

# Phage Pierces the Host Cell Membrane with the Iron-Loaded Spike

Christopher Browning,<sup>1</sup> Mikhail M. Shneider,<sup>1,2</sup> Valorie D. Bowman,<sup>3</sup> David Schwarzer,<sup>1,4</sup> and Petr G. Leiman<sup>1,\*</sup>

<sup>1</sup>Institut de Physique des Systèmes Biologiques, Laboratory of Structural Biology and Biophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), BSP-415, CH-1015, Lausanne, Switzerland

<sup>2</sup>Laboratory of Molecular Bioengineering, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 117997, Moscow, GSP-7, Ul. Miklukho-Maklaya, 16/10, Russia

<sup>3</sup>Department of Biological Sciences, Purdue University, 915 West State Street, West Lafayette, IN 47907-2054 USA

<sup>4</sup>Present address: Medizinische Hochschule Hannover, Institut für Zelluläre Chemie (OE 4330) Carl-Neuberg-Straße 1, D-30625 Hannover, Germany

\*Correspondence: [petr.leiman@epfl.ch](mailto:petr.leiman@epfl.ch)

DOI 10.1016/j.str.2011.12.009

## SUMMARY

Bacteriophages with contractile tails and the bacterial type VI secretion system have been proposed to use a special protein to create an opening in the host cell membrane during infection. These proteins have a modular architecture but invariably contain an oligonucleotide/oligosaccharide-binding (OB-fold) domain and a long  $\beta$ -helical C-terminal domain, which initiates the contact with the host cell membrane. Using X-ray crystallography and electron microscopy, we report the atomic structure of the membrane-piercing proteins from bacteriophages P2 and  $\phi$ 92 and identify the residues that constitute the membrane-attacking apex. Both proteins form compact spikes with a  $\sim 10\text{\AA}$  diameter tip that is stabilized by a centrally positioned iron ion bound by six histidine residues. The accumulated data strongly suggest that, in the process of membrane penetration, the spikes are translocated through the lipid bilayer without undergoing major unfolding.

## INTRODUCTION

Bacteriophages with contractile tails, R-type pyocins, the *Serratia entomophila* antifeeding prophage, the *Photobacterium* virulence cassette, and the type VI secretion system (T6SS) of Gram-negative bacteria belong to the recently defined class of contractile injection systems based on their apparent shared ancestry (Bönemann et al., 2010; Leiman and Shneider, 2012; Pukatzki et al., 2007). A contractile injection system is a complex macromolecular machine composed of hundreds of protein subunits of 10–20 different types that form the contractile sheath, the internal noncontractile tube and the baseplate (Leiman and Shneider, 2012). The machine is able to penetrate the host cell envelope and to deliver specific proteins and/or DNA into the cell.

Most of the information regarding the function and structure of proteins involved in host attachment, as well as their location in

the machine for all contractile injection systems, comes from studies of the phage T4 tail (Aksyuk et al., 2009; Kostyuchenko et al., 2005, 2003; Leiman et al., 2010, 2004). It was proposed that a special spike-shaped protein complex, which forms an extension of the tail tube (the “cell-puncturing device”), is used to pierce the outer cell membrane in the process of tail sheath contraction (Kanamaru et al., 2002). The cell-puncturing device is formed by T4 gene product (gp) 27, gp5 and a “tip protein,” which caps the C-terminal  $\beta$ -helical needle domain of gp5 and is responsible for the first contact with the host cell membrane (Kostyuchenko et al., 2003). Despite the availability of the detailed description for the rest of the system, the identity of this piercing tip protein has not been established.

The cell-puncturing devices are a conserved feature of contractile tail-like systems (Leiman and Shneider, 2012). Even though they are not annotated as such in the database, they can be identified with modern bioinformatic tools (e.g., HHpred (Söding et al., 2005)). Crystal structures of T4 gp27 orthologs have been solved for some phages and prophages mostly within the framework of structural genomics (Protein Data Bank [PDB] IDs: 3cdd, 1wru, 3d37, 3gs9) (Kondou et al., 2005). No structural information regarding any of the T4 gp5 orthologs from these or other phages is available.

One of the most interesting orthologs of the T4 cell-puncturing device is found in the T6SS. It is called VgrG (valine/glycine-repeat protein G) (Pukatzki et al., 2007). VgrG contains T4 gp27, gp5, and—sometimes—an effector domain within a single gene and represents one of the T6SS signature secreted proteins. Using its T6SS, a bacterium can translocate VgrG across the membrane directly into the cytoplasm of the eukaryotic host cell (Pukatzki et al., 2007). The crystal structure of the N-terminal part for one of the VgrG proteins containing the complete gp27 module and the N-terminal domain of gp5 is available (PDB ID 2p5z), but the structure of its C-terminal membrane-piercing part is unknown (Leiman et al., 2009).

We addressed the question of the identification and structure of the membrane-piercing proteins by studying two contractile tail bacteriophages whose baseplates are somewhat simpler than that of T4: phages P2 and  $\phi$ 92. Phage P2 is one of the few “model” phages with contractile tails that have been studied in the laboratory for many years (Haggård-Ljungquist et al., 1995; Nilsson and Haggård-Ljungquist, 2007). The protein composition of the P2 tail is well established, although very little is known

**Table 1. HHpred Analysis Shows that P2 gpV and  $\phi$ 92 gp138 Contain an OB-Fold Domain, Also Found in the Central Spike Proteins T4 gp5 and VgrG**

Best matches (PDB ID)	P2 gpV (211 Residues)				$\phi$ 92 gp138 (245 Residues)			
	Aligned Residues (gpV/Target)	E Value	p Value	Sequence Identity	Aligned Residues (gp138/Target)	E Value	p Value	Sequence Identity
VgrG (2p5z)	18-99/386-484	1.6E-21	6.9E-26	17%	21-95/383-462	0.22	9.8E-06	18%
T4 gp5 (1k28)	20-100/12-115	0.00092	4E-08	16%	Not matched			

See also Table S1.

about the structure of the baseplate (Haggård-Ljungquist et al., 1995). Phage  $\phi$ 92 infects a wide range of enterobacteria, including, among others, the polysialic acid encapsulated *Escherichia coli* strains K1 and K92 (Kwiatkowski et al., 1982). The recently sequenced  $\phi$ 92 genome (148.4 kb, EBI Accession No. FR775895) contains 248 open reading frames (ORFs), many of which are related to those of phages rv5 and PVP-SE1 (Santos et al., 2011). Putative tail genes are listed in Table S1, available online (the full annotation of the  $\phi$ 92 genome will be published elsewhere).

Here, we present the crystal structures of P2 gpV and  $\phi$ 92 gp138 and show that these proteins form the membrane-piercing components of the phage with the help of cryoelectron microscopy (cryoEM) of  $\phi$ 92. Both proteins are spike-shaped trimers consisting of an N-terminal oligonucleotide/oligosaccharide-binding (OB-fold) domain, a tapering  $\beta$  helix, and the C-terminal apex domain. The membrane-piercing tips of the two proteins have similar topologies notwithstanding the last swapped  $\beta$  strand. The apex domains of both spikes contain an iron ion, coordinated by six histidine residues in the octahedral fashion, with two histidines donated by each of the three symmetry-related chains. Except for the iron-coordinating histidines, the residues that form the piercing tip of the structure are not conserved, although both phages infect *E. coli*. The structure and biophysical data strongly support the earlier hypothesis that the spike proteins serve as rigid piercing needles and do not undergo refolding upon interaction with the membrane.

## RESULTS

### Identification of the Phage P2 and Phage $\phi$ 92 Cell-Piercing Proteins

An earlier study with immunogold labeling showed that gpV of P2 is located in the central part of the baseplate and exposed to the milieu (Haggård-Ljungquist et al., 1995). In the literature, this protein is called the “P2 tail spike,” which is almost identical to the term “tailspike” used to describe host cell receptor-binding molecules of phages with short noncontractile tails such as P22, Sf6, K1F, and so forth (Israel et al., 1967; Leiman and Molineux, 2008; Steinbacher et al., 1994). The latter have an appearance of stubby spikes in electron microscopy (EM) images tailspikes and are positioned at the tail’s periphery (Israel et al., 1967). As the two types of spikes are clearly different, here we use the term “central tail spike” or “central spike” to refer to P2 gpV and its orthologs.

We surmised that the P2 central tail spike is an ortholog of T4 gp5; that is, both are membrane piercing but not host cell-binding proteins, even though they show only 16% sequence

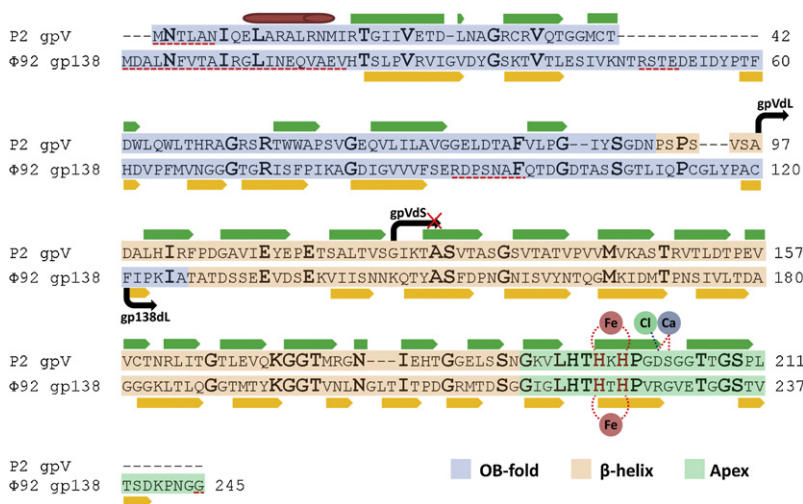
identity over the first 100 residues (the C termini are more diverged) and vary greatly in size (P2 gpV contains 211 residues and T4 gp5 585). Despite these differences, both proteins contain a valine/glycine-rich C-terminal region, which forms a  $\beta$  helix in gp5 and is likely to form multiple consecutive  $\beta$  strands in gpV according to PSIPRED (McGuffin et al., 2000). More important, HHpred analysis showed that, despite the low sequence identity, the N-terminal domain of gpV is very likely to have an OB-fold also found in gp5 and VgrG (Table 1).

In a parallel effort, we obtained the cryoEM reconstruction of the phage  $\phi$ 92 baseplate, which showed a central spike-shaped structure, similar to those found in the phage T4 and phage phiKZ baseplates. Using the same reasoning as in the analysis of the P2 genome, we hypothesized that the  $\phi$ 92 central spike is formed by a protein homologous to T4 gp5 (and to P2 gpV); thus, its N-terminal domain is likely an OB-fold and its C-terminal domain is a  $\beta$  helix. The amino acid sequences of all 248  $\phi$ 92 genes were analyzed for the presence of valine/glycine-rich regions known to form repetitive  $\beta$  strands and  $\beta$  helices (Kajava and Steven, 2006). The HHpred analysis was then carried out on several selected proteins. Gp138 was the only protein with the N-terminal part predicted to have an OB-fold, albeit with a lower degree of confidence than that of P2 gpV (Table 1). The amino acid sequences of P2 gpV and  $\phi$ 92 gp138 have 19% identity and contain a conserved LHTHxHP motif near the C terminus (Figure 1).

### Crystallographic Analysis of P2 gpV and $\phi$ 92 gp138 Structure Determination and Structure-Aided Design of Deletion Mutants

We solved the structure of P2 gpV using the single-wavelength anomalous diffraction (SAD) technique with the help of a Selenomethionine (SeMet) derivative. For both SeMet and native proteins, we were able to grow very large (>1 mm in size) crystals, which showed no imperfections when observed under the microscope. The best crystals diffracted X-rays to about 2 Å resolution (Table 2), but the diffraction images had a blotchy background indicating of internal disorder. The extent of this disorder became apparent in the electron density (Figure S1A). It was not possible to build the atomic model for about one quarter of the crystallized protein (residues 1–6 and 164–211 were disordered). Notably, the disordered regions participate in crystal lattice contacts (Figure S1A). A similar type of disorder was observed in the crystals of the bacteriophage T4 short tail fiber fragments (Thomassen et al., 2003; van Raaij et al., 2001).

Using the information from the partial model, we designed two deletion constructs, gpVdS and gpVdL, which contained the disordered C-terminal part of the protein and corresponded to



**Figure 1. Sequence Comparison of P2 gpV and  $\phi$ 92 gp138**

The sequence alignment, calculated with the help of ClustalW (Thompson et al., 1994), was instrumental in designing the deletion mutants of gpV and gp138, which are labeled with their corresponding names and black arrows. Also shown is the information derived from the crystallographic studies. The secondary structure elements are indicated with arrows ( $\beta$  strands) and a red bar (the N-terminal  $\alpha$  helix). The three domains constituting the structure are labeled with distinct colors. The residues disordered in all crystal forms and mutants are underlined with a red dashed line. The ions buried within the apex domain are shown as circles labeled Fe, Cl, and Ca.

residues 122–211 and 96–211, respectively (Figure 1). The shorter mutant (gpVdS) was insoluble on expression, but the longer mutant (gpVdL) was soluble and formed a very stable trimer in solution. Crystals of gpVdL had a very long unit cell (49 Å × 49 Å × 464 Å; Table 2), but nevertheless diffracted X-rays to better than 0.9 Å resolution. The PILATUS 6M photon-counting detector was used to finely slice the reciprocal space, making it possible to collect a 100% complete data set to 0.94 Å resolution (Table 2). We obtained initial phases from the anomalous scattering signal of sulfur atoms of the methionine residues using the data collected at a wavelength of 1.254 Å corresponding to the Ta  $L_{III}$  edge. These phases were improved by the atom-recycling procedure as implemented in ARP/wARP (Morris et al., 2003) and allowed building the atomic model for the C-terminal domain. The latter was then used to complete the missing part of the full-length structure by superimposing residues 97–140 found in both structures.

Despite extensive attempts and expected structural homology, we were unable to solve the structure of full-length gp138 using any part of gpV as a molecular replacement (MR) model in automatic searches or in the brute force search procedure. SeMet-substituted gp138 failed to crystallize, and heavy atom derivatization attempts were unsuccessful. Thus, we designed a deletion mutant called gp138dL (residues 121–245), which corresponded to gpVdL based on the sequence alignment of gpV with gp138 (Figure 1). Crystals of gp138dL had an even longer cell than those of gpVdL (48 Å × 48 Å × 553 Å; Table 2) but nevertheless diffracted X-rays to 1.3 Å resolution. We solved the structure of gp138dL using the SAD technique with the anomalous scattering signal of strontium ions present in one of the crystallization conditions. The structure of full-length gp138 was solved by MR with gp138dL as a search model.

#### Crystal Structure of gpV and gp138

Phage P2 gpV and phage  $\phi$ 92 gp138 are spike-shaped trimers about 125 Å and 120 Å long, respectively, and 42 Å and 53 Å wide in their widest parts, respectively (Figure 2). The three polypeptide chains constituting the trimers are highly intertwined. About 42% and 25% of all surface-exposed residues in the monomers of gpV and gp138, respectively, are buried within the trimer interface. The bioinformatic analysis outlined earlier

proved to be correct: both proteins have a similar architecture and are composed of an N-terminal OB-fold domain and a C-terminal  $\beta$ -rich part. The latter consists of an extended  $\beta$  helix, which tapers toward the C terminus and ends in a sharp tip structure constituting the apex domain.

#### The N-Terminal OB-Fold Domain

Residues 1–17 of gpV form a short coiled-coil segment with a typical hydrophobic core. The corresponding region of gp138 (residues 1–22) is disordered in the crystal structure but, according to PSIPRED, is likely to form an  $\alpha$  helix.

Residues 18–90 and 23–126 constitute the N-terminal OB-fold domain in gpV and gp138, respectively (Murzin and Chothia, 1992). The OB-fold of gp138 contains an intramolecular disulfide bond between C114 and C120. The distance between the two corresponding  $S_{\gamma}$  atoms is 2.07 Å, which is within the experimental error of the average disulfide bond length of 2.05 Å. The most significant difference of gpV and gp138 OB-folds, compared to other OB-folds, is a 20 Å-long  $\beta$ -hairpin insertion between strands  $\beta$ 3 and  $\beta$ 4 (Figure 3). This long arm forms the first strand of the  $\beta$  helix and is also found in the structure of two other known orthologs, T4 gp5 and T6SS VgrG (Kanamaru et al., 2002; Leiman et al., 2009). The second and two subsequent strands of the  $\beta$ -helical domain are donated by a different polypeptide chain but create a smooth  $\beta$  sheet (Figure 3).

Similar to gp5 and VgrG, the long axis of the gpV OB-fold barrel is roughly parallel to the trimer axis. The OB-fold domain of gp138 is tilted away from the trimer axis, making the diameter of the gp138 OB-fold trimer significantly greater. This difference in conformation is unlikely to be caused by the crystal lattice contacts. The gp138 OB-fold contains a unique 26-residue insertion between strands  $\beta$ 4 and  $\beta$ 5. Residues 92–117 form a large loop that wraps around each of the OB-fold domains, pushing them apart (Figure 3). The first ordered residue of gp138 OB-fold (H22) is 14 Å away from its equivalents in the other two chains, which is far too long to form an interaction to create a coiled coil, partially explaining the reason for the disordered N-terminal  $\alpha$  helix.

#### The $\beta$ Helix

Residues 98–190 in gpV and 138–215 in gp138 form the  $\beta$  helix or  $\beta$  prism domain. The highly intertwined all- $\beta$  domains are called “ $\beta$  helices,” “ $\beta$  prisms,” and “ $\beta$  rolls” in the literature (Kajava and Steven, 2006). Here we call the all- $\beta$  domains of gpV and gp138 “ $\beta$  helices” for simplicity, despite their different

**Table 2. Crystallographic Data Collection and Refinement Statistics**

Phage	P2			$\phi$ 92		
	gpV	gpVdL		gp138	gp138dL	
Protein/Mutant Name	SeMet $f''$ Peak <sup>a</sup>	High- Resolution Native	Low-Energy S-SAD Phasing	Native	High- Resolution Native	Strontium $f''$ Peak
Data collection						
Beamline	SLS PXI	SLS PXI	SLS PXII	SLS PXIII	SLS PXI	SLS PXIII
Wavelength (Å)	0.9795	0.8000	1.2537	1.0000	1.0000	0.7688
Space group	P321	H32	H32	H32	H32	H32
Cell dimensions	a = 68.54 Å	a = 49.07 Å	a = 49.28 Å	a = 65.14 Å	a = 48.08 Å	a = 48.04 Å
	c = 132.18 Å	c = 463.90 Å	c = 464.51 Å	c = 329.38 Å	c = 553.20 Å	c = 552.21 Å
Asymmetric unit content (polypeptide chains)	1	2	2	1	2	2
Resolution range (Å)	66.09–2.03	27.33–0.94	42.49–1.30	55.61–2.64	41.52–1.29	41.20–1.93
Unique reflections	24032	141144	48286	8322	61805	35668
Redundancy	17.4	6.0	16.4	10.5	16.6	10.2
Completeness (%)	99.9 (100.00) <sup>b</sup>	99.9 (99.9)	88.1 (18.4)	99.7 (98.1)	98.6 (96.6)	99.5 (97.1)
$R_{\text{sym}}$ (%)	5.9 (34.3)	5.8 (25.6)	4.9 (27.5)	8.4 (46.9)	8.2 (75.9)	8.6 (43.9)
$I/\sigma(I)$	27.6 (7.5)	16.6 (5.8)	50.6 (4.5)	18.4 (4.5)	17.6 (4.3)	19.6 (3.6)
Refinement						
$R$ (%)	25.97	10.23		19.56	12.29	
$R_{\text{free}}$ (%)	30.50	11.20		25.33	15.98	
B factor (Å <sup>2</sup> )	35.67	6.41		43.81	12.50	
Rmsd of bond lengths (Å)	0.004	0.017		0.004	0.016	
Rmsd of bond angles (°)	0.996	1.991		0.836	1.751	
No. of residues	156	228		208	212	
No. of water molecules	224	477		165	420	
Other ligands: Fe:Ca:Cl:Na:K	0:0:1:0:0	2:2:2:1:0		2:0:0:0:1	2:0:0:7:0	
PDB ID	3QR8	3QR7		3PQI	3PQH	

See also Table S1.

<sup>a</sup>  $f''$  is the imaginary component of the atomic scattering factor.

<sup>b</sup> The number in parentheses describes the statistics for the highest resolution shell.

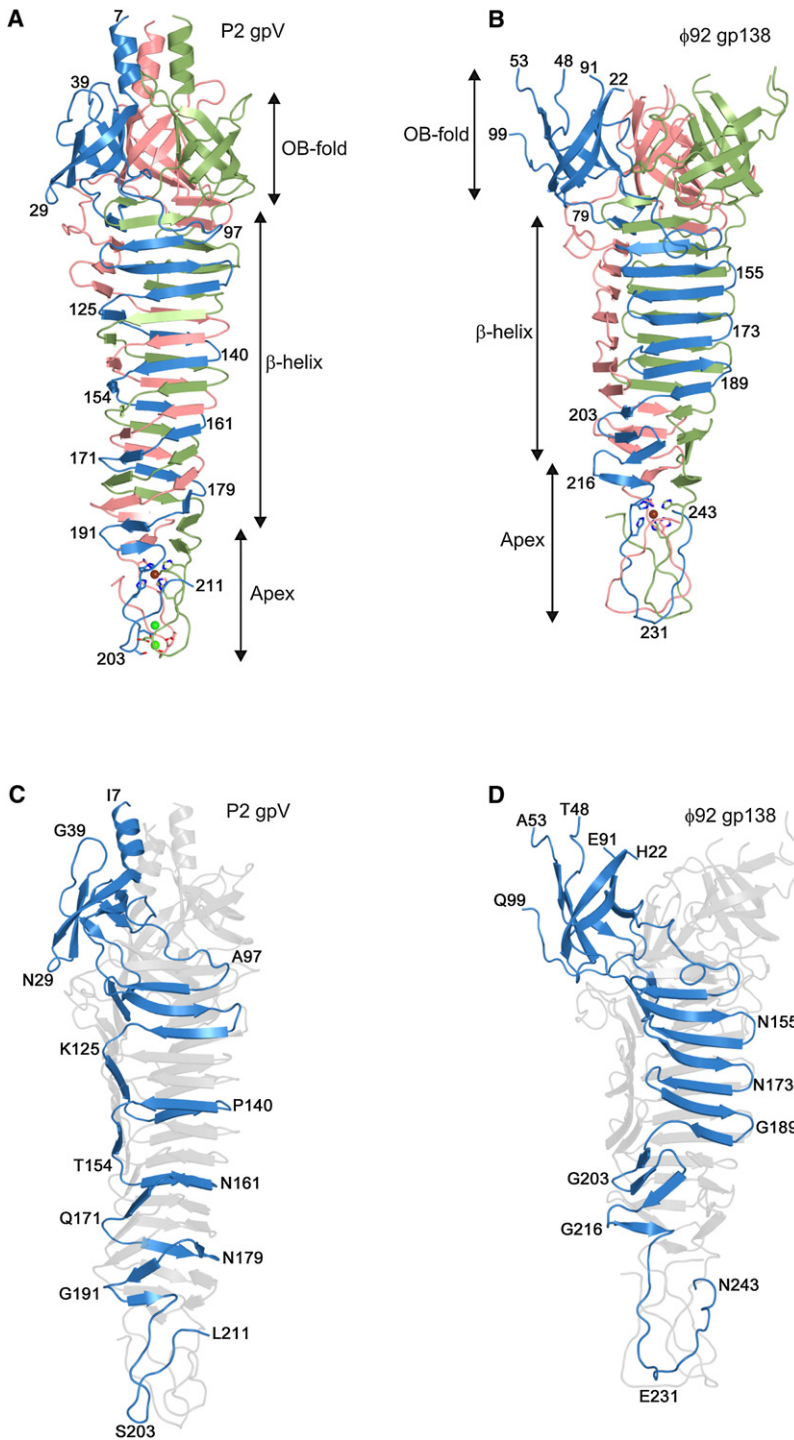
topologies. The three chains of the gpV  $\beta$  helix are fully intertwined with each chain, forming a long corkscrew structure (similar to T4 gp5), whereas the gp138  $\beta$  helix is composed of three antiparallel  $\beta$  sheets, which swap three C-terminal  $\beta$  strands with each other around the threefold axis (Figure 2). Despite the low sequence similarity (Figure 1), the two domains are likely to have a common ancestor from which they both diverged by many domain/chain/ $\beta$  strand swapping events.

The  $\beta$  helix tapers toward the C terminus from about 32 Å in diameter (the distance between the outermost  $C\alpha$  atoms) to 20 Å or 25 Å for gpV and gp138, correspondingly. There are six residues per  $\beta$  strand in the beginning of the  $\beta$  helix and five starting from about its middle part to the end. The C-terminal part of the helix contains several large charged residues (arginines, lysines, and glutamates) that extend into the solution, increasing the actual diameter of the molecular surface. Nevertheless, most of the  $\beta$  helix does not display a prominent electrostatic charge. A significant negatively charged patch is located at the point of junction of the  $\beta$  helix with the OB-fold domain in both gpV and gp138 (Figures S2A and S2B). Most of the large charged residues are found near the corners of the triangular cross-section of the  $\beta$  helix (i.e., in the beginning or at the end of  $\beta$  strands).

Neither the gpV nor gp138  $\beta$  helix contains an obvious binding site for an oligosaccharide or oligopeptide. No heavy atom, cryoprotector, or other small molecule was found to bind to either of the  $\beta$  helices in any of the available crystallization conditions and many soaking experiments.

The interior of the  $\beta$  helix is hydrophobic and formed by valines, leucines, isoleucines, and occasional phenylalanines, methionines, and cysteines (Kajava and Steven, 2006).  $\beta$ -helical proteins are generally characterized by an increased content of valines and glycines compared to the average found for all known proteins, and this skewed distribution is often used to predict a  $\beta$  helix-like fold (Kajava and Steven, 2006). In agreement with this observation, the glycine content of the gpV and gp138  $\beta$  helices is greater than the UniProt database average (10%–14% versus ~7%) (Figure 4). However, their valine content does not fully follow this trend. The gpV  $\beta$  helix contains a high percentage of valines (~13% versus the UniProt average of ~7%), whereas the valine content of the gp138 helix (~5%) is significantly below the database average (Figure 4).

It is surprising that both the gpV and gp138  $\beta$  helices contain an unusually large number of threonines (~14% versus the database average of ~5%). Most of these threonines are located in



**Figure 2. Structure of P2 gpV and φ92 gp138**

Trimers of P2 gpV (A) and φ92 gp138 (B) are shown as ribbon diagrams. Residue numbers are indicated in strategic positions. The three domains constituting the structure are labeled. (C) and (D) show the topology of one of the three constituent polypeptide chains in gpV and gp138, respectively. See also Figure S2.

because of the insufficient number of residues to amount to good statistics. Notably, the gp138 β helix is characterized by a high content of asparagines: ~10% versus the UniProt average of ~4% and ~3% for the gpV β helix (Figure 4). Most of these asparagines are part of or near the sharp turns of the polypeptide chain between the β strands, with their head groups forming interchain hydrogen bonds and thus stabilizing the trimer. In summary, amino acids of three most prevalent types—glycines, threonines, and valines—constitute a total of ~38% and 33% of all residues in gpV and gp138 β helices, respectively.

**The Structure of the Apex Domain**

The β helices end in the apex domain (residues 191–211 and 216–245 in gpV and gp138, respectively) (Figure 5). The monomeric unit of this trimeric structure is composed of three β strands, with the first short strand being perpendicular to the trimer axis and the two subsequent strands forming two antiparallel β strands roughly aligned with the trimer axis. The structure of the apex domain in both proteins is conserved apart for the last β strand being swapped between the symmetry-related chains.

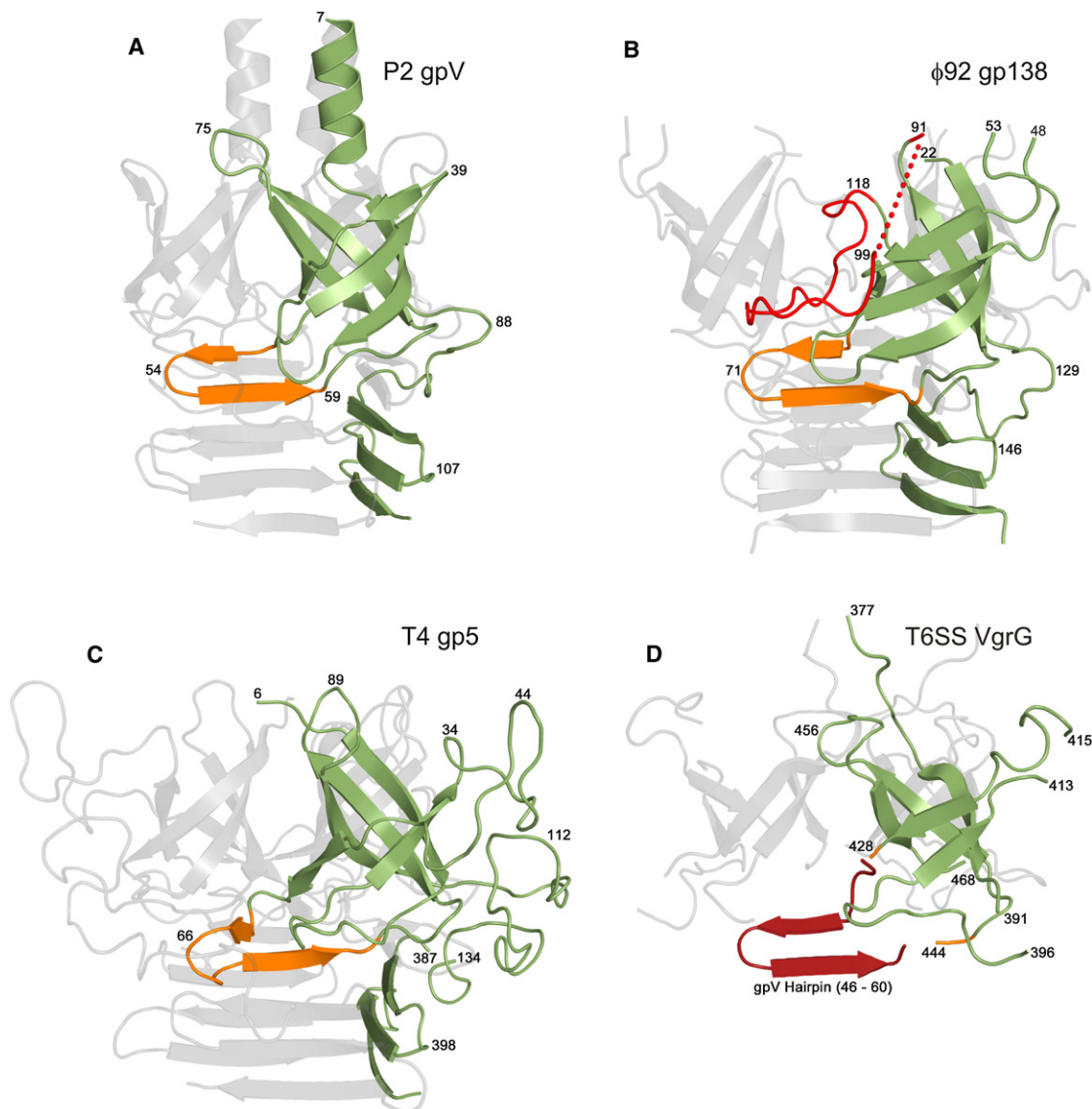
In both structures, the apex domain tapers from about 23 Å to 9 Å in diameter. The gpV apex tip is formed by three β-hairpins, which come together like petals in a flower bud. Residues D203–S203–G204 comprise the sharp turn of the β-hairpins, ensuring the small diameter of the tip. The last β strand of gp138 is threaded under the symmetry-related polypeptide chain, making the tip appear to be less sharp than that of gpV. This impression is incorrect, as the tip of gp138 contains a well-defined apex residue E231, which points along the axis of the trimer and form the actual extremity. The

middle of the β strands with their side chains pointing outward. The elevated threonine and glycine content makes them TG-rich, but not VG-rich, domains, a feature that is also apparent in the amino acid composition of the complete protein sequences of gpV and gp138 (Figure S3).

The content of other amino acids in the two β helices is different, most possibly because of their different topologies (the corkscrewlike gpV versus the β prism-like gp138) and

head groups of the three symmetry-related glutamates (E231) are only 3.4 Å away from each other and 1.9 Å away from the trimer axis (Figure 5).

The external surface of the apex's tip in gpV and gp138 is formed by small residues S203–G204–G205–T206 and T232–G233–G234–S235–T236, respectively. However, large positively charged residues protrude away from the backbone further away from the tip. In gpV, these are three symmetry-related



**Figure 3. The Conserved Organization of the Spike Proteins**

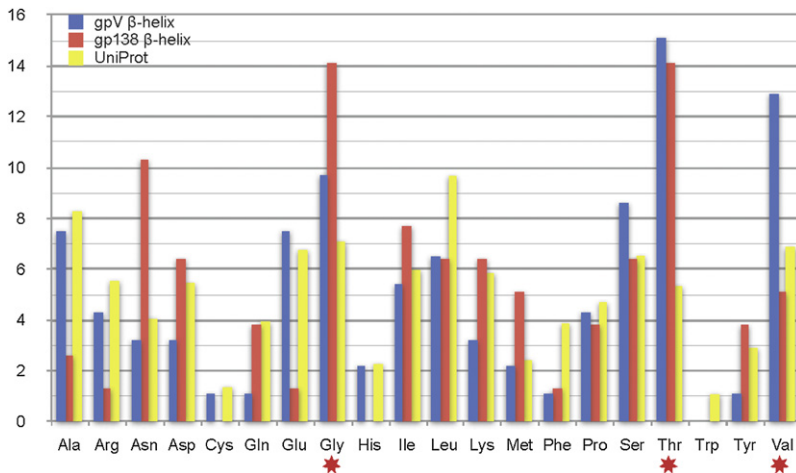
OB-fold domains and the first strands of the  $\beta$  helix of P2 gpV,  $\phi$ 92 gp138, T4 gp5, and T6SS VgrG are shown in (A), (B), (C), and (D), respectively. The conserved  $\beta$ -hairpin, which starts the  $\beta$  helix, is highlighted in orange. This  $\beta$ -hairpin is disordered in the crystal structure of T6SS VgrG. As the gpV  $\beta$ -hairpin has the same number of residues, it was used to model the VgrG  $\beta$ -hairpin, shown in red in (D). The extra loop, which changes the orientation of the OB-fold domain in  $\phi$ 92 gp138, is shown in red in (B).

K198, which are  $\sim 15$  Å from the tip. In gp138, R228 is  $\sim 8$  Å from the  $C\alpha$  atoms of the tip and  $\sim 14$  Å from E231 head groups. D240 and K241 are found further down the gp138 surface. They are  $\sim 19$  Å and  $\sim 24$  Å away from the E231 head groups, respectively (Figure 5).

#### The Apex Domain Contains an Iron Atom

The most remarkable feature of the gpV and gp138 apex domains is the organization of their interior. Each of the three chains constituting the apex domain contains two histidine residues—H197 and H199 in gpV and H223 and H225 in gp138—that point to the structure interior. Three pairs of these histidines create a perfect octahedral coordination shell for the iron ion positioned on the axis of the trimer (Figure 5). The average

distance between the  $N\epsilon 2$  atom of histidines and the iron ion is  $2.20 \pm 0.01$  Å in gpV and  $2.23 \pm 0.01$  Å in gp138. This is longer than the average found in other iron-containing proteins (2.08 Å). The presence of the iron atom was unexpected because the proteins, their deletion mutants, or their crystals did not show any color or any unusual spectroscopic properties in the UV-Vis spectrum associated with all known iron-containing proteins. The presence of iron was confirmed by X-ray fluorescent emission spectra of gpV, gpVdL, gp138, and gp138dL and later using Raman spectroscopy. Preliminary combined X-ray diffraction/Raman measurements showed no obvious changes in the structure of the iron coordination site after prolonged exposure to intense X-rays indicating tight binding.



**Figure 4. The Histogram of Amino Acid Composition of P2 gpV and  $\phi$ 92 gp138  $\beta$ -Helices**

UniProt is the histogram of amino acid usage frequencies averaged over the entire database of the Universal Protein Resource (<http://www.uniprot.org/>) in January 2011. Red stars identify the most prevalent amino acids in gpV and gp138  $\beta$ -helices. See also Figure S3.

oxygen atoms (3.30 Å away), and the calcium ion mentioned earlier (4.02 Å away). The gp138 apex tip is hydrophobic and is free from these or any ions because of its more interwoven topology.

### Structure of the Phage $\phi$ 92 Baseplate

In an effort parallel to those of the crystallographic studies of the central spike proteins,

we analyzed the structure of the  $\phi$ 92 particle using cryo-electron microscopy and image reconstruction (cryoEM). The structure of the distal part of the tail, containing the baseplate and part of the sheath was reconstructed to a resolution of  $\sim 26$  Å (Figure 6). The structures of the capsid and the rest of the tail will be reported elsewhere. The hand of the reconstruction was determined with the help of atomic force microscopy (AFM) (Figure S4) and by fitting the crystal structure of a sheath protein homolog (PDB ID 3HXL) into the cryoEM map. Only one hand of the sheath protein subunit was assumed to be roughly similar to that found in the T4 tail sheath.

The  $\phi$ 92 baseplate is significantly different from that of the phage T4 and phage phiKZ studied earlier (Fokine et al., 2007; Kostyuchenko et al., 2003). Several long fibers point away from the  $\phi$ 92 baseplate like different tools in an open Swiss army knife (Figure 6). The  $\phi$ 92 genome contains a large cluster of genes, which are responsible for forming these fibers (Table S1), but not every phage particle might contain a full complement of fibers/tailspikes leading to the observed difference in their occupancies. It is also possible that one type of fibers is less mobile and, hence, much better defined in the cryoEM map than the others. Still another alternative explanation is that one large fibrous structure on the baseplate periphery adapts at least three distinct conformations, with the conformations “down” and “up” being the most and the least populous, respectively. Because of the limited size of the data set, we were unable to classify particles according to the fiber occupancy and/or conformation. We are addressing this question using cryotomography.

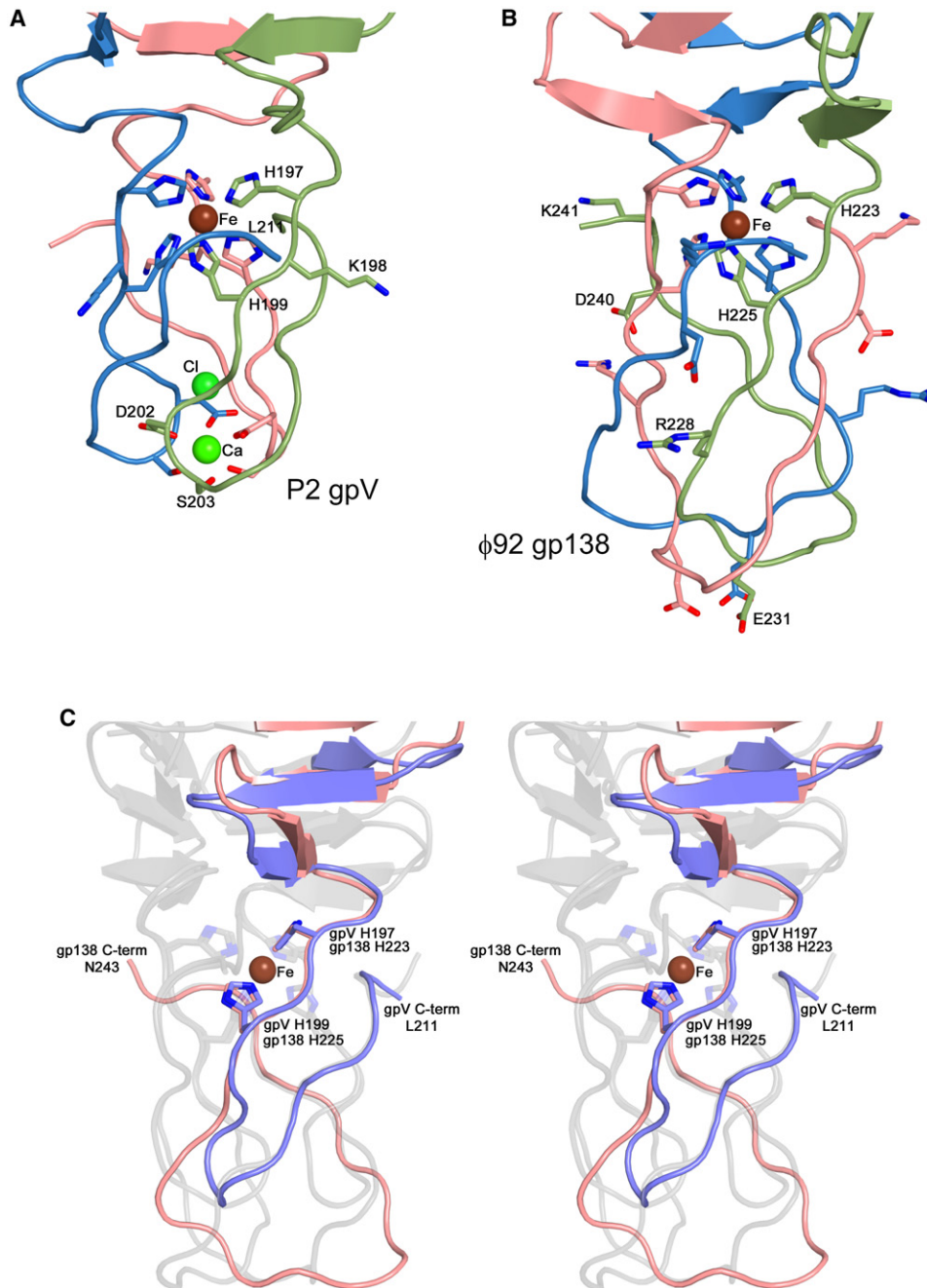
The central part of the  $\phi$ 92 baseplate is well ordered. It is possible to uniquely assign the central spike density to the gp138 trimer. The axial orientation of gp138 can be established with twofold ambiguity because the density is sixfold averaged, whereas the structure is trimeric. In contrast, the translation of gp138 along the spike density is determined uniquely. The crystal structure of gp138 fully occupies the density of the central spike, with the iron-containing apex domain constituting the tip of the spike (Figure 6). As the structure of gpV is homologous to that of gp138, gpV is likely to occupy the same position in P2 baseplate with the apex domain forming the tip of the central tail spike.

A similar octahedral hexahistidine coordination of iron ions was found recently in the structure of the C-terminal domain (called gp37C) of the phage T4 long tail fiber protein gp37 (Bartual et al., 2010). The needle-shaped gp37C trimer contains seven iron atoms dispersed through the length of the molecule. As in gpV and gp138, the gp37C iron ions are positioned on the threefold axis of the protein and are bound to the HxH motifs. Again, similar to gpV and gp138, the UV-Vis spectral properties of gp37C protein solution or crystals gave no indication of their high iron content. The average distance between the iron ions and the N $\epsilon$ 2 atom of histidines in gp37C is  $2.31 \pm 0.06$  Å, which is even greater than that of gpV or gp138. It is possible that the long distances, which characterize the iron-binding sites in gpV, gp138, and gp37C, contribute to the absence of color in these proteins.

The position and binding of the iron ions in gpV and gp138 have one significant difference compared to those of gp37C. None of the iron ions in gp37C is found at the extremities of the molecule. In fact, all iron atoms are part of gp37C middle domain. The distal and proximal ends of gp37C are formed by globular domains, which are free from any metal ions.

Despite obtaining the electron density map of exceptional quality, in which many of hydrogen atoms are visible (0.94 Å resolution and  $R_{\text{free}}$  of 11%; Figure S1F), we were unable to reliably estimate the oxidation state of the iron ion using refinement of crystallographic data. Both of the two common oxidation states of iron (+2 or +3) gave similar crystallographic residuals and approximately similar residual difference map peaks, albeit the residual peaks in the Fe $^{3+}$  map were slightly smaller. Electronic paramagnetic resonance experiments turned out to be inconclusive. Therefore, we made a tentative assignment of the iron oxidation state as +3.

Beside the iron ion, the interior of the gpV apex domain contains two other ions—a calcium and a chlorine—both of which are positioned on the trimer axis, one below the other. The calcium ion is coordinated by the head groups of three D202 and S203 and is located just below the molecular surface formed by the O $\gamma$  atoms of S203. The geometry of the calcium-binding site is not unusual (Figure 5), but the chlorine-binding site is quite remarkable. It is created by the backbone nitrogen atoms of D202 (which are 3.30 Å away), the D202 head group



**Figure 5. Structure of the Apex Domain**

(A) and (B) show a close-up view of the ions buried within the apex domains of P2 gpV and  $\phi$ 92 gp138, respectively. (C) is a stereo diagram of the two superimposed apex domains demonstrating the conserved topology and swapping of the last  $\beta$  strand. See also Figure S4.

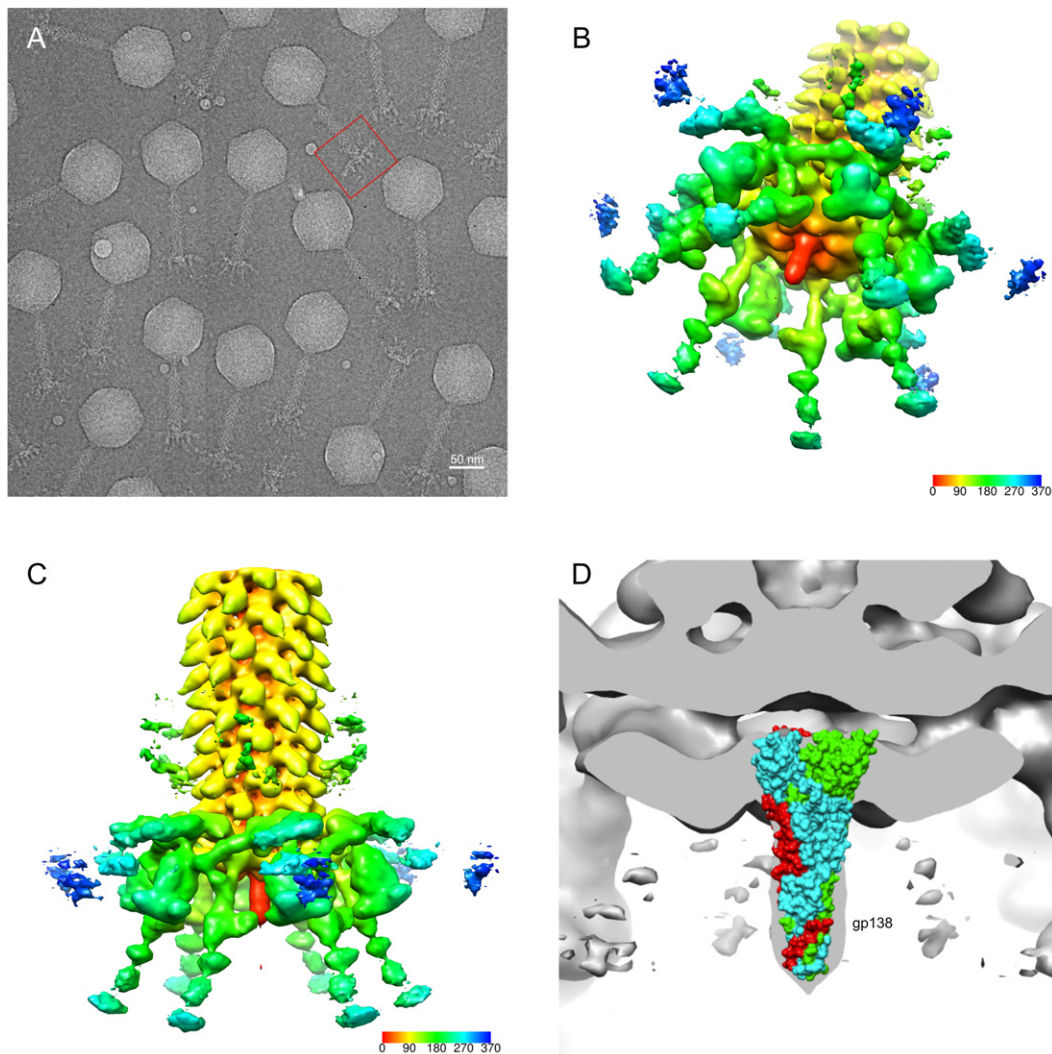
## DISCUSSION

### P2 gpV and $\phi$ 92 gp138 Form the Membrane-Piercing Spikes

Information available from earlier studies and modern bioinformatic tools allowed us to surmise that P2 gpV and  $\phi$ 92 gp138 are orthologs of the phage T4 central spike protein gp5. The combined crystallographic and cryoEM analysis presented

here confirms this hypothesis and provides key additional information about the structure and organization of the central spike proteins of contractile tail-like systems (phage tails, R-type pyocins, T6SS, etc.). The membrane-piercing tip of the central spike proteins contains a buried, centrally positioned, hexahistidine-coordinated iron ion, which is bound to the conserved HxH motif. The iron ion is likely to have a dual function: it participates in the folding of gpV into a trimer by helping to keep the three





### Figure 6. CryoEM Structure of $\phi 92$ Baseplate

(A) CryoEM micrograph of  $\phi 92$  phage particles. The red box shows the reconstructed part of the particle.

(B and C) Tilted and side views of the cryoEM density map contoured at  $2\sigma$  above the mean. The map is colored in a rainbow pattern starting from the center (red) to the periphery (blue). A key relating the color to the radius in Å is given in the lower right corners of (B) and (C).

(D) Fitting of the gp138 crystal structure into the cryoEM map. The molecular surface of gp138 is colored according to the trace of each of the three polypeptide chains constituting the trimer. A slice ( $\sim 60$  Å thick) of the cryoEM map is shown. See also Table S1 and Figure S5.

polypeptide chains, which are about to be folded, in register; and it provides an increased stability to the tip structure once the protein has folded.

Both phages P2 and  $\phi 92$  require calcium for infection. A calcium ion is found at the tip of P2 gpVdL, crystallized in several conditions containing no calcium. Furthermore, the protein was incubated with EDTA at a concentration of 1.5 mM overnight during purification (see [Experimental Procedures](#)). Therefore, this calcium ion is unlikely to have a regulatory function in phage infectivity but is important for the structural stability of gpV. This reasoning is further confirmed by the structure of the gp138 apex domain, which is free from calcium or any other divalent metal ions when crystallized in the presence of potassium, calcium, or strontium at concentrations as high as 0.4 M. The conformation of the protein backbone in all of these crystallization condi-

tions is unchanged. Therefore, calcium is most probably required for the initial adsorption of either phage to the host cell surface, which is mediated by other baseplate proteins. It is possible that P2 and  $\phi 92$  baseplates have a similar calcium-dependent activation mechanism proposed for noncontractile tail phages infecting lactic acid bacteria ([Sciara et al., 2010](#)).

Alignment of the amino acid sequences of P2 gpV and  $\phi 92$  gp138 to the protein database shows that the double-histidine H×H motif responsible for iron binding at the C-terminal end of the protein is very well conserved ([Figure S5](#)), suggesting that the apex domains are likely to contain iron and to have the same fold with a possibility for occasional chain swapping, similar to gpV and gp138. In some spike proteins, for example in R-type pyocins, the double-histidine motif forms the C terminus of the structure. In some others, the double-histidine motif is followed

by a sequence containing more than 20 residues. A stretch of small residues (G, S, T, D, and P) is always present in the middle of this sequence, indicating the position where the polypeptide chain turns back, and forms the pointy tip of the structure. In general, the sequence of the apex tip residues for phages infecting the same bacterial species is not conserved, suggesting that no specific binding of the tip to the host cell membrane occurs.

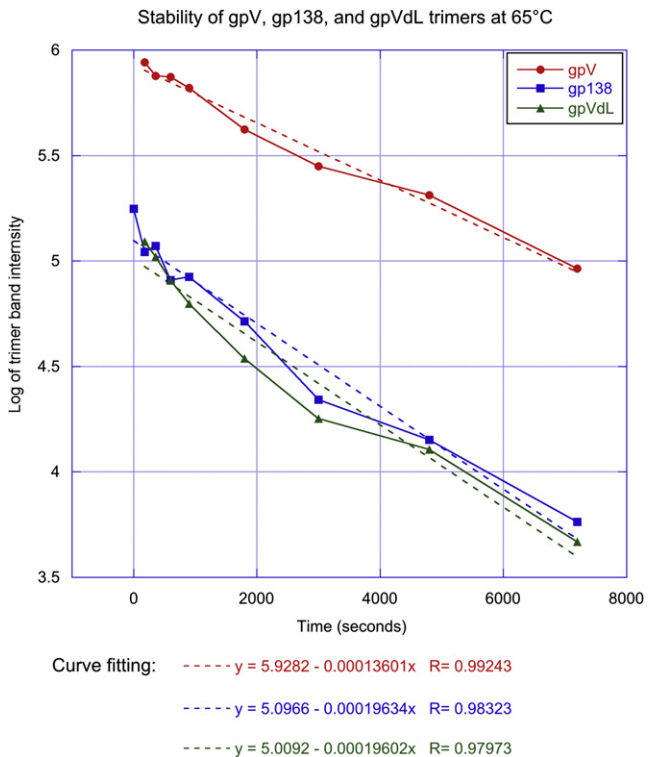
Earlier studies showed that a G64→E mutation in gpV leads to a temperature-sensitive phage phenotype (Haggård-Ljungquist et al., 1995). This glycine is located in the beginning of strand  $\beta$ 5 of the OB-fold domain on the surface of the protein. The glycine residue has backbone angles corresponding to those found in an  $\alpha$  helix and can be easily replaced with a glutamate with the head group pointing into the solution (Figure S2C). Multiple sequence alignment shows that a glycine is a predominant residue in this position, but an aspartate is also found in some homologs (this glycine corresponds to position 93 in Figure S5A). Thus, the G64→E mutation is likely to negatively influence the interaction of gpV with other baseplate proteins, in particular with gpJ (the P2 ortholog of T4 gp6) to a greater degree than the folding of gpV.

#### Interaction of the Spike Protein with the Host Membrane

Host cell attachment is followed by contraction of the tail sheath, which causes the tail tube tip to cross the host cell membrane (Liu et al., 2011; Simon and Anderson, 1967a, 1967b). The central spike protein forms the membrane-attacking tip of the tube and plays a critical role in this process (Kanamaru et al., 2002; Kostyuchenko et al., 2003). A hypothesis that the central spike complex of phage T4 functions to pierce the outer membrane was proposed (Kanamaru et al., 2002), albeit later studies showed that the T4 spike protein gp5 is capped with an unknown protein significantly undermining the membrane-piercing concept (Kostyuchenko et al., 2003).

The structure of the apex domain of P2 gpV and  $\phi$ 92 gp138 shows that the tip of the spike is a very sharp and stable structure formed by three intertwined and iron-coordinated polypeptide chains. GpV, gp138, and their deletion mutants constituting the membrane-attacking regions are resistant to dissociation into monomers in the presence of denaturing agents such as urea and SDS. They also display high stability to heat denaturation. The half-life of the full-length gpV and gp138 trimers at 65°C is  $\sim$ 37 and  $\sim$ 26 min, respectively (Figure 7, Figure S6). Furthermore, gpV, gpVdL, gp138, and gp138dL do not aggregate at concentrations exceeding 20 mg/ml in a neutral pH buffer containing little or no salt. In summary, gpV, gp138, and their C-terminal fragments are exceptionally stable and soluble proteins, which show no lipid-binding propensity.

These properties of the spike proteins very strongly support the piercing hypothesis that the central spike protein of a contractile tail is translocated through the membrane without major unfolding or any specific interaction with it but instead is used as a rigid needle to create an opening in the membrane. The fine details of the apex domain interaction with the membrane are not clear, but their progressively greater diameter starting from the tip and going to the N-terminal OB-fold domain (Figure 6D, Figure S2) suggests that the initial opening can be enlarged by the spike as it is translocated through the membrane by the incoming tail tube.



**Figure 7. Thermal Stability of gpV, gp138, and gpVdL**

See also Figure S6.

The spike protein is located in the center of the baseplate and is inserted into a donut-shaped baseplate hub protein (a T4 gp27 ortholog), which terminates the tube (Kostyuchenko et al., 2003). The spike protein thus serves as the distal plug of tail tube channel and might regulate the process of the phage DNA and protein ejection into the host cell. It is interesting that the density corresponding to the region of interaction between the spike and the hub protein is weak in the cryoEM reconstructions of both the T4 and  $\phi$ 92 tails (Figure 6D) (Kostyuchenko et al., 2003). It is possible that the spike protein interacts with the hub protein (and the tube) weakly and is poised for an easy dissociation once the membrane is breached. In T4, the dissociation of the spike and hub proteins (gp5 and gp27) can be triggered by lowering the pH of the solution, which mimics the conditions of the periplasm.

The energy for the translocation of the spike and the tube through the membrane and the movement itself is generated by the sheath, whose contraction results a lower energy state of the entire system. Participation of other factors that can supply additional energy for the translocation (e.g., DNA, which is packaged under high pressure inside the phage capsid) is unlikely. The cell-puncturing complex, identical to that of phage P2, is found in many contractile injection systems (e.g., in R-type phages) that constitute a “capsidless” and DNA-free phage tail-like structure (Nakayama et al., 2000).

#### The Conserved Organization of Spike Proteins

The four available crystal structures of the spike proteins (P2 gpV,  $\phi$ 92 gp138, T4 gp5, and T6SS VgrG) allow us to analyze the

**Table 3. Putative Spike Proteins from Several Diverse Contractile Tail Systems**

Contractile System Name	Type of Susceptible Organism	Protein Name	GenBank Reference	N-terminal Domain	Putative C-terminal Domain Structure	C-terminal HxH Motif Present?
<i>Pseudomonas</i> phage SN	Gram-negative	PPSN_gp43	YP_002418849	OB-fold	$\beta$ helix	Yes
Iodobacteriophage $\phi$ PLPE	Gram-negative	phiPLPE_59	ACG60381	OB-fold	$\beta$ helix	Yes
<i>Cronobacter</i> phage ESSI-2	Gram-negative	ADX32369.1	ADX32369	T4 gp27 plus OB-fold	$\beta$ helix	Yes
<i>Haemophilus</i> phage HP1	Gram-negative	orf35	NP_043506	T4 gp27 plus OB-fold	$\beta$ helix	Yes
<i>Listeria</i> phage A511	Gram-positive	gp99	AAV52880	OB-fold	Coiled coil	No
<i>Staphylococcus</i> phage A5W	Gram-positive	gp ORF085	ACB89078	OB-fold	$\alpha/\beta$ structure	No
<i>Brochothrix</i> phage A9	Gram-positive	BrPhBA9_gp101	YP_004301434	OB-fold	$\beta$ helix	No
<i>Staphylococcus</i> phage Twort	Gram-positive	ORF010	YP_238569	OB-fold	$\alpha/\beta$ structure	No
<i>Bacillus</i> phage SPO1	Gram-positive	gp12.2	YP_002300362	OB-fold	$\beta$ helix	No
<i>Geobacillus</i> bacteriophage D6E	Gram-positive thermophile	D6E-ORF13A <sup>a</sup>		OB-fold	$\beta$ helix	Yes/no <sup>b</sup>
<i>Natrialba</i> phage PhiCh1	Halobacterium (Archeon)	orf28	NP_665946	OB-fold	$\beta$ helix	Yes
<i>Serratia entomophila</i> antifeeding prophage (afp)	Grass grub <i>Costelytra zealandica</i>	afp8 <sup>c</sup>	YP_026144	T4 gp27 plus OB-fold	$\beta$ helix	No
<i>Photorhabdus luminescens</i> strain W14 necrotizing factor	<i>Heterorhabditidae</i> nematode	orf19 <sup>c</sup>	AAN64223	T4 gp27 plus OB-fold	$\beta$ helix	No

All gpV orthologs were identified with the help of HHpred. In all cases, the T6SS spike protein VgrG (PDB ID 2p5z) was the best match to the search query.

<sup>a</sup> D6E-ORF13A is not annotated as a gene product in the current edition of the phage DE3 genome (GenBank ID: GU568037). The genome location and amino acid sequence of D6E-ORF13A is given in the [Supplemental Information](#).

<sup>b</sup> Instead of HxH, D6E-ORF13A contains the HRxH motif, which might be able to bind iron or some other metal ion.

<sup>c</sup> These ORFs are correctly annotated in the database as an ortholog of P2 gpD plus P2 gpV. P2 gpD is a homolog of T4 gp27.

conservation of the spike structure in great detail. Besides the OB-fold domain, the following features are common to all structures: (1) the hairpin insertion between OB-fold strands  $\beta$ 3 and  $\beta$ 4, which serves as the start of the  $\beta$  helix; (2) the spatial relationship of the OB-fold and its hairpin insertion to the  $\beta$  helix; and (3) the topology of the first four  $\beta$  strands of the  $\beta$  helix (Figure 3). The OB-fold domains of gpV, gp138, T4 gp5, and VgrG, however, display very low sequence similarity, making the conservation of these structural features particularly striking (Table S2).

The OB-fold domain with its hairpin insertion represents a very distinct structural marker, which can be used to find orthologs of the spike protein in other contractile tail-like systems with the help of HHpred and gene clustering analysis. gpV orthologs are present in phages infecting Gram-positive bacteria and in contractile ejection systems acting on eukaryotes (Table 3). In some systems, the spike protein is part of a larger gene that also encodes the baseplate hub protein (T4 gp27 ortholog) immediately in front of the OB-fold domain, such as, for example, in T6SS VgrG proteins. The requirement for the conserved iron-containing structure of the membrane-piercing apex domain appears to be common to bacteriophages infecting Gram-negative bacteria (Table 3). It is possible that this requirement extends to all contractile ejection systems acting on Gram-negatives.

The exact composition of the membrane-binding tip of the spike complex in Gram-positive contractile tail phages and eukaryotic contractile systems is unknown, but their spike proteins appear to have significantly diverged from those of

Gram-negative phages (Table 3). In many Gram-positive contractile tail phages the OB-fold is followed by a large coiled coil domain, which is sometimes capped by another protein or domain lacking the double-histidine motif (e.g., gp99 from phage A511; Table 3). In eukaryotic contractile systems, the  $\beta$  helix is present at the C terminus of VgrG proteins, but its fold, topology, and whether it is capped by another protein are unknown. In many VgrG proteins, the  $\beta$  helix is extended by a functional domain, which confers pathogenicity to the entire T6SS machine (Pukatzki et al., 2007). The structure of the gpV and gp138 apex domains shows that three polypeptide chains can form a very sharp tip and turn back, taking the C termini away from the tip and making them open for fusion with functional domains.

The T4 gp5 spike protein studied earlier contains a lysozyme domain, which is located between the OB-fold and the  $\beta$  helix in the protein sequence and interacts with the N-terminal part of the  $\beta$  helix in the gp5 trimer (Kanamaru et al., 2002). It was proposed to be required to digest the intermembrane peptidoglycan layer during infection, although more recent studies showed that T4 particle containing a mutated lysozyme domain with a 10-fold reduced activity retains the wild-type level of infectivity (Kanamaru et al., 2005). Thousands of phage genomes are deposited into the GenBank and only few—mostly very close relatives of T4—contain a lysozyme domain within their spike protein. It is interesting that a relatively close relative of T4, the broad range *Vibrio* phage KVP40, contains no lysozyme domain in its spike protein (gene 335, GenBank ID: 34419568) or in any

other baseplate protein (Rossmann et al., 2004). KVP40 shows a very high similarity to T4 over most of its virion genes and therefore must have the same infection mechanism, which somehow does not involve a lysozyme domain.

The other feature of the T4 spike is the mysterious additional small protein that caps the spike's  $\beta$  helix and forms the actual membrane-piercing tip (Kostyuchenko et al., 2003). Analysis of the entire T4 genome shows that there is only one gene with the double-histidine iron-binding motif (H $\times$ H) near its 3' end: gene 26. Gp26 has been proposed to participate in T4 baseplate assembly, but it is not found in the mature phage particle. Gene 26 contains an internal in-frame translation initiation site corresponding to a protein with a molecular weight (MW) of  $\sim$ 10 kDa (GenBank ID: 20067601). This is in agreement with the MW assigned to the tip protein in the earlier cryoEM study of the baseplate (Kostyuchenko et al., 2003). The small MW explains why gp26 was not identified as a virus particle component in previous studies. Gp26 and its double-histidine motif is well conserved in T4 relatives with very few orthologs containing a HxxH motif and even fewer—only one histidine. Whether these distant gp26 orthologs lacking the HxH motif are part of the spike complex is unclear.

### The Alternative Crystal Structure of the gpV C-Terminal Domain

When this article was formatted for journal submission, a paper describing the crystal structure of the C-terminal fragment of P2 gpV containing residues 87–211 came out (Yamashita et al., 2011) (PDB ID: 3AQJ). The 3AQJ structure and the four crystal structures reported here were deposited into the PDB at approximately the same time. Having independently identified and crystallized essentially the same fragment of gpV (residues 87–211 versus 96–211), Yamashita et al. reported the same chemical composition (the hexahistidine coordinated iron ion and the C-terminally located calcium and chlorine ions) and the same intertwined  $\beta$  helix fold. However, lacking the structure of the N-terminal domain or any other structural data, Yamashita et al. assumed that six gpV trimers are located at the periphery of the baseplate and participate in host cell binding (Yamashita et al., 2011). This contradicts the earlier study by immuno-EM, which clearly shows that gpV is found at the center of the baseplate (Haggård-Ljungquist et al., 1995). Our results are in line with the immuno-EM study. Furthermore, we show that there is only one copy of the spike protein per particle and that it comprises the membrane-piercing spike in its entirety. As indicated earlier, we were unable to detect any membrane-binding propensity of gpV as reported by the same laboratory in an earlier paper (Kageyama et al., 2009). The spike proteins and their membrane-piercing apex domains from phages infecting the same bacterial species do not show any detectable sequence conservation, suggesting that their weak membrane-binding properties might be due to nonspecific binding by several surface-exposed positively charged residues, some of which are mentioned in the Results section.

### EXPERIMENTAL PROCEDURES

#### Cloning, Expression, and Purification

PCR-amplified fragments corresponding to gpV, gpVdL, gpVdS, gp138, and gp138dL were cloned into the pEEVa2 expression vector using the primers

given in the Supplemental Information. pEEVa2 was derived from pET23a by introducing a TEV protease cleavage site in front of the BamHI restriction site. The resulting proteins contained the MGSSHHHHHSSGQNLVYFQGS tag at the N terminus, which can be cleaved off with TEV. The native N-formyl-methionine was replaced with the last S from the aforementioned sequence.

Protein expression was performed in the 2 $\times$ TY medium supplemented with ampicillin (100 mg/ml). The plasmid-carrying cells were grown to the optical density OD<sub>600</sub> of 0.6 and induced with IPTG to a final concentration of 0.75 mM. The cells were then incubated overnight at 18°C with rigorous shaking. The cells were pelleted at 8000g, resuspended in a lysis buffer (50 mM TrisCl, pH 8.0, 200 mM NaCl), and disrupted by sonication at 4°C–10°C. The insoluble fraction was removed by high-speed centrifugation for 20 min at 27000g. The clarified cell lysate was loaded on to a 5-ml GE HisTrap FF Ni-charged column, and the protein was eluted with imidazole-containing buffer (50 mM TrisCl, pH 8.0; 200 mM NaCl; 250 mM Imidazole) using either a continuous or step gradient. The eluted protein was dialyzed overnight into 50 mM TrisCl buffer, pH 8.0. The His-tag was removed by TEV digestion, which was carried out at 20°C overnight in the presence of 3 mM DTT and 1.5 mM EDTA. The TEV:target protein ratio was 1:40. The digested protein was dialyzed into 10 mM TrisCl, pH 8.0, overnight and then further purified by ion-exchange chromatography (MonoQ 10/100 GL column). The fractions containing the target protein were pulled together and passed through a 1-ml GE HisTrap Ni column to remove the undigested protein. The final purification step was size exclusion chromatography with the Superdex 200 HiLoad 16/60 column in a buffer containing 20mM TrisCl, pH 8.0, and 150 mM NaCl.

The SeMet derivatives of gpV and gp138 were obtained with the help of the Met-auxotrophic *E. coli* 834/DE3 strain grown in the presence of SeMet. The SeMet mutants were purified by following the procedure established for the native proteins.

#### Crystallization and Structure Determination

The purified proteins were concentrated to levels in the range of 20–50 mg/ml in a buffer containing 20 mM TrisCl, pH 8.0, and 50 mM NaCl and were subjected to 960-condition crystallization screening using crystallization kits from Jena Biosciences. This search resulted in several hits for each protein, which were further refined to produce large diffraction quality crystals (Table S3).

The X-ray fluorescent emission spectra of gpV, gpVdL, gp138, and gp138dL crystals showed the presence of iron, although crystals had no spectroscopic features often associated with iron-containing proteins, such as color or additional absorption bands in the near-UV to visible optical range. Except for one crystal form of gp138dL, the protein threefold axis always coincided with a crystallographic threefold, placing the iron atom on a crystallographic threefold. This negated the anomalous scattering contribution of the iron atoms to the diffraction pattern, and phasing with iron atoms resulted in poor maps.

The diffraction data were integrated with XDS (Kabsch, 2010), MOSFLM (Leslie, 2006), or HKL2000 (Otwinski and Minor, 2006), and the data set displaying the best statistics was chosen for subsequent analysis. The heavy atom sites were found with the help of SHELX (Sheldrick, 2008). In cases when the SHELX solvent flattened map was not good enough for model building, the SHELX sites were used in PHENIX (Adams et al., 2011) for a new round of phasing and density modification. The CCP4 program suite (Winn et al., 2011), SHELX, and PHENIX were used in data processing and refinement. The atomic models were built with COOT (Emsley and Cowtan, 2004).

In general, the structure solution procedures were routine except for two cases. First, despite being emerald green in color, a TaBr-soaked crystal of gpVdL showed a very weak anomalous diffraction signal at the Ta L<sub>III</sub> edge (1.2537 Å), suggesting that bound Ta atoms did not have crystalline order. Nevertheless, the structure was solved with this data set. We located the “heavy” atoms in this data set and calculated the electron density map with the help of SHELX. Of the 10 found heavy atoms, only 4 were not on the crystallographic threefold axis. They were later identified to be the sulfur atoms of the four methionines contained in the asymmetric unit. As most of the phasing atoms were located on the crystallographic threefold axis, we attempted to use noncrystallographic symmetry averaging by processing the data and calculating the electron density in the C2 space group in addition to the true space group H32. Both the C2 and H32 electron density maps were equally

fragmented, noisy, and uninterpretable on visual inspection, but the C2 map could be improved to become interpretable with the help of the atom recycling procedure as implemented in ARP/wARP (Morris et al., 2003).

The second nonroutine point of crystallographic analysis was the MR solution of the full-length gp138 data set using the gp138dL mutant structure as a search model. Several MR programs successfully placed the gp138dL structure into the gp138 unit cell, but the resultant phases proved to be very difficult to improve to produce an interpretable density for the N-terminal domain. In the end, the only procedure that resulted in an interpretable map was MR with PHASER (McCoy et al., 2007) followed by density modification with PARROT (Cowtan, 2010).

### Electron Microscopy

The *E. coli* strain Bos12 (O16:K92:H-) was used for propagation of phage  $\phi$ 92, which was purified using a procedure similar to the one described in (Gerardy-Schahn et al., 1995). *E. coli* Bos12 (ATCC 35860) and phage  $\phi$ 92 ("phi 92"; ATCC 35860-B1) were obtained from ATCC. High-purity samples used in EM were obtained by two successive CsCl gradient ultracentrifugation runs and subsequent dialysis into the SM buffer (50 mM TrisCl, pH 7.5; 100 mM NaCl; 8 mM MgSO<sub>4</sub>).

Low-dose cryoEM was performed as described in (Baker and Henderson, 2006). The images were recorded on Kodak SO-163 film using a Philips CM300 FEG microscope at a magnification of  $\times 33000$ , with a radiation dose of  $\sim 20 \text{ e}^-/\text{\AA}^2$  and a defocus of  $-2.0$  to  $-3.5 \text{ }\mu\text{m}$ . The images were scanned using a Zeiss SCAI scanner, with a  $7 \text{ }\mu\text{m}$  step size, and were binned to obtain a pixel size of  $2.108 \text{ \AA}$ .

A total of 985 images of the distal part of the tail containing the baseplate were picked from 38 micrographs. The contrast transfer function was determined and then corrected for with EMAN (Ludtke et al., 1999). The images were binned further to give a pixel size of  $4.216 \text{ \AA}$  and a map of  $200 \times 200 \times 200$  voxels. The T4 baseplate-tail tube complex was used as the initial model. The SPIDER (Frank et al., 1996) and EMAN software suites were used in image processing. We found the resolution of the reconstruction to be  $26 \text{ \AA}$  using the Fourier shell correlation coefficient method, assuming the limit of resolution to be when the correlation coefficient drops to below 0.5.

Fitting of the gp138 crystal into the  $\phi$ 92 cryoEM map was performed with UCSF Chimera (Pettersen et al., 2004). The cryoEM map of  $\phi$ 92 was deposited into the EM databank with the accession number 10191.

### Thermal Stability Assay of gpVdL, gpV, and gp138

The thermal stability assay was based on the procedures carried out by (Chen and King, 1991). Briefly,  $100 \text{ }\mu\text{l}$  of  $5 \text{ mg/ml}$  samples of each gpVdL, gpV, and gp138 were prepared in a low-salt neutral pH buffer (50 mM TrisCl, pH 8.0; 20 mM NaCl). Thermal denaturation was carried out at  $65^\circ\text{C}$  in a heating block, and  $10\text{-}\mu\text{l}$  samples were withdrawn at the following time points: 0, 180, 360, 600, 900, 1,800, 3,000, 4,800, and  $7,200 \text{ s}$ , and mixed with  $2 \times$  Laemmli SDS sample buffer and placed on ice to stop the reaction until analysis was carried out with SDS-PAGE. Gels were stained with Coomassie Blue R-250, digitally scanned, and band intensity quantified with the software GelQuant.NET (<http://biochemlabsolutions.com/GelQuantNET.html>). The results were plotted with KaleidaGraph (<http://www.synergy.com/>).

### Molecular Graphics

Figures 2, 3, and 5 were prepared with Pymol (Schrödinger, LLC). Figure 6 was made with UCSF Chimera.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, three tables, and Supplemental Experimental Procedures and can be found with this article online at doi: 10.1016/j.str.2011.12.009.

### ACKNOWLEDGMENTS

Phage P2 DNA was kindly provided by Prof. Haggård-Ljungquist (Stockholm University). We express our sincere thanks to the entire staff of the protein crystallography beam lines at the Swiss Light Source and, in particular, to Dr. Meitian Wang, Dr. Takashi Tomizaki, Dr. Vincent Olieric, Dr. Martin Fuchs,

Dr. Florian Dworkowski, and Dr. Clemens Schulze-Briese. We are very grateful to Victor Kostyuchenko, who wrote a script for identification of all protein sequences containing the HxH motif in the T4 genome. We thank Jozef Adamcik for AFM imaging of  $\phi$ 92. This work was supported by EPFL, the Swiss National Science Foundation (Grant 31003A-127092/1), and the Scientific and Technological Cooperation Programme Switzerland-Russia. D.S. was supported in part by the Deutsche Forschungsgemeinschaft (DFG) in the framework of DFG Research Unit 548 (Grants Ge801/7-1 and 7-2).

Received: August 30, 2011

Revised: December 2, 2011

Accepted: December 4, 2011

Published: February 7, 2012

### REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W., et al. (2011). The Phenix software for automated determination of macromolecular structures. *Methods* 55, 94–106.
- Aksyuk, A.A., Leiman, P.G., Kurochkina, L.P., Shneider, M.M., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2009). The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J.* 28, 821–829.
- Baker, T.S., and Henderson, R. (2006). Electron cryomicroscopy. In *International Tables for Crystallography Volume F. Crystallography of biological macromolecules*, M.G. Rossmann and E. Arnold, eds. (Dordrecht, Netherlands: Kluwer Academic Publishers), pp. 451–463.
- Bartual, S.G., Otero, J.M., Garcia-Doval, C., Llamas-Saiz, A.L., Kahn, R., Fox, G.C., and van Raaij, M.J. (2010). Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc. Natl. Acad. Sci. USA* 107, 20287–20292.
- Bönemann, G., Pietrosiuk, A., and Mogk, A. (2010). Tubules and donuts: a type VI secretion story. *Mol. Microbiol.* 76, 815–821.
- Chen, B., and King, J. (1991). Thermal unfolding pathway for the thermostable P22 tailspike endorhamnosidase. *Biochemistry* 30, 6260–6269.
- Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* 66, 470–478.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Fokine, A., Battisti, A.J., Bowman, V.D., Efimov, A.V., Kurochkina, L.P., Chipman, P.R., Mesyanzhinov, V.V., and Rossmann, M.G. (2007). Cryo-EM study of the *Pseudomonas* bacteriophage phiKZ. *Structure* 15, 1099–1104.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116, 190–199.
- Gerardy-Schahn, R., Bethe, A., Brennecke, T., Mühlenhoff, M., Eckhardt, M., Ziesing, S., Lottspeich, F., and Frosch, M. (1995). Molecular cloning and functional expression of bacteriophage PK1E-encoded endoneuraminidase Endo NE. *Mol. Microbiol.* 16, 441–450.
- Haggård-Ljungquist, E., Jacobsen, E., Rishovd, S., Six, E.W., Nilssen, O., Sunshine, M.G., Lindqvist, B.H., Kim, K.J., Barreiro, V., Koonin, E.V., and Calendar, R. (1995). Bacteriophage P2: genes involved in baseplate assembly. *Virology* 213, 109–121.
- Israel, J.V., Anderson, T.F., and Levine, M. (1967). *in vitro* MORPHOGENESIS OF PHAGE P22 FROM HEADS AND BASE-PLATE PARTS. *Proc. Natl. Acad. Sci. USA* 57, 284–291.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.
- Kageyama, Y., Murayama, M., Onodera, T., Yamada, S., Fukada, H., Kudou, M., Tsumoto, K., Toyama, Y., Kado, S., Kubota, K., and Takeda, S. (2009). Observation of the membrane binding activity and domain structure of gpV, which comprises the tail spike of bacteriophage P2. *Biochemistry* 48, 10129–10135.
- Kajava, A.V., and Steven, A.C. (2006). Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv. Protein Chem.* 73, 55–96.

- Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553–557.
- Kanamaru, S., Ishiwata, Y., Suzuki, T., Rossmann, M.G., and Arisaka, F. (2005). Control of bacteriophage T4 tail lysozyme activity during the infection process. *J. Mol. Biol.* **346**, 1013–1020.
- Kondou, Y., Kitazawa, D., Takeda, S., Tsuchiya, Y., Yamashita, E., Mizuguchi, M., Kawano, K., and Tsukihara, T. (2005). Structure of the central hub of bacteriophage Mu baseplate determined by X-ray crystallography of gp44. *J. Mol. Biol.* **352**, 976–985.
- Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2003). Three-dimensional structure of bacteriophage T4 baseplate. *Nat. Struct. Biol.* **10**, 688–693.
- Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2005). The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat. Struct. Mol. Biol.* **12**, 810–813.
- Kwiatkowski, B., Boscchek, B., Thiele, H., and Stirm, S. (1982). Endo-N-acetylneuraminidase associated with bacteriophage particles. *J. Virol.* **43**, 697–704.
- Leiman, P.G., and Molineux, I.J. (2008). Evolution of a new enzyme activity from the same motif fold. *Mol. Microbiol.* **69**, 287–290.
- Leiman, P.G., and Shneider, M.M. (2012). Contractile tail machines of bacteriophages. In *Viral Molecular Machines*, M.G. Rossmann and V.B. Rao, eds. (New York: Springer), in press.
- Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2004). Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419–429.
- Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzi, S., Burley, S.K., Almo, S.C., and Mekalanos, J.J. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl. Acad. Sci. USA* **106**, 4154–4159.
- Leiman, P.G., Arisaka, F., van Raaij, M.J., Kostyuchenko, V.A., Aksyuk, A.A., Kanamaru, S., and Rossmann, M.G. (2010). Morphogenesis of the T4 tail and tail fibers. *Virology* **417**, 304–311.
- Leslie, A.G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 48–57.
- Liu, J., Chen, C.Y., Shiomi, D., Niki, H., and Margolin, W. (2011). Visualization of bacteriophage P1 infection by cryo-electron tomography of tiny *Escherichia coli*. *Virology* **417**, 304–311.
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674.
- McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.
- Morris, R.J., Perrakis, A., and Lamzin, V.S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244.
- Murzin, A.G., and Chothia, C. (1992). Protein architecture: new superfamilies. *Curr. Opin. Struct. Biol.* **2**, 895–903.
- Nakayama, K., Takashima, K., Ishihara, H., Shinomiya, T., Kageyama, M., Kanaya, S., Ohnishi, M., Murata, T., Mori, H., and Hayashi, T. (2000). The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol. Microbiol.* **38**, 213–231.
- Nilsson, A.S., and Haggård-Ljungquist, E. (2007). Evolution of P2-like phages and their impact on bacterial evolution. *Res. Microbiol.* **158**, 311–317.
- Otwinowski, Z., and Minor, W. (2006). DENZO and SCALEPACK. In *International Tables for Crystallography. Volume F: Crystallography of Biological Macromolecules*, M.G. Rossmann and E. Arnold, eds. (Dordrecht, the Netherlands: Kluwer Academic Publishers), pp. 226–235.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
- Pukatzi, S., Ma, A.T., Revel, A.T., Sturtevant, D., and Mekalanos, J.J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl. Acad. Sci. USA* **104**, 15508–15513.
- Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., and Leiman, P.G. (2004). The bacteriophage T4 DNA injection machine. *Curr. Opin. Struct. Biol.* **14**, 171–180.
- Santos, S.B., Kropinski, A.M., Ceysens, P.J., Ackermann, H.W., Villegas, A., Lavigne, R., Krylov, V.N., Carvalho, C.M., Ferreira, E.C., and Azeredo, J. (2011). Genomic and proteomic characterization of the broad-host-range *Salmonella* phage PVP-SE1: creation of a new phage genus. *J. Virol.* **85**, 11265–11273.
- Sciara, G., Bebeacua, C., Bron, P., Tremblay, D., Ortiz-Lombardia, M., Lichère, J., van Heel, M., Campanacci, V., Moineau, S., and Cambillau, C. (2010). Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proc. Natl. Acad. Sci. USA* **107**, 6852–6857.
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122.
- Simon, L.D., and Anderson, T.F. (1967a). The infection of *Escherichia coli* by T2 and T4 bacteriophages as seen in the electron microscope. I. Attachment and penetration. *Virology* **32**, 279–297.
- Simon, L.D., and Anderson, T.F. (1967b). The infection of *Escherichia coli* by T2 and T4 bacteriophages as seen in the electron microscope. II. Structure and function of the baseplate. *Virology* **32**, 298–305.
- Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33** (Web Server issue), W244–8.
- Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994). Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265**, 383–386.
- Thomassen, E., Gielen, G., Schütz, M., Schoehn, G., Abrahams, J.P., Miller, S., and van Raaij, M.J. (2003). The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331**, 361–373.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- van Raaij, M.J., Schoehn, G., Burda, M.R., and Miller, S. (2001). Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137–1146.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242.
- Yamashita, E., Nakagawa, A., Takahashi, J., Tsunoda, K.I., Yamada, S., and Takeda, S. (2011). The host-binding domain of the P2 phage tail spike reveals a trimeric iron-binding structure. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 837–841.