:z_{29}	100 (ol)	—	03-00230	
	100 (ol)	—	08-00724	
	100 (ol)	—	08-02001	
S. subsp. I Typhi 9,12:d:- (Vi-PT E1a)	100 (ol)	—	08-02103	
S. subsp. I Typhimurium 4,12:i:1,2 (DT1)	100 (ol)	—	11-06612	
S. subsp. I Typhimurium 4,12:i:1,2 (DT 120)	100 (ol)	—	04-00287	
S. subsp. II 42:z:1,5	100 (ol)	—	11-07007	
S. subsp. IIIb 48:k:z53	100 (ol)	—	11-7900	
S. subsp. IIIb 50:k:z	100 (ol)	—	06-02251	
S. subsp. IV 44:z4,z32:-	100 (ol)	—	11-06873	
Serological rough	100 (ol)	—	99-00210	
Other bacteria			04-00224	
Pseudomonas putida	—	—		
B. subtilis DSM 5545	—	—	DSMZ 50906	
B. pumilus G1R	—	—	DSMZ 5545	
			DSMZ 5549	

[a] *Salmonella* species designations and antigen and phage data are indicated as follows: O-antigen: phase 1 H-antigen: phase 2 H-antigen (phage type). Tet, tetracycline.
[b] Values represent relative titers compared to the original strain *E. coli* K92 or *S.* Enteritidis. (ol), opaque lysis of *Salmonella* strains. —, absence of a phage lysis halo or phage plaques at the highest concentrations of 10^{10} PFU/ml.

This report illustrates that phages display an enormous potential to change host recognition proteins by intra- and interspecies recombination and by mutations, resulting in their ability to recognize new host receptors. This phenomenon plays an important role in the natural ecology of phage-host interactions and can be exploited in the selection of therapeutic phages having desired "target ranges."

ACKNOWLEDGMENTS

This study was supported by funding from EPFL to the Laboratory of Structural Biology and Biophysics and by the Deutsche Forschungsgemeinschaft (DFG, Bonn, Germany) in the framework of DFG Research Unit 548 (Ge801/7-1 and Ge801/7-2).

Franziska Begett (RKI Wernigerode) and Florian Tönsing (Hannover Medical School) are acknowledged for expert technical assistance. Hazel Leanne Sarah Fuchs (Institut für Zelluläre Chemie, Hannover Medical School) is acknowledged for carefully reading the manuscript. Kristina Kampmann (GATC, Konstanz, Germany) is acknowledged for help with the sequencing. Ana M. Cerdeño-Tárraga and the European Nucleotide Archive EMBL-Bank Curation Team (EBI, Hinxton, Cambridge, United Kingdom) are gratefully acknowledged for support with the genome annotation.

REFERENCES

1. **Altschul SF, et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

2. **Altschul SF, et al.** 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J. **272**:5101–5109.
3. **Baker TS, Henderson R.** 2006. Electron cryomicroscopy, p 451–463. *In* International tables for crystallography, vol. F. Springer, New York, NY.
4. **Barry GT, Goebel WF.** 1957. Colominic acid, a substance of bacterial origin related to sialic acid. Nature **179**:206.
5. **Bateman A, Eddy SR, Mesyanzhinov VV.** 1997. A member of the immunoglobulin superfamily in bacteriophage T4. Virus Genes **14**:163–165.
6. **Belfort M.** 1990. Phage T4 introns: self-splicing and mobility. Annu. Rev. Genet. **24**:363–385.
7. **Besemer J, Lomsadze A, Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. **29**:2607–2618.
8. **Blum G, Marre R, Hacker J.** 1995. Properties of Escherichia coli strains of serotype O6. Infection **23**:234–236.
9. **Browning C, Shneider MM, Bowman VD, Schwarzer D, Leiman PG.** 2012. Phage pierces the host cell membrane with the iron-loaded spike. Structure **20**:326–339.
10. **Buchan DW, et al.** 2010. Protein annotation and modelling servers at University College London. Nucleic Acids Res. **38**:W563–W568.
11. **Buettner FF, Konze SA, Maas A, Gerlach GF.** 2011. Proteomic and immunoproteomic characterization of a DIVA subunit vaccine against Actinobacillus pleuropneumoniae. Proteome Sci. **9**:23. doi:10.1186/1477-5956-9-23.
12. **Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J.** 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics **25**:119–120.
13. **Casjens SR, Molineux IJ.** 2012. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv. Exp. Med. Biol. **726**:143–179.
14. **Clarke BR, Esumeh F, Roberts IS.** 2000. Cloning, expression, and purification of the K5 capsular polysaccharide lyase (KflA) from coliphage K5A: evidence for two distinct K5 lyase enzymes. J. Bacteriol. **182**:3761–3766.
15. **Cross AS.** 1990. The biologic significance of bacterial encapsulation. Curr. Top. Microbiol. Immunol. **150**:87–95.
16. **Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL.** 2012. Long noncontractile tail machines of bacteriophages. Adv. Exp. Med. Biol. **726**:115–142.
17. **De Lappe N, Doran G, O'Connor J, O'Hare C, Cormican M.** 2009. Characterization of bacteriophages used in the Salmonella enterica serovar Enteritidis phage-typing scheme. J. Med. Microbiol. **58**:86–93.
18. **Effantin G, Boulanger P, Neumann E, Letellier L, Conway JF.** 2006. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. J. Mol. Biol. **361**:993–1002.
19. **Egan W, et al.** 1977. Structural studies on the sialic acid polysaccharide antigen of Escherichia coli strain Bos-12. Biochemistry **16**:3687–3692.
20. **Emanuelsson O, Brunak S, von Heijne G, Nielsen H.** 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. **2**:953–971.
21. **Fokine A, et al.** 2004. Molecular architecture of the prolate head of bacteriophage T4. Proc. Natl. Acad. Sci. U. S. A. **101**:6003–6008.
22. **Fokine A, et al.** 2005. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. Proc. Natl. Acad. Sci. U. S. A. **102**:7163–7168.
23. **Frank J, et al.** 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. **116**:190–199.
24. **Frosch M, et al.** 1987. Serotyping and genotyping of encapsulated Escherichia coli K1 sepsis isolates with a monoclonal IgG anti K1 antibody and K1 gene probes. Microb. Pathog. **2**:319–326.
25. **Furowicz AJ, Orskov F.** 1972. Two new Escherichia coli O antigens, O150 and O157, and one new K antigen, K92, in strains isolated from veterinary diseases. Acta Pathol. Microbiol. Scand. B Microbiol. Immunol. **80**:441–444.
26. **Gerardy-Schahn R, et al.** 1995. Molecular cloning and functional expression of bacteriophage PK1E-encoded endoneuraminidase Endo NE. Mol. Microbiol. **16**:441–450.
27. **Grigorieff N.** 2007. FREALIGN: high-resolution refinement of single particle structures. J. Struct. Biol. **157**:117–125.
28. **Gross RJ, Cheasty T, Rowe B.** 1977. Isolation of bacteriophages specific for the K1 polysaccharide antigen of Escherichia coli. J. Clin. Microbiol. **6**:548–550.
29. **Hatfull GF.** 2008. Bacteriophage genomics. Curr. Opin. Microbiol. **11**:447–453.
30. **Hyatt D, et al.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. doi:10.1186/1471-2105-11-119.
31. **Iwasaki K, et al.** 2000. Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein, Soc. Virology **271**:321–333.
32. **Jakobsson E, Schwarzer D, Jokilammi A, Finne J.** 1 August 2012. Endosialidases: versatile tools for the study of polysialic acid. Top. Curr. Chem. doi:10.1007/128_2012_349.
33. **Jones DT.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292**:195–202.
34. **Kanamaru S, et al.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature **415**:553–557.
35. **Kostyuchenko VA, et al.** 2005. The tail structure of bacteriophage T4 and its mechanism of contraction. Nat. Struct. Mol. Biol. **12**:810–813.
36. **Kostyuchenko VA, et al.** 2003. Three-dimensional structure of bacteriophage T4 baseplate. Nat. Struct. Biol. **10**:688–693.
37. **Kwiatkowski B, Boschek B, Thiele H, Stirm S.** 1982. Endo-N-acetylneuraminidase associated with bacteriophage particles. J. Virol. **43**:697–704.
38. **Kwiatkowski B, Boschek B, Thiele H, Stirm S.** 1983. Substrate specificity of two bacteriophage-associated endo-N-acetylneuraminidases. J. Virol. **45**:367–374.
39. **Kwiatkowski B, Stirm S.** 1987. Polysialic acid depolymerase. Methods Enzymol. **138**:786–792.
40. **László VG, Csorian ES, Paszti J.** 1985. Phage types and epidemiological significance of Salmonella enteritidis strains in Hungary between 1976 and 1983. Acta Microbiol. Hung. **32**:321–340.
41. **Leiman PG, et al.** 2010. Morphogenesis of the T4 tail and tail fibers. Virol. J. **7**:355.
42. **Leiman PG, et al.** 2007. The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. J. Mol. Biol. **371**:836–849.
43. **Leiman PG, Molineux IJ.** 2008. Evolution of a new enzyme activity from the same motif fold. Mol. Microbiol. **69**:287–290.
44. **Leiman PG, Shneider MM.** 2012. Contractile tail machines of bacteriophages. Adv. Exp. Med. Biol. **726**:93–114.
45. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.
46. **Ludtke SJ, Baldwin PR, Chiu W.** 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. **128**:82–97.
47. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26**:1107–1115.
48. **McGuire EJ, Binkley SB.** 1964. The structure and chemistry of colominic acid. Biochemistry **3**:247–251.
49. **Miller ES, et al.** 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. **67**:86–156.
50. **Mindell JA, Grigorieff N.** 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. **142**:334–347.
51. **Morais MC, et al.** 2005. Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. Mol. Cell **18**:149–159.
52. **Mühlenhoff M, Stummeyer K, Grove M, Sauerborn M, Gerardy-Schahn R.** 2003. Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. J. Biol. Chem. **278**:12634–12644.
53. **Navasa N, Rodriguez-Aparicio L, Martinez-Blanco H, Arcos M, Ferrero MA.** 2009. Temperature has reciprocal effects on colanic acid and polysialic acid biosynthesis in E. coli K92. Appl. Microbiol. Biotechnol. **82**:721–729.
54. **Nissle A.** 1918. Die antagonistische Behandlung chronischer Darmstörungen mit Colibakterien. Med. Klin. (Munich) **2**:29–33.
55. **Parent KN, Gilcrease EB, Casjens SR, Baker TS.** 2012. Structural evolution of the P22-like phages: comparison of Sf6 and P22 procapsid and virion architectures. Virology **427**:177–188.
56. **Parent KN, et al.** 2010. P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. Structure **18**:390–401.

57. **Pettersen EF, et al.** 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. **25**:1605–1612.
58. **Prevelige PE, Fane BA.** 2012. Building the machines: scaffolding protein functions during bacteriophage morphogenesis. Adv. Exp. Med. Biol. **726**: 325–350.
59. **Qin L, Fokine A, O'Donnell E, Rao VB, Rossmann MG.** 2010. Structure of the small outer capsid protein, Soc: a clamp for stabilizing capsids of T4-like phages. J. Mol. Biol. **395**:728–741.
60. **Rabsch W.** 2007. *Salmonella* Typhimurium phage typing for pathogens, p 177–211. *In* Schatten H, Eisenstark A (ed), Methods in molecular biology. Salmonella and protocols. Humana Press, Totowa, NJ.
61. **Sambrock J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
62. **Santos SB, et al.** 2010. Selection and characterization of a multivalent Salmonella phage and its production in a nonpathogenic Escherichia coli strain. Appl. Environ. Microbiol. **76**:7338–7342.
63. **Santos SB, et al.** 2011. Genomic and proteomic characterization of the broad host range Salmonella phage PVP-SE1—the creation of a new phage genus. J. Virol. **85**:11265–11273.
64. **Scholl D, Adhya S, Merril C.** 2005. Escherichia coli K1's capsule is a barrier to bacteriophage T7. Appl. Environ. Microbiol. **71**:4872–4874.
65. **Scholl D, Merril C.** 2005. The genome of bacteriophage K1F, a T7-like phage that has acquired the ability to replicate on K1 strains of Escherichia coli. J. Bacteriol. **187**:8499–8503.
66. **Schulz EC, et al.** 2010. Crystal structure of an intramolecular chaperone mediating triple-beta-helix folding. Nat. Struct. Mol. Biol. **17**:210–215.
67. **Schulz EC, et al.** 2010. Structural basis for the recognition and cleavage of polysialic acid by the bacteriophage K1F tailspike protein EndoNF. J. Mol. Biol. **397**:341–351.
68. **Schwarzer D, Stummeyer K, Gerardy-Schahn R, Mühlenhoff M.** 2007. Characterization of a novel intramolecular chaperone domain conserved in endosialidases and other bacteriophage tail spike and fiber proteins. J. Biol. Chem. **282**:2821–2831.
69. **Söding J.** 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics **21**:951–960.
70. **Söding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33**:W244–W248.
71. **Stirm S, Bessler W, Fehmel F, Freund-Mölbert E.** 1971. Bacteriophage particles with endo-glycosidase activity. J. Virol. **8**:343–346.
72. **Stummeyer K, Dickmanns A, Mühlenhoff M, Gerardy-Schahn R, Ficner R.** 2005. Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. Nat. Struct. Mol. Biol. **12**:90–96.
73. **Stummeyer K, et al.** 2006. Evolution of bacteriophages infecting encapsulated bacteria: lessons from *Escherichia coli* K1 specific phages. Mol. Microbiol. **60**:1123–1135.
74. **Taylor CM, Roberts IS.** 2005. Capsular polysaccharides and their role in virulence. Contrib. Microbiol. **12**:55–66.
75. **Veesler D, Cambillau C.** 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol. Mol. Biol. Rev. **75**:423–433.
76. **Vimr ER, Troy FA.** 1985. Regulation of sialic acid metabolism in Escherichia coli: role of N-acylneuraminate pyruvate-lyase. J. Bacteriol. **164**: 854–860.
77. **Ward LR, de Sa JD, Rowe B.** 1987. A phage-typing scheme for Salmonella enteritidis. Epidemiol. Infect. **99**:291–294.
78. **Whichard J, et al.** 2010. Complete genomic sequence of bacteriophage Felix O1. Viruses **2**:710–730.
79. **Whitfield C.** 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annu. Rev. Biochem. **75**:39–68.
80. **Wikoff WR, et al.** 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science **289**:2129–2133.
81. **Xiang Y, et al.** 2009. Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Mol. Cell **34**:375–386.
82. **Yang F, et al.** 2000. Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat. Struct. Biol. **7**:230–237.
83. **Zhang Y, Muyrers JP, Testa G, Stewart AF.** 2000. DNA cloning by homologous recombination in Escherichia coli. Nat. Biotechnol. **18**: 1314–1317.