

# Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin

Petr G. Leiman<sup>a,1,2</sup>, Marek Basler<sup>b,1</sup>, Udupi A. Ramagopal<sup>c</sup>, Jeffrey B. Bonanno<sup>c</sup>, J. Michael Sauder<sup>d</sup>, Stefan Pukatzki<sup>e</sup>, Stephen K. Burley<sup>d</sup>, Steven C. Almo<sup>c</sup>, and John J. Mekalanos<sup>b,3</sup>

<sup>a</sup>Department of Biological Sciences, Purdue University, West Lafayette, IN 47906; <sup>b</sup>Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115; <sup>c</sup>Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461; <sup>d</sup>SGX Pharmaceuticals, Inc. 10505 Roselle Street, San Diego, CA 92121; and <sup>e</sup>Department of Medical Microbiology and Immunology, University of Alberta, 1-63 Medical Sciences Building, Edmonton, Alberta T6G2H7

Contributed by John J. Mekalanos, December 30, 2008 (sent for review October 24, 2008)

**Protein secretion is a common property of pathogenic microbes. Gram-negative bacterial pathogens use at least 6 distinct extracellular protein secretion systems to export proteins through their multilayered cell envelope and in some cases into host cells. Among the most widespread is the newly recognized Type VI secretion system (T6SS) which is composed of 15–20 proteins whose biochemical functions are not well understood. Using crystallographic, biochemical, and bioinformatic analyses, we identified 3 T6SS components, which are homologous to bacteriophage tail proteins. These include the tail tube protein; the membrane-penetrating needle, situated at the distal end of the tube; and another protein associated with the needle and tube. We propose that T6SS is a multicomponent structure whose extracellular part resembles both structurally and functionally a bacteriophage tail, an efficient machine that translocates proteins and DNA across lipid membranes into cells.**

bacteriophage | membrane | nanomachine | translocation | virulence

**G**ram-negative bacteria use secretion systems to modify their environment by translocating proteins and DNA into the external medium and neighboring cells. Six types of secretion systems have been characterized (1). The recently defined type VI secretion system (T6SS) (2–4) has been implicated in virulence-related processes in multiple bacterial species (5–10). T6SS genes are clustered in pathogenicity islands found in about 25% of all sequenced Proteobacteria and contain 20 or more ORFs (3). Many bacterial genomes contain 1–6 apparently complete copies of the T6SS clusters as well as many more copies of incomplete clusters or individual T6SS genes (6). The T6SS genes are predicted to encode cytoplasmic and membrane-associated proteins, ATPases, lipoproteins, and various substrates, which are typically recognized by virtue of their extracellular secretion ([supporting information \(SI\) Table S1](#)) (11).

The substrates secreted by T6SS lack N-terminal hydrophobic signal sequences and appear in culture supernatant as unprocessed polypeptides (2). Interestingly, the conserved Hcp and VgrG proteins are both secreted and required for the function of the T6SS machine (2, 4, 12, 13). Many VgrG orthologs carry a wide range of C-terminal putative effector domains, which can function as pathogenic factors (13). The appearance of covalently cross-linked actin in J774 macrophages provides the most definitive evidence to date that the *Vibrio cholerae* T6SS can translocate the VgrG-1 protein containing an actin cross-linking domain (ACD) into target cells (13).

Earlier bioinformatic analysis suggested that VgrGs are homologs of the bacteriophage T4 cell-puncturing device, also called the needle or spike (13). The needle consists of 2 proteins, gene product (gp) 5 and gp27, which associate into

the (gp5)<sub>3</sub>-(gp27)<sub>3</sub> complex used by the phage to penetrate through the bacterial cell envelope during infection (14). The distinctive needle-like shape of the complex is due to the C-terminal domain of gp5, which forms a long triple-stranded  $\beta$ -helix. Based on a reasonable sequence similarity between the *V. cholerae* VgrG-1 and bacteriophage Mu gp44 protein, which is very similar structurally to T4 gp27, and on a high probability that VgrG-1 contains a  $\beta$ -helix, it was proposed that VgrG-1 is homologous to the T4 gp5-gp27 complex (13).

The crystal structure of *Pseudomonas aeruginosa* Hcp1, the most abundant T6SS secreted protein, shows that it is a donut-shaped hexamer with external and internal diameters of 85 Å and 40 Å, respectively (4). These hexamers stack on top of each other head-to-tail to form continuous tubes in the crystals. The contacting surfaces of the hexamers can be modified with cysteine mutations competent to forming covalent bonds across these hexameric rings, resulting in chemically stable tubes of various lengths made entirely of Hcp1 hexamers (15). The external and internal diameters of the tubes are virtually identical to those of the bacteriophage T4 tail tube (Fig. 1 *A* and *B*), which is composed of gp19 and is terminated with the gp5-gp27 complex (16, 17).

In this paper, we show further striking similarities between the T6SS proteins and bacteriophage tails. We report the crystal structure of an N-terminal fragment of the *Escherichia coli* CFT073 VgrG protein encoded by ORF c3393. Despite only 13% sequence identity, it shows remarkable structural similarity to the gp5-gp27 complex, as predicted by Pukatzki *et al.* (13). Furthermore, we reexamine the crystal structure of Hcp1 (4) and find that Hcp1 is homologous to the tandem “tube” domain of gp27, which interacts with the T4 tail tube. We also present bioinformatic and biochemical analyses of the ORFs c3385–c3402 from the *E. coli* CFT073 T6SS cluster and their homologs from *P. aeruginosa* and *V. cholerae*. We find that another conserved T6SS protein (e.g., c3402, PA0087,

Author contributions: P.G.L., M.B., and J.J.M. designed research; P.G.L., M.B., U.A.R., J.B.B., J.M.S., S.P., S.K.B., and S.C.A. performed research; P.G.L., M.B., U.A.R., J.B.B., J.M.S., S.K.B., and S.C.A. analyzed data; and P.G.L., M.B., and J.J.M. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The atomic coordinates have been deposited in Protein Data Bank, www.pdb.org (PDB ID code 2P5Z).

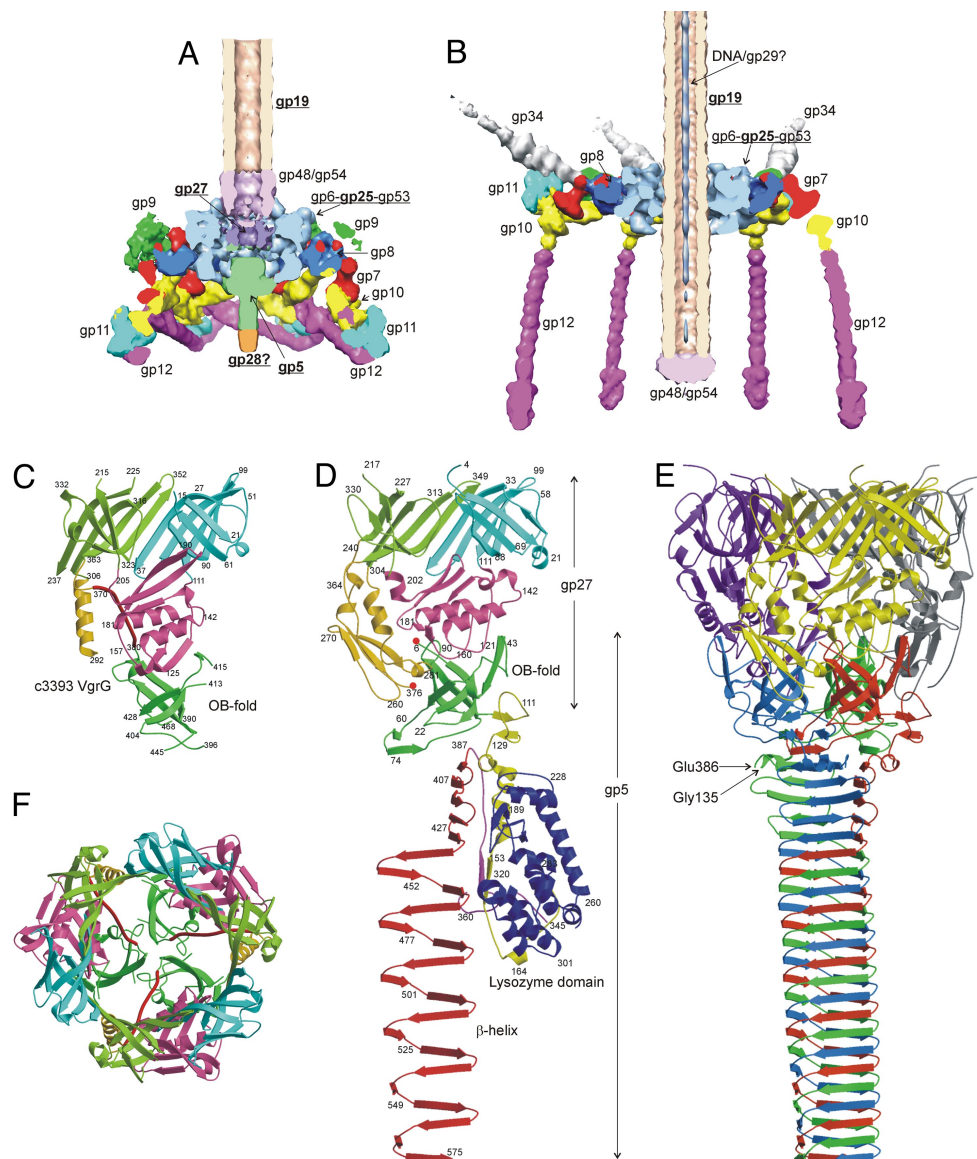
See Commentary on page 4067.

<sup>1</sup>P.G.L. and M.B. contributed equally to this work.

<sup>2</sup>Present address: École Polytechnique Fédérale de Lausanne, Institut de physique des systèmes biologiques, Cubotron, CH-1015, Lausanne, Switzerland.

<sup>3</sup>To whom correspondence should be addressed. E-mail: john.mekalanos@hms.harvard.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0813360106/DCSupplemental](http://www.pnas.org/cgi/content/full/0813360106/DCSupplemental).



**Fig. 1.** Structure of the bacteriophage T4 baseplate and comparison of the *E. coli* CFT073 c3393 VgrG with its T4 homologs, gp5 and gp27. CryoEM reconstructions of the T4 baseplate before (A) and after (B) attachment to the host cell. Component proteins are labeled with their respective gene numbers. The T6SS protein homologs are highlighted in bold and underlined. (C) The crystal structure of the c3393 VgrG. Different domains are colored in distinct colors. The gp27 tube domains are colored cyan and light green. The fragment of the polypeptide chain connecting the gp27 and gp5 modules is shown as a thick red tube. (D) The structures of gp5 and gp27 monomers extracted from the (gp5)<sub>3</sub>-(gp27)<sub>3</sub> complex. The terminal ends of the gp5 and gp27 polypeptide chains, which become fused in the VgrG structure, are highlighted with red dots. (E) A model of the prototypical VgrG is created from the entire (gp5)<sub>3</sub>-(gp27)<sub>3</sub> complex by removing the lysozyme domain. (F) End-on view of the crystal structure of the c3393 VgrG trimer.

VCA0109) is homologous to the T4 baseplate protein gp25. As a component of the T4 baseplate, gp25 is located near the interface between the gp5-gp27 complex and the tail tube (Fig. 1 A and B)(17, 18). Analysis of Hcp protein sequences suggested an evolutionary relationship to the bacteriophage tail tube proteins, including the T4 tail tube protein gp19. This finding is further supported by our observation that *P. aeruginosa* Hcp3, an Hcp variant, polymerizes into tube-like structures in vitro. Finally, the accompanying paper reports that gpV tail tube protein of phage lambda, a protein that is functionally analogous to T4 phage gp19 tube protein, has a structure that is similar to that of Hcp1 (19). Collectively, these data provide strong evidence that the proteins comprising the T6SS apparatus assemble into a large multicomponent structure that is structurally and potentially functionally similar to a bacteriophage tail.

## Results

**Crystal Structure of VgrG Encoded by the *Escherichia coli* CFT073 Gene c3393 Is an Excellent Match to the T4 Cell-Puncturing Device gp5-gp27 Complex Structure.** The N-terminal fragment of VgrG from the uropathogenic *E. coli* CFT073 encoded by gene c3393 (Table S1), consisting of residues 1 through 483 out of 824, was crystallized. The crystal structure has been determined and refined to a resolution of 2.6 Å (PDB ID 2P5Z). The c3393 VgrG structure consists of 2 modules and 5 domains, all of which have counterparts in the structure of the T4 cell-puncturing device, the gp5-gp27 complex (PDB ID 1K28) (14) (Fig. 1). The 2 structures can be superimposed onto each other with a root mean square deviation (RMSD) of 2.7 Å between the 232 equivalent Cα atoms despite exhibiting only 13% overall sequence identity (Table S2). Thus, this VgrG represents a fusion of T4 gp27 and gp5 with the gp27 module being very similar and the gp5 having slight modifications.



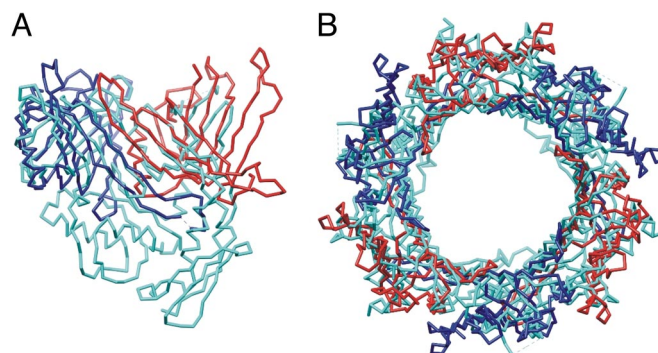
Crystal structures of 2 other VgrG homologs are currently available: gp44 from *E. coli* bacteriophage Mu (PDB ID 1WRU) and the 43 kDa tail protein from *Shewanella oneidensis* MR-1 prophage MuSO2 (PDB ID 3CDD). Both of these bacteriophages have contractile tails similar to T4, but their baseplates are significantly less complex than that of T4. These 2 bacteriophage proteins are homologous at the amino acid sequence level (34% sequence identity), and their structures are very similar (Table S2). Although the precise location of these proteins in the phage tails is unknown, their structural similarity to gp27 strongly suggests that these proteins form a centerpiece of the tail baseplate and that they are translocated across the cell outer membrane into the periplasm during tail contraction.

Residues 1–370 of the c3393 VgrG form the N-terminal, gp27-equivalent module (Fig. 1 C and D). Similar to gp27, this part has a rather complex topology and consists of 4 domains. Two domains with similar structures (residues 1–111 and 205–237 plus 306–370) are called “tube domains” in gp27 because they participate in binding of the entire gp5-gp27 complex to the tail tube. These 7 stranded antiparallel  $\beta$ -barrels can be superimposed onto each other with a nearly perfect 60° rotation around the axis of the trimer, thus forming a pseudohexamer that can interact with the hexameric tail tube. The second function for these tube domains is to serve as an adapter between the trimeric protein complex and the 6-fold symmetric surrounding structure (the rest of the baseplate in T4) (Fig. 1A).

T4 gp5 consists of 3 domains connected via long linkers: the N-terminal oligosaccharide/oligonucleotide-binding (OB)-fold domain, the middle lysozyme domain, and the C-terminal triple-stranded  $\beta$ -helix with a well defined 8 residue-long repeat VxGxxxx. The equivalent of the gp5 OB-fold domain in the structure of c3393 VgrG is the domain of unknown function 586 (DUF586), comprising residues 380–470 and conserved in all known VgrGs. The OB-fold domain is a 5-stranded antiparallel  $\beta$ -barrel with a Greek-key topology, which was originally defined and named as the OB domain (20). It is clear now that this robust fold shows a great variability of binding specificity, albeit the substrate binding site location on the surfaces of many OB-folds overlap roughly (21). The c3393 DUF586 function is similar to that of the gp5 N-terminal domain: both serve as an adapter between the gp27-like module and the  $\beta$ -helical domain.

The secondary structure prediction for the C-terminal part of c3393 VgrG (residues 490–820), which immediately follows the OB-fold domain, shows repetitive, 5 to 10 residue-long  $\beta$ -strands flanked by glycines. There are up to 30 such strands interrupted by a 25 residue-long  $\alpha$ -helix (residues 568–592). These repeated  $\beta$ -strands most probably form a  $\beta$ -helix that is equivalent to the membrane-penetrating triple stranded  $\beta$ -helix in trimeric gp5 (Fig. 1D). The long  $\alpha$ -helix is likely not part of the  $\beta$ -helix but instead forms an “arm” that may interact with another T6SS protein such as ORF c3402 (see below). Thus, VgrG lacks the lysozyme domain and its  $\beta$ -helix is fused directly to the OB-fold domain. Since VgrG is translocated into eukaryotic cells, it does not require a functional glycosidase domain to cross the cell envelope, unlike gp5. Interestingly, gp5 from Vibriophage KVP40 (a close relative of T4) also lacks the lysozyme domain, and the KVP40  $\beta$ -helix is fused to the OB-fold domain, very much like that in VgrG (22). In KVP40, the glycosidase function is most probably encoded in a different protein.

There are several disordered regions in the c3393 crystal structure. In particular, a 55 residue-long stretch between residues 237 and 292 is missing from the atomic model (Fig. 1C). The equivalent residues in the T4 gp5-gp27 structure form a 3-stranded antiparallel  $\beta$ -sheet. Another disordered region of the c3393 structure is between residues 428 and 445. In the gp5-gp27 complex, these residues form an extended loop, which is involved in intersubunit interactions. Remarkably, the c3393 polypeptide chain is in register with that of the gp5-gp27 complex immediately preceding and



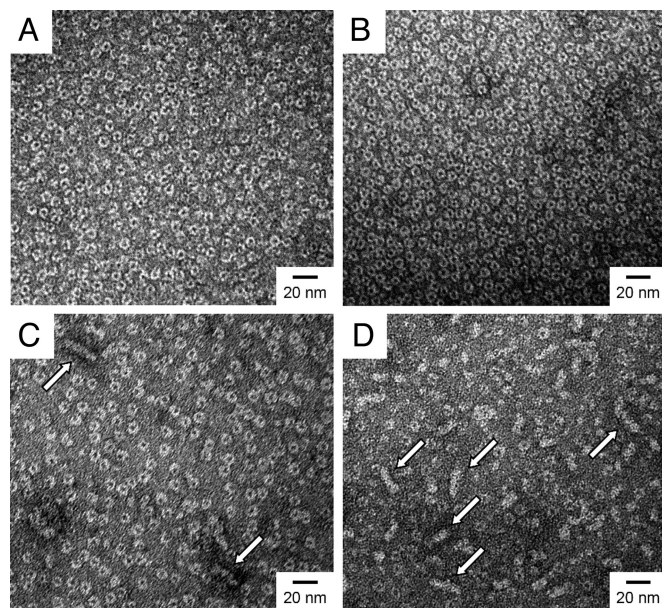
**Fig. 2.** Superposition of the Hcp1 and gp27 structures. (A) Superposition of the crystallographic Hcp1 dimer (red and blue) onto the gp27 monomer (cyan). (B) An end-on view of the superposition of the Hcp1 hexamer (red and blue) onto the entire gp27 trimer (cyan).

succeeding all of the disordered regions (Fig. 1 C and D). The reason for the disorder is unclear, but the disordered regions in c3393 might become ordered upon binding to other components of the secretion system, and the conformations these regions adopt are possibly equivalent to those of the gp5-gp27 complex structure. These assumptions were used to complete an atomic model of the c3393 VgrG N-terminal domain via comparative modeling (23) to include the missing residues (SI Text, Fig. S1A).

Given the trimeric crystal structure of the N-terminal fragment of c3393 (Fig. 1F), it seems highly probably that full length VgrG proteins will adopt a trimeric structure analogous to the (gp5)<sub>3</sub>-(gp27)<sub>3</sub> complex (14). Fig. 1E shows a hypothetical model of a generic VgrG trimer that was created from the (gp5)<sub>3</sub>-(gp27)<sub>3</sub> complex by removing residues 136–385 of gp5, constituting the lysozyme domain and associated linkers. Remarkably, this excision introduces a gap of only 3.9 Å into the polypeptide chain (the Gly-135 and Glu-386 C $\alpha$  distance), suggesting that the entire lysozyme domain can be “cut out” from the gp5 structure with little distortion to the remaining structure. Additional properties of the VgrG sequence, structure, and electrostatic charge distribution are reported in the SI Text and Fig. S2.

Sequence analysis shows that the effector domains are fused to the C termini of many VgrG proteins (e.g., the ACD of VgrG-1 from *V. cholerae*; ref. 13). In the cryoEM map of the bacteriophage T4 baseplate (Fig. 1A), the gp5 C-terminal  $\beta$ -helix also has an extension, which forms an extension of the gp5 C terminus and corresponds to a protein with a molecular weight of approximately 23 kDa matching that of gp28 (17, 24). This protein might be an equivalent of a VgrG effector domain, although the function of gp28 is unknown and its presence at the tip of the  $\beta$ -helix appears to be nonessential for infection of laboratory *E. coli* strains (14, 25). Thus, the VgrG-1 of *V. cholerae* is a fusion of T4 gp5, gp27, and gp28 into a single polypeptide chain, with the gp28 module being the pathogenic factor. Many VgrG genes do not appear to contain any additional domains at their C termini, suggesting that in such T6SS clusters the pathogenicity factor is encoded by a separate protein.

**Hcp Is a Homolog of the Phage Tail Tube Protein.** A close examination of the *P. aeruginosa* Hcp1 crystal structure showed that 2 polypeptide chains of Hcp1 related by crystallographic symmetry can be superimposed, as a rigid body, onto the 2 homologous gp27  $\beta$ -barrel tube domains. 137 C $\alpha$  atoms (a total of 43% of all Hcp1 dimer residues) participate in this superposition, resulting in an RMSD of 2.5 Å (Fig. 2A). The amino acid sequence identity found upon superposition is only 4%. The 2 tube domains of gp27, which show only 3% sequence identity, superimpose with similar agreement. Furthermore, the entire hexamer of Hcp1 can be superimposed

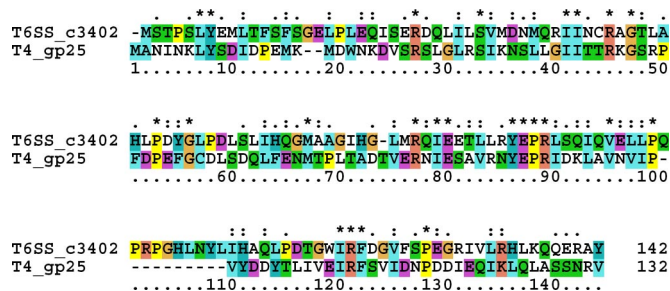


**Fig. 3.** Electron microscopy analysis of oligomerization properties of Hcp proteins. Hcp proteins fused at C terminus to a 3xAla-6xHis-tag were overexpressed in *E. coli* BL21 Star (Invitrogen) and purified by affinity chromatography using Ni-NTA Agarose (QIAGEN). For electron microscopy, the protein samples were diluted to a final concentration of 0.02 mg/ml and negatively stained with uranyl formate. Electron micrographs were recorded with an FEI Tecnai G2 Spirit BioTWIN electron microscope. Arrows point to polymeric structures. (A) PA1512 (Hcp2, *P. aeruginosa* PAO1), (B) c3391 (Hcp, *E. coli* CFT073), (C) PA2367 (Hcp3, *P. aeruginosa* PAO1), (D) An aliquot of purified PA2367 protein sample was denatured by adding solid urea (Fluka) to 8 M concentration. The protein was refolded by 500× dilution into a buffer without urea.

onto the trimeric pseudohexamer formed by the tube domains of the gp27 trimer with an RMSD of 2.8 Å (Fig. 2B).

HHpred (26) analysis shows that *E. coli* CFT073 Hcp ortholog (Table S1) is weakly similar to putative phage tail protein family PF09540 (e-val = 1.5e-4). As revealed by Hidden Markov Models (HMM) -HMM comparison performed by HAlign (27), this protein family exhibits significant homology (e-val = 9.3e-10) to the family of T4-like tail tube proteins gp19 (PF06841). Moreover, the molecular weights of 90% of more than 450 Hcp proteins in InterPro database release 18.0 (28) are in the range of 16.8 kDa to 21.7 kDa (average 18.5 kDa) and bracket the T4 gp19 protein (18.4 kDa). Interestingly, *P. aeruginosa* Hcp1 protein was packed in its crystal matrix into a tube formed by hexamers with 40 Å and 85 Å internal and external diameters and with a pitch of about 41 Å (4). These dimensions are nearly identical to those of the T4 tail tube, which is composed of stacked gp19 hexamers (29). Finally, the authors of the accompanying paper (19) report the structure of the bacteriophage lambda tail tube protein, gpV, and show that it is a structural homolog of Hcp. Taken together, these data strongly support the evolutionary relationship among the tube-forming and tube-associated proteins Hcp, gp27, VgrG, and gpV.

An apparent functional and structural similarity of Hcp to the phage tail tube proteins led us to investigate the in vitro tube polymerization properties of Hcp proteins from *P. aeruginosa* PAO1 (genes PA1512, PA2367) and *E. coli* CFT073 (gene c3391). As shown in Fig. 3, despite the low sequence conservation between Hcp homologs (PA0085 is 17% identical to PA1512, 29% to PA2367, and 23% to c3391), the ability to form ring structures in solution is conserved, and the dimensions are virtually identical to those reported for previously characterized *P. aeruginosa* Hcp1 (PA0085) (4). Only one specimen, Hcp3 or PA2367, showed a weak ability to form short tube-like anamorphous oligomers (Fig. 3C).



**Fig. 4.** ClustalX sequence alignment of T4 gp25 and the T6SS ORF c3402 from *E. coli* CFT073.

Notably, the T4 tail tube protein gp19 (30) and the lambda tail protein gpV (19) are monomers in solution. Once exposed to the assembly nucleus, gp19 becomes competent to form tubes of various lengths because gp19 obtained by dissociation from the tail tube-baseplate complex can self-assemble into tubes (31). Thus, an assembly nucleus and, possibly, a chaperone, which makes the hexameric rings competent for stacking, are likely required for polymerization of Hcp into a tube under physiological conditions.

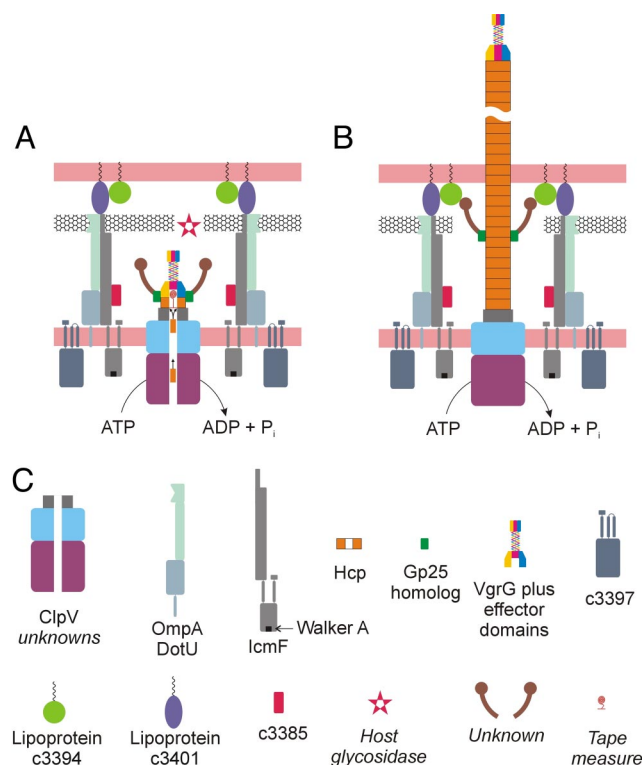
The ClpV protein channel as well as other proteins in the T6SS machine are responsible for translocation of Hcp across the bacterial cell envelope to the external milieu (4). This translocation is likely accompanied by Hcp refolding that can make it competent for polymerization into a tube. To test whether the oligomerization properties of Hcp3 can be altered by refolding it in vitro, we denatured the protein by adding solid urea to a concentration of 8 M and subsequently diluted the sample 500 times into a buffer without urea. As shown in Fig. 3D, the fraction of the tube-like structures, as well as other aggregates, increased dramatically. We performed the same procedure with other Hcp proteins, but no tube-like structures were detected. Instead, the hexameric ring structures similar to those found before refolding were observed (data not shown). It is possible that the tube-like structures formed by the native and refolded Hcp3 might be non-physiological aggregates. However, it appears that the oligomerization and tube-forming properties of some Hcp proteins can be adjusted by refolding them under certain conditions.

**T6SS Gene Cluster Contains a Protein Homologous to the T4 Baseplate Protein gp25.** One of the smallest T6SS proteins, encoded by the *E. coli* gene c3402 (Table S1), is a 15 kDa protein, which belongs to the PF04965 (GPW\_gp25) family. This family includes T4 gp25, which is a structural component of the T4 baseplate. The *E. coli* c3402 protein shows 16.9% sequence identity and 39.4% sequence similarity to the T4 gp25 (Fig. 4).

The rough location of gp25 in the T4 baseplate has been established earlier (Fig. 1A and B) (18). In the native conformation of the baseplate, 6 gp25 subunits surround the central baseplate hub and interact with gp48, which forms the interface between the gp5-gp27 complex and the tail tube (17, 18). Contraction of the tail sheath causes the central hub displaying the gp5-gp27 needle complex to be pushed out from the baseplate by the tail tube. In this contracted configuration, 6 gp25 subunits form a ring encompassing the gp19 tube (16).

Gp25 is one of the most conserved proteins in all phages with contractile tails, which agrees well with its key role as a connector between the central and peripheral parts of the baseplate. Thus, the presence of a gene encoding a gp25-like protein in T6SS gene clusters suggests that it plays a role in producing a baseplate-like structure, which interacts with the corresponding VgrG needle and Hcp tube during assembly or function of the T6SS apparatus.





**Fig. 5.** Structure and assembly of the T6SS apparatus. (A and B) Two steps are shown. See text for details. Although the model is based on some reported protein-protein interactions, predicted membrane topologies, and subcellular localization, the majority of the detail presented is speculative. (C) Explanation of color coding and labeling. The proteins, whose names are given in italic, have not been identified in the T6SS cluster yet.

## Discussion

In this report we have provided evidence that three components of the T6SS apparatus (VgrG, Hcp, and the T4 gp25-like protein) are homologous to bacteriophage tail proteins that are located in the close proximity in the T4 tail. These findings suggest a common evolutionary origin of a major part of the T6SS machine and membrane-penetrating tails of bacteriophages. Based on the known structure of T4 bacteriophage tail complex, on the protein localization predictions, and on the already known protein-protein interactions (Table S1), we have constructed a model for the assembly and function of the T6SS apparatus (Fig. 5). The model borrows from the working models of T2SS and T4SS, where a small subunit of a filament such as a pilin serves the role of a “piston” or extension device to drive proteins out of the cell as it polymerizes (1).

The IcmF and DotU proteins are the conserved inner membrane components of T6SS, which interact in the assembly of functional T4SS of *Legionella* (32). The IcmF N-terminal domain of about 350 residues is predicted to be in the cytoplasm. It contains a conserved ATP/GTP binding site, albeit, in the case of *Edwardsiella tarda*, the conserved Walker A motif is not necessary for T6SS functionality (12). A transmembrane helix connects the IcmF N-terminal domain to the putative periplasmic C-terminal domain of about 600 residues. In *E. tarda*, the C-terminal domain interacts with a conserved lipoprotein, DotU, and at least one other conserved protein of unknown function, EvpA (12). The observed protein-protein interactions suggest that the lipoprotein, DotU, and EvpA are most likely localized to the periplasm. As shown recently, a lipoprotein from the *E. coli* T6SS cluster is localized to the outer membrane and, indeed, faces to the periplasm (33). Thus IcmF and the lipoprotein might form a continuous channel interconnecting the

inner and outer membranes. Interestingly, about 30% of T6SS-related DotU homologs are fused to an OmpA domain responsible for peptidoglycan binding and thus DotU-OmpA, interacting with IcmF, might be involved in the stabilization of the periplasmic part of the T6SS complex (Fig. 5A). In contrast to IcmF, DotU, and the lipoprotein, no signal sequence can be predicted for EvpA that is therefore likely a T6SS substrate targeted to periplasm. In many organisms, secretion of T6SS substrates was shown to be dependent on ClpV, a ClpB-like ATPase (4, 12, 34). However, unlike ClpB, ClpV failed to solubilize aggregated proteins in vitro (35). Nevertheless, ClpV may function to unfold and translocate the T6SS components across the cytoplasmic membrane when complexed with other T6SS components.

It is known that an assembled T4 phage baseplate structure is necessary to initiate tube protein polymerization (30, 31). We propose that ClpV, IcmF, DotU, and other components of the T6SS apparatus assemble in and near the inner membrane and transport VgrG, Hcp, EvpA, and the gp25-like protein into the periplasm by an ATP-dependent process. These proteins assemble into a structure analogous to the centerpiece of the baseplate of T4 phage, thus initiating polymerization of Hcp monomers into a tube. New Hcp monomers are translocated into the periplasm and are added to the last ring of the Hcp tube closest to the inner membrane. Eventually, the VgrG/Hcp complex passes through a channel in the outer membrane, which is either formed by the lipoprotein or induced by VgrG penetration (Fig. 5B). Further tube elongation pushes VgrG carrying the C-terminal effector domain into the cytoplasm of a target cell. We can also speculate that ClpV recognizes other substrates that are then translocated through the Hcp tube and delivered into target cells. This process is analogous to membrane penetration by tailed phages, which is presumed to use the energy of the contracting sheath to push the tail needle complex through the cell outer membrane. It is not clear whether the Hcp tube polymerization itself would provide enough force to puncture the outer membrane and translocate the VgrG trimer across the target cell membrane or whether a yet unidentified T6SS component functionally analogous to a phage sheath protein is involved in this process.

The remarkable homology of several T6SS components with proteins of bacteriophage tails raises the question of whether these components of a membrane-penetrating nanomachine evolved first in the context of a phage tail or as a bacterial secretion system. Such diversification of function is a common fact in the evolution of molecular assemblies and is present in the context of bacterial secretion systems. The structural homology displayed by filamentous phages, type IV pili, and the T2S apparatus provides one example, while the similarities found between bacterial flagella and the T3S machine provide another (1). Evolution drives diversity of function from initially successful prototypes. Understanding how these structurally related machines perform similar but clearly distinct tasks will likely provide fascinating insights into the evolutionary process, function, and regulation of these complex assemblies.

## Materials and Methods

**Cloning, Crystallization and Structure Determination of c3393.** The structure of the *E. coli* C3393 VgrG was determined as part of the NYSXRC initiative (the target gene name is NYSXRC-10105b). The target gene containing residues 2–481 was amplified via PCR from the *E. coli* CFT073 genomic DNA and inserted into the pET26b vector (modified for TOPO directed cloning), which drives the expression of the protein with a non-cleavable C-terminal (His)<sub>6</sub>-tag (Invitrogen). BL21(DE3)-Codon+RIL cells (Stratagene) were transformed with this vector and grown overnight in HY medium (Mediatech, Inc.) at 37 °C until OD<sub>600</sub> reached approximately 1. The temperature was reduced to 22 °C for 20 min and the SeMet buffer (Mediatech, Inc.) was added. Growth was continued for 20 min and expression was induced by addition of 1 mM (isopropyl-β-D-thiogalactopyranoside) IPTG. After additional 21 h of growth, the cells were harvested and frozen at –80 °C.

The cells were resuspended and lysed by sonication; the lysate was clarified by

centrifugation at 38,900 g for 30 min. The protein solution was applied to a Ni-NTA column (Qiagen), was washed with Buffer A (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 10 mM imidazole, 10 mM methionine, and 10% glycerol), and eluted with buffer A containing 500 mM imidazole. The solution was concentrated by Amicon ultrafiltration (Millipore), applied to an S200 gel filtration column, and chromatographed with 10 mM Hepes pH 7.5, 150 mM NaCl, 10 mM methionine, 10% glycerol, and 5 mM DTT. The protein was concentrated to 10 mg/ml.

Single crystals were obtained by mixing 1  $\mu$ l of the protein at 10 mg/ml with 1  $\mu$ l of precipitant (100 mM Hepes pH 7.5, 21% PEG 3350) followed by vapor diffusion equilibration against 100  $\mu$ l of the same precipitant at room temperature. Following cryoprotection with mother liquor supplemented with 25% ethylene glycol, the crystals were immersed in liquid nitrogen. Single wavelength anomalous diffraction data extending to 2.6 Å resolution were collected at the selenium peak wavelength at the Argonne National Laboratory Advanced Photon Source Sector 31-ID. Diffraction from these crystals was consistent with space group P6<sub>3</sub> (a = b = 116.35 Å, c = 80.58 Å), with one molecule per asymmetric unit and the 3-fold axis of the biological trimer corresponding to a crystallographic 3-fold.

Six selenium sites were located using SHELXD and density modified phases were calculated with SHELXE (36). Following rounds of automated and manual model building with ARP/wARP (37) and Coot (38), refinement with Refmac (39) converged at  $R_{\text{work}} = 21.7\%$  and  $R_{\text{free}} = 26.4\%$ . The final model of *E. coli* C3393 consisted of amino acids H15 to K468 (disordered regions correspond to 2–14, 71–80, 216–224, 238–291, 392–395, 414, 429–443, 469–481, and all cloning artifacts at both termini) and 62 waters. The structure has been deposited in the Protein Data Bank as entry 2PSZ.

**Expression, Purification, and Electron Microscopy of Hcp Proteins.** The ORFs PA1512 (hcp2) and PA2367 (hcp3) from *P. aeruginosa* PAO1 and c3391 (hcp1) from *E. coli* CFT073 were cloned into the pET24C6H expression vector, which introduces a C-terminal fusion of the protein to a 3xAla-6xHis-tag. The proteins were over-expressed using *E. coli* BL21 Star (Invitrogen). Cells were grown at 37 °C in a Luria Broth medium in the presence of kanamycin at 50  $\mu$ g/ml and were induced with 0.5 mM IPTG when the optical density at 600 nm reached 0.6. The expression was performed, at 37 °C for 4 h. The cells were harvested and resuspended in 50 mM Tris-HCl (pH 7.4), 500 mM NaCl, 30 mM imidazole, 20 mM

$\beta$ -mercaptoethanol; and lysed by sonication. The proteins were purified by affinity chromatography using Ni-NTA Agarose (QIAGEN). An aliquot of purified protein sample was denatured by adding solid urea (Fluka, 02493 BioUltra) to 8 M concentration. The proteins were refolded by 500 $\times$  dilution into a buffer containing 50 mM Tris-HCl pH 7.4, 0.15 M NaCl and 1 mM  $\beta$ -mercaptoethanol.

For electron microscopy, the protein samples were diluted to a final concentration of 0.02 mg/ml in a solution containing 50 mM Tris-HCl pH 7.4, 0.15 M NaCl, and 1 mM  $\beta$ -mercaptoethanol and negatively stained with uranyl formate. Electron micrographs were recorded with an FEI Tecnai G2 Spirit BioTWIN electron microscope.

**Molecular Graphics.** Fig. 1 is drawn using MolScript (40) and Raster3d (41). Fig. 2 is drawn using UCSF Chimera (42).

**Note Added in Proof.** The *V. cholerae* ClpV protein was recently shown to be involved in remodeling of a tubular structure formed by two conserved and essential members of T6SS, called VipA and VipB (43). We have found that the dimensions, symmetry, and organization of the VipA/B structure resemble those of the contracted T4 tail sheath or “polysheath” (44), suggesting that the VipA/B tube is a structural and possibly functional homolog of the T4 phage tail sheath. Thus, VipA/B may provide energy to the T6SS-mediated secretion and membrane insertion process through conformational changes that mimic those that occur during phage tail contraction. However, it is unclear whether such changes occur in the bacterial cytosol, in the envelope, or on the cell surface.

**ACKNOWLEDGMENTS.** We gratefully acknowledge the efforts of all members of the NYSGXRC, past and present. The NYSGXRC is supported by NIH Grant U54 GM074945. Use of the Advanced Photon Source was supported by the Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the SGX Collaborative Access Team (SGX-CAT) beam line facilities at Sector 31 of the Advanced Photon Source was provided by SGX Pharmaceuticals, Inc., who constructed and operates the facility. Work in J.J.M.’s laboratory on Type VI secretion has been supported by Grant AI-18045 and AI-26289 from National Institutes of Allergy and Infectious Disease. P.G.L.’s work in Michael Rossmann’s laboratory is supported by the National Science Foundation grant MCB-0443899. M.B. was supported by EMBO fellowship ALTF 350-2008.

- Economou A, et al. (2006) Secretion by numbers: Protein traffic in prokaryotes. *Mol Microbiol* 62:308–319.
- Pukatzki S, et al. (2006) Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the Dictyostelium host model system. *Proc Natl Acad Sci USA* 103:1528–1533.
- Bingle LE, Bailey CM, Pallen MJ (2008) Type VI secretion: A beginner’s guide. *Curr Opin Microbiol* 11:3–8.
- Mougous JD, et al. (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* 312:1526–1530.
- Folkesson A, Lofdahl S, Normark S (2002) The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res Microbiol* 153:537–545.
- Schell MA, et al. (2007) Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol Microbiol* 64:1466–1485.
- Shalom G, Shaw JG, Thomas MS (2007) In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology* 153:2689–2699.
- Rao PS, Yamada Y, Tan YP, Leung KY (2004) Use of proteomics to identify novel virulence determinants that are required for *Edwardsiella tarda* pathogenesis. *Mol Microbiol* 53:573–586.
- Gray CG, Cowley SC, Cheung KK, Nano FE (2002) The identification of five genetic loci of *Francisella novicida* associated with intracellular growth. *FEMS Microbiol Lett* 215:53–56.
- Nano FE, Schmerk C (2007) The *Francisella* pathogenicity island. *Ann NY Acad Sci* 1105:122–137.
- Filloux A, Hachani A, Blevess S (2008) The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology* 154:1570–1583.
- Zheng J, Leung KY (2007) Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* 66:1192–1206.
- Pukatzki S, et al. (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.
- Kanamaru S, et al. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature* 415:553–557.
- Ballister ER, et al. (2008) In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci USA* 105:3733–3738.
- Leiman PG, et al. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* 118:419–429.
- Kostyuchenko VA, et al. (2003) Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol* 10:688–693.
- Watts NR, Coombs DH (1990) Structure of the bacteriophage T4 baseplate as determined by chemical cross-linking. *J Virol* 64:143–154.
- Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR (2009) The structure of the phage lambda major tail protein: A common evolution for all long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci USA*, 10.1073/pnas.0900044106.
- Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *EMBO J* 12:861–867.
- Arcus V (2002) OB-fold domains: A snapshot of the evolution of sequence, structure and function. *Curr Opin Struct Biol* 12:794–801.
- Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG (2004) The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol* 14:171–180.
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.
- Kostyuchenko VA, et al. (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat Struct Mol Biol* 12:810–813.
- Kanamaru S, et al. (1999) The C-terminal fragment of the precursor tail lysozyme of bacteriophage T4 stays as a structural component of the baseplate after cleavage. *J Bacteriol* 181:2739–2744.
- Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–248.
- Soding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960.
- Mulder NJ, et al. (2007) New developments in the InterPro database. *Nucleic Acids Res* 35:D224–D228.
- Moody MF, Makowski L (1981) X-ray diffraction study of tail-tubes from bacteriophage T2L. *J Mol Biol* 150:217–244.
- Wagenknecht T, Bloomfield VA (1977) In vitro polymerization of bacteriophage T4D tail core subunits. *J Mol Biol* 116:347–359.
- Poglavozov BF, Nikolskaya TI (1969) Self-assembly of the protein of bacteriophage T2 tail cores. *J Mol Biol* 43:231–233.
- Sexton JA, et al. (2004) *Legionella pneumophila* DotU and IcmF are required for stability of the Dot/Icm complex. *Infect Immun* 72:5983–5992.
- Aschtgen MS, et al. (2008) SciN is an outer membrane lipoprotein required for Type VI secretion in enterohaggregative *Escherichia coli*. *J Bacteriol* 190:7523–7531.
- Weibezahn J, et al. (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* 119:653–665.
- Schlieker C, Zentgraf H, Dersch P, Mogk A (2005) ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria. *Biol Chem* 386:1115–1127.
- Pape T, Schneider TR (2004) HKL2MAP: A graphical user interface for phasing with SHELX programs. *J Appl Crystallogr* 37:843–844.
- Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
- Kraulis P (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Merritt EA, Bacon DJ (1997) Raster3D: Photorealistic molecular graphics. *Methods Enzymol* 277:505–524.
- Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
- Bönemann G, et al. (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for Type VI protein secretion. *EMBO J* 28:315–325.
- Moody MF (1967) Structure of the sheath of bacteriophage T4. I. Structure of the contracted sheath and polysheath. *J Mol Biol* 25:167–200.